

Joint CADDY Steering Group

Format Specification

CADDY 1.1

Document Interchange Format
For Pesticides Registration Applications

September 23, 1997

Copyright

© 1995 Hewlett Packard GmbH.

Licenses and Trademarks

TIFF is a trademark of Aldus Corporation. MS-DOS is a registered trademark of Microsoft Corporation. UNIX is a trademark of Bell Laboratories.

Instead of putting a trademark symbol in every occurrence of other trademarked names, we state that we are using the names only in an editorial fashion, and to the benefit of the trademark owner, with no intention of infringement of the trademark.

Acknowledgments

This document references other standards like TIFF, ECMA standards and others. We like to thank all the people that have worked on these standards.

Experience from the DAMOS project was helpful too. The goal of this project was to standardize a transfer interface for pharmaceutical dossiers. DAMOS stands for Drug Application Methodology with Optical Storage. We like to thank all the contributors of DAMOS for their valuable work.

The content of this document bases on a lot of discussions and talks with people from registration departments and authorities. Thanks to all for lending their time and knowledge to this effort.

This document was prepared by

Dipl. Inform. Michael Groß
Hewlett Packard GmbH
Professional Services
Hewlett-Packard-Straße
D-61352 Bad Homburg
Germany

For questions regarding the joint steering group and the standardization process please contact

Alberik Scharpé
Commission of the European Communities
Directorate General VI - Agriculture
Rue de la Loi 84
B-1040 Brussels
Belgium

Dr. Jürgen Wenzelburger
Bayer AG
Geschäftsbereich Pflanzenschutz
Pflanzenschutzzentrum Monheim
D-51368 Leverkusen
Germany

Table of contents

- Introduction5
 - About this format specification 5
 - CADDY Version 1.0 5
 - CADDY Version 1.1 5
 - Future versions 5
 - Changes to CADDY 1.0 5
 - Electronic submissions of plant protection products 6
 - Standardization objectives 6
 - Major assumptions 7
 - Retrieval possibilities 7
 - Application of the standard for pesticides registrations 8
- Media 9
- Information filing 10
 - Volume sets 10
 - Distribution of page files to the volumes (recommendation) 11
 - Volume set ID 11
 - Volume names 11
 - Label files 12
 - Volume labels 13
 - Creating new versions by adding supplements 13
 - Index files 14
 - Basic index files 14
 - Data dictionary 14
 - Changes history 15
 - Page files 15
 - Unique page ID 15
 - Confidential pages and volumes 16
 - Copies of page files 16
 - Additional files not covered by CADDY (optional) 16
 - Frequently asked questions 16

- Dossier database 17
 - Database table attributes 19
 - Dossier header 19
 - Report attributes 20
 - Table of contents 21
 - Page table 21
 - Confidentiality of pages, reports and TOC nodes 21
 - Frequently asked questions 21
- Index file format 22
 - The basic data types 22
 - Storing the tables to the index files 24
 - General index file format 25
 - Frequently asked questions 25
- Image file format 26
 - TIFF 6.0 baseline bilevel images 27
 - Informational fields 28
 - Compression schemes 28
 - Paper formats 29
- Conformity Testing 30
- References 32
- Appendix A Structure definition of CADDY 1.1 index files 33
 - Additional remarks on fields 33
 - Index structure changes from CADDY 1.0 to CADDY 1.1 35
 - Frequently asked questions 36
 - Index structure of CADDY 1.1 37

Introduction

About this format specification

This document describes CADDY, an electronic dossier interchange and archiving format for pesticides applications. CADDY stands for Computer Aided Dossier and Data supply. No statement is made concerning organizational or legal aspects. This technical specification has to be seen in the context of the whole standardization work.

CADDY Version 1.0

The first version released was called CADDY 1.0. This version was produced as a result of the discussion between EC, EU Member States and ECPA members about a common dossier interchange format on electronic media. The specification for CADDY 1.0 (Version 0.7) was published after being accepted by EC, EU Member States and ECPA.

CADDY Version 1.1

The current version is called CADDY 1.1. This enhanced version was created from CADDY 1.0 by adding some minor adjustments that makes it usable in the United States and Canada too. This specification (CADDY 1.1) contains minor adjustments that were agreed upon by the Joint CADDY Steering Group since the CADDY 1.0 (Version 0.7) was released.

Future versions

An attempt will be made to add functionality in such a way as to minimize compatibility problems with CADDY dossiers that were generated according to CADDY 1.1. The readability of CADDY 1.1 dossiers by future CADDY 2.0 viewers has to be achieved.

Changes to CADDY 1.0

To allow easy finding of changed parts, new and changed paragraphs are marked by a line on the right margin.

The following changes to the Format Specification CADDY 1.0 / V0.7 were agreed upon by the Joint CADDY Steering Group:

- CIPAC number should be added to index data
- CAS number should be added to index data and on label
- More than one active substance can be covered (change of entity relationship model !)
- Hierarchical number in TOC was added
- Correction of country code in volume set ID examples

After circulation of a draft version the following changes were added:

- Clarification of the confidential flag handling for pages
- Adjustment of the regulation field length

Changed chapters

- Introduction
- Information Filing
- Dossier Database
- Appendix A

Electronic submissions of plant protection products

At the beginning of 1995 a discussion about the development of an electronic dossier exchange standard for pesticides registration applications in Europe started. The need for standardization of an electronic format for the submission of pesticides dossiers is obvious. Representatives from regulatory authorities and industry expressed their strategic goal in the following statement:

To facilitate

- the provision of dossiers for pesticides to regulatory authorities,
- the long-term archiving of such dossiers,
- and the accessibility of information contained in such dossiers

in a cost-effective manner using electronic media.

The Joint EU Member States / ECPA Data Transfer Steering Group (DTSG) consisting of five experts of the EU and five experts of the ECPA registration task force was established in June 1995 to work out a solution to meet this strategic goal.

In mid 1996 the US and Canadian authorities and industry joined the CADDY steering group. Representatives from the Pesticide Regulatory Management Agency of Canada (PMRA) and the Canadian industry and representatives from the Environmental Protection Agency (EPA) of the United States and the American Crop Protection Association (ACPA) joined the group. The common group (Joint CADDY Steering Group (JCSG)) has developed a specification for CADDY retrieval software. During this development some changes to the format specification were discussed and agreed. These changes are included in the current document.

Standardization objectives

It was clearly outlined by authorities and industry, that the new electronic submissions are not intended to replace paper, but rather to reduce the amount of necessary paper copies for a submission. Therefore, the electronic submission has to represent the content and format of the complete paper package.

The discussions have led to a list of objectives for the format specification:

- The first release should be very simple and is not intended to cover more than absolutely necessary.
- The user of a dossier can use an electronic submission instead of a paper copy if desired.
- The page file format should be readable by a wide variety of standard applications.
- The index file format should be readable by a wide variety of standard applications.

- To implement compilation software and interfaces to existing in-house systems should be as simple as possible.
- To implement retrieval software should be as simple as possible.
- The standard format should have a modular design to allow for future extensions.

Major assumptions

The following basic decisions were made by the DTSG:

- The medium chosen is CD-ROM.
- All pages of the dossier are represented in TIFF, which is readable by a wide variety of standard imaging applications.
- The index information is represented in a text format, which is readable by standard database applications.

Retrieval possibilities

As a result of the discussions the necessary retrieval possibilities were identified.

- Retrieval of reports, using a report list that covers all reports of the dossier.
- Retrieval of reports and pages, using a table of contents (TOC) for the dossier. This TOC should reflect the structure of the entire dossier.
- Retrieval on page level enables the user to display the desired pages directly if the page number is known.

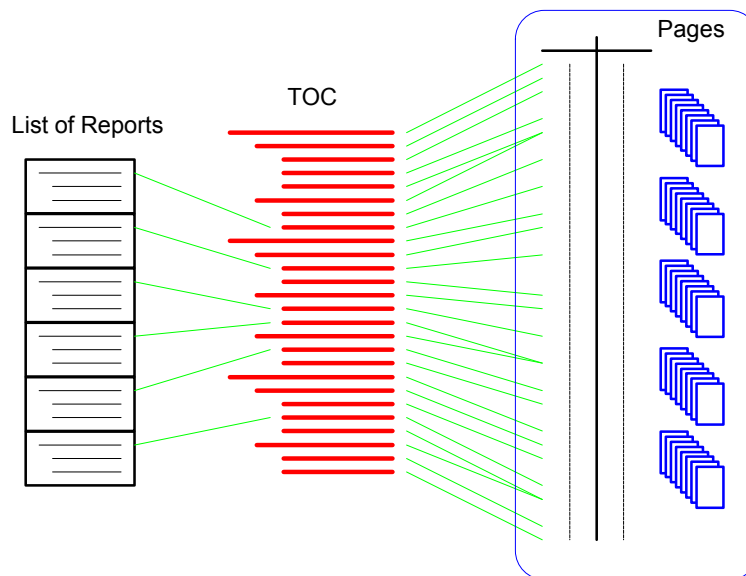


Figure 1 Retrieval possibilities

Application of the standard for pesticides registrations

This document describes the format of pesticides registration dossiers on electronic media. A guidance for the application of this format is not given here.

Data Security

The DTSG has discussed the topic of data security, too. As in the past with paper documents the responsibility for data security is with the receiver whereto the CADDY dossier is delivered. Volumes which are classified as confidential and the ownership of reports are different matters. More information on data security will be found in the Application Guide.

Next steps

The following documents and software should be available, before CADDY 1.1 is used to submit pesticides registration dossiers:

- ⇒ Example application on CD-ROM
(Short example for the application of the CADDY standard)
- ⇒ Application Guide: How to use the format specification to file
a pesticides registration dossier on electronic media
- ⇒ CADDY Retrieval software
- ⇒ CADDY Conformity test software

Media

There was an urgent need to choose a medium for transfer and storage of pesticides registration dossiers which is easy to handle by the user, secure and has sufficient storage capacity to store dossiers.

The medium which was chosen for transfer and storage of the data is CD-ROM. A CD-ROM is an evolution of the CD-Audio ("Red Book") technology. It is readable by a wide variety of computers and operating systems. Furthermore, it is well standardized and offers enough storage capacity and data stability. For the CADDY standard it is also allowed to use the recordable CD-Write-Once (CD-WO), also known as CD-Recordable (CD-R), beside the CD-ROM disk technology. The medium, physical formats, and basic CD-R recording system are specified in part II of the "Orange Book". The CD-R technology is well suited for the creation of low volume CD distributions in a normal office environment.

The "Yellow Book" is an established standard for recording data on CD-ROM. It defines two modes. Mode 1 is an error protected mode which is used for most of today's CD-ROM applications. The "Orange Book" supports the two following recording methods: disk-at-once (DAO) and incremental. With DAO the whole disc has to be recorded in one uninterrupted stream. For compatibility reasons this is the required recording method of CADDY.

CADDY media have to conform to

Yellow Book, Mode 1 (and in case of CD-R: Orange Book, DAO)
--

The ECMA standard ECMA-119 describes a commonly used logical format of CD-ROMs. To be compatible to a wide range of existing hardware and software products, interchange level 1 was chosen as a restriction. The 2nd Edition of ECMA-119 is technically identical to the ISO standard ISO 9660.

CADDY media have to conform to

ECMA-119, Interchange Level 1

In the future it might be necessary to choose additional media types for new CADDY versions. But even in this case, it will be possible to read existing dossiers on CADDY CD-ROM or to copy them to the new media as long as the hardware for reading CD-ROMs is available.

Information filing

A single medium is called a volume. The entire dossier is contained on a set of volumes. There are three different types of files stored on the volumes of a volume set:

- **Page Files**
All pages of the dossier are stored in image format. For every page there is a single image file on one of the volumes. The image file format is described in chapter "Image file format".
- **Index Files**
A set of index files is stored on the last volume of the submission. All the index files have a common structure format (general index file format), which can be easily read by common database applications. The index file format is described in chapter "Index file format".
- **Label File**
A label file is stored on every volume to allow an easy identification of the volume.

Using the index files it is possible to build up a database for the dossier. This database, called dossier database, is used for retrieval purposes. It contains the table of contents and the attributes of the dossier and the reports. Furthermore, the dossier database contains location information for the access to the page files.

Volume sets

A submission consists of 1 to n volumes (volume set).

Pages are stored in TIF-Format on the volumes. This format is described in detail in chapter "Image file format". The pages are spread over the volumes of the volume set.

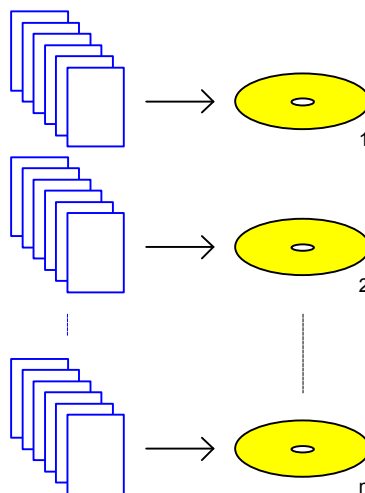


Figure 2 Spread pages on volumes

A single page file cannot be spread across two volumes.

The last volume is called the current index volume. On this volume the current index files are stored in addition to the page files.

Distribution of page files to the volumes (recommendation)

To minimize the necessity of changing the media in the drives we recommend to choose the storage locations for the pages using the following algorithm:

Stored on volume 1	pages 1 .. a
Stored on volume 2	pages a+1 .. b
..	
Stored on the last volume	pages y+1 .. z

It is highly recommended not to spread the pages of a report to different volumes. However it may be necessary to choose another distribution on the volumes. So the retrieval software cannot rely on any special distribution of the page files.

Volume set ID

To identify a volume set, every volume set will get an unique volume set ID, which is build using the following three subparts:

[volume set ID] ::= [notifying company shortcode] [country shortcode] [current dossier number]

[notifying company shortcode] ::= Notifying company or in case of a task force the leading company (short form with 3 characters)

[country shortcode] ::= Country code of the notifying company (short form with 2 characters) (ISO 3166 - Alpha-2-code)

[current dossier number] ::= Current number of dossier generated by the notifying company (padded with leading zeros to 3 characters)

This ensures, that the notifying company is able to guarantee the uniqueness of the volume set IDs of its submissions.

Example: DOEGB001

Volume names

The single volumes are identified by their volume names. These volume names consist of the unique volume set ID and the volume number. The first volume will get the number 1. The other volumes will be numbered consecutively.

[volume name] ::= [volume set ID] "-" [volume number]

[volume number] ::= volume number padded with leading zeros to 3 characters
The volume number is unique within the volume set.

Example: DOEGB001-001

Label files

Every volume has to contain a file named "label.txt" in its root directory. This file contains the following lines of text:

	[volume name]
	[volume type]
	[version string]
	[notifying company]
	[country name]
	[active substance]
	[CIPAC code]
	[CAS code]
[volume type]	Either " confidential " or " index volume " or "."
[version string]	CADDY version number followed by the version date in English format. (see example below and structure definition in Appendix A)
[notifying company]	up to 30 characters (longer names are to be shortened)
[country name]	up to 30 characters (longer names are to be shortened) contains the country name of the notifying company
[active substance]	up to 30 characters contains the clipped name of the active substance
[CIPAC code]	up to 4 characters contains the CIPAC code of the active substance
[CAS code]	up to 11 characters contains the CAS code of the active substance

If an information item like CIPAC code is not available, the line has to contain "." instead.

The file "label.txt" has to be coded using the general index file format described in chapter "Index file format". This file corresponds to a table containing only one column of type [text] (Restriction: The use of the escape character and the delimiter character is not allowed inside this column). The advantage of using this format for the label file is that the file can be easily read by the receiver of a CD-ROM using a simple text viewer and it can be easily read into a database using the same mechanism as it is used for the index files.

Example:

```

DOEGB001-001
.
1, January 1, 1996
DowElanco Europe
United Kingdom
Fluroxypyr
431
123456-37-5
    
```

Volume labels

Every medium has to be labeled by a printed label containing the information inside the file "label.txt".

Example (for three printed labels):

BAYDE001-001 Version 1, April 7, 1995 BAYER AG Germany Cyfluthrin / 385 / 68359-37-5	BAYDE001-002 confidential Version 1, April 7, 1995 BAYER AG Germany Cyfluthrin / 385 / 68359-37-5	BAYDE001-003 index volume Version 1, April 7, 1995 BAYER AG Germany Cyfluthrin / 385 / 68359-37-5
--	---	---

Creating new versions by adding supplements

It is possible to add an additional set of volumes to an already existing volume set. In this case the additional volumes are called a supplement. When adding the supplement, you will get a new version of the whole dossier and the last volume of the supplement will be the current index volume.

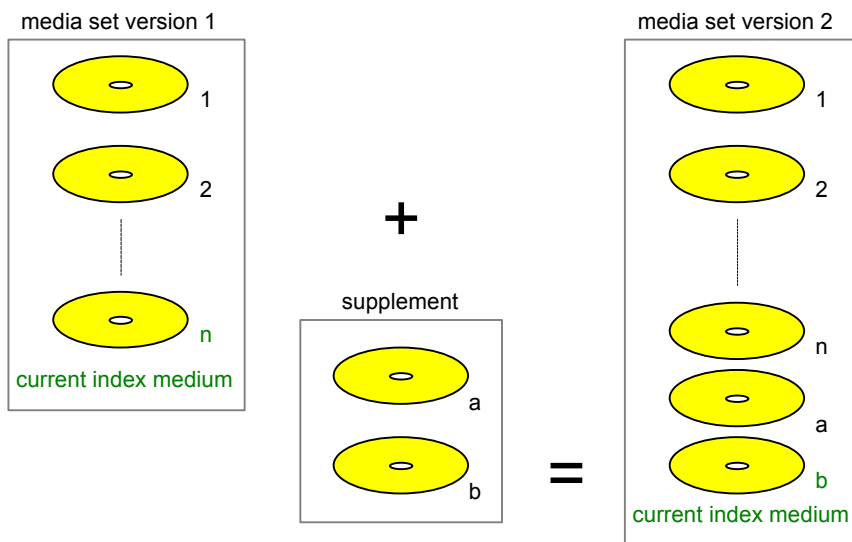


Figure 3 Adding supplements

Index files

All index files are stored in a directory with the name "INDEX" on the current index volume. The parent directory of the directory "INDEX" has to be the root directory. If one of the index files is empty, this file is not stored on the current index volume.

There are three groups of index files

- Basic index files
are necessary to build the dossier database
- Data dictionary
will simplify the downward compatibility of software for future CADDY versions
- Changes history
contains information about changes made in the dossier versions

Basic index files

report list

contains a list of all reports contained in the current version

table of contents

contains a hierarchical table of contents of the current version

page table

contains a list of all pages of the current version

version history

contains a list of all versions of the dossier

volume list

contains a list of all volumes belonging to the current volume set

header attribute sets

contains the history of dossier attributes for all versions of a dossier

header companies

contains the owner companies of the dossier

header products

contains the product names covered by the dossier

header active substances

contains the active substances covered by the dossier

relation product - substance

contains the relation between products and active substances

Data dictionary

A data dictionary for all the index files is contained in the directory "INDEX". This data dictionary is constant for the version 1.0 of CADDY. It may be changed in future versions. Two index files are used to store the data dictionary:

table list

contains a list of all tables which can be contained in directory "INDEX"

column list

contains a list of all columns of the tables listed in the table list

Changes history

On supplemental volumes, an additional index file can be stored:

Changes history

contains a list of all pages that are deleted and a list of all reports that are changed or deleted

This additional file has to be present if pages were deleted or if reports were deleted or changed. It is useful if the differences to a previous version should be determined. The file can also be produced by using the page table and the report list of all old versions and the page table and the report list of the current version. Since these tables can be very big, it is easier for the retrieval software to use this file, than to compare the tables.

The rows of the changes history table will never be updated in new dossier versions. If necessary new rows for changes have to be added.

Page files

Unique page ID

Every page in a dossier gets a unique page ID called UPID. This page ID is a system key for internal use only. The UPID is to be generated by the compilation software. It can never be used again (for other pages in supplements), even if the original page was deleted. There should be no possibility for users to change the value of this key. The UPID should not be displayed by the retrieval software to avoid confusion with other data.

The retrieval of page files is done by using the page table. This table contains information about the storage location (volume) of the file. All page files are stored on the volumes in subdirectories of a directory with the name "PAGES". The parent directory of the directory "PAGES" has to be the root directory.

The basename of a page consists of the UPID padded with leading zeros to 8 characters. The file name of a page is formed using its basename and the extension ".TIF".

basename.TIF

The relative path name of the subdirectory, containing the file, is calculated using the basename. This path name is called subpath. To be sure to meet the special requirements of the media a subdirectory structure for the pages is given, which is built out of the first six characters (digits) of the basename.

The subpath is build as follows: Character one and two of the basename followed by a pathdelimiter, the characters three and four of the basename, a second delimiter and the characters five and six of the basename.

Example: subpath = "00/53/91" (if basename = "00539128" and pathdelimiter = "/")

The access to a page file can be done using one of the following possibilities:

- The path name in a DOS environment is built as follows: (pathdelimiter = "\\")
CD-ROM drive letter + ":\\" + "PAGES" + "\\\" + subpath + "\\\" + basename + ".TIF"
- The path name in a UNIX environment is built as follows: (pathdelimiter = "/")
CD-ROM mounting point + "/" + "PAGES" + "/" + subpath + "/" + basename + ".TIF"

Confidential pages and volumes

If there are confidential pages they have to be stored on one or more extra volumes.

These volumes are to be labeled clearly as confidential volumes (see subchapter "Label files"). The handling of confidential volumes in terms of file system structure and retrieving page information is the same as with "normal" volumes. If a volume contains one or more confidential pages, then the confidentiality flag for this volume has to be set. It is not allowed to store confidential and non confidential information on the same volume. The index information files are always stored on non confidential volumes.

Copies of page files

It is possible to store additional copies of already submitted page files on supplemental volumes. The purpose of this is to minimize the necessity of media changes during retrieval. If a single page in a report was replaced, it may be better to store the whole report on the supplemental volume. This technique can be used to copy all pages of a single volume to the new supplemental volume. In this case the user of a submission does not need to insert the old volume any more, when pages are retrieved. If all pages of a volume have been deleted or copied to a new volume, then the old volume is marked in the volume list as not containing current page files any more. Even if copies of page files are submitted, the references to the original page files are kept in the database for audit trail reasons.

Additional files not covered by CADDY (optional)

It is possible to have an additional directory on the volumes. This directory must have the name "FILES" and can be located in the root directory of any volume. Inside this directory any files and subdirectories can be stored, which are not covered by CADDY. This option is helpful for files, which are to be exchanged together with the dossier. Especially for the preparation of a Monograph it may be helpful to have additional wordprocessor files accessible together with the dossier. The content of the directory "FILES" is not standardized, which means, that there have to be a mutual agreement between the submitter of a volume and the receiver of the volume on the used file formats. A need for harmonization was recognized, but is not within the scope of this format specification.

Frequently asked questions

Is it possible to store confidential files on the current index volume ?

No

Is it possible to choose other file names as outlined above ?

No

Is it possible to have volumes with no pages stored on ?

Yes, the index volume is allowed to contain no page files.

Is there a version control on the files in the directory "FILES" ?

No

How many page files can be stored in one directory ?

Up to 100 page files can be stored in one directory. They are only stored in leaf directories of the directory tree PAGES.

Dossier database

The index information is organized as a relational database. The tables are stored in the basic index files.

The following tables contain the history of all versions of the dossier and the dossier attributes. With one exception in the table volume list, the rows of these tables will never be updated in new dossier versions. It is necessary to add new rows for every new version.

version history (version of the dossier)

This table contains a version history. To keep track of supplemental volumes and dossier versions the information stored in this table can be used.

volume list (volume)

This table stores the access information for the volumes. Every volume entry corresponds to a volume. Only the attribute "contains current page files", that informs the user if he needs this volume for retrieval, may be updated in a new dossier version.

header attribute set (dossier attributes)

This table contains a history of all dossier attributes that were attached to the versions of a dossier.

header companies (dossier owners)

This table contains a history of all company names that were attached to the attribute sets of the dossier.

header products (in Europe: annex III products)

This table contains a history of all product names that were attached to the attribute sets of the dossier.

header active substances (in Europe: annex II substance)

This table contains a history of all active substance names that were attached to the attribute sets of the dossier.

relation product - substance

This table contains the relation between products and active substances.

The following three tables reflect only the content of the current version. They are used for retrieval purposes.

report list (report attributes)

This table contains a list of all reports contained in the current version of the dossier. Every report entry has to have exactly one corresponding entry in the TOC table. This means the list of reports is a kind of additional attribute information to report TOC entries. If one physical report is used more than once in a dossier, then the report entry has to be copied as often as needed. The primary key of this table is the unique report ID (URID).

table of contents (TOC entry)

The TOC table represents a hierarchical table of contents (TOC tree). Every entry in this table represents a node in the TOC tree. One entry represents the root of the tree. All other entries are subnodes. The level of the root entry is defined as one. The levels of other TOC entries are defined to be the level of their parent entry incremented by 1. This table can never be empty. At least one entry for the root is required. The primary key of this table is the unique node ID (UNID).

page table (page)

This table stores the access information for the pages. Every entry corresponds to a page file on the volume set. The table is ordered by the order of the pages in the dossier. This page order is not explicitly stored on the volume. The primary key of this table is the unique page ID (UPID).

The following figure shows an entity relationship diagram for the dossier database.

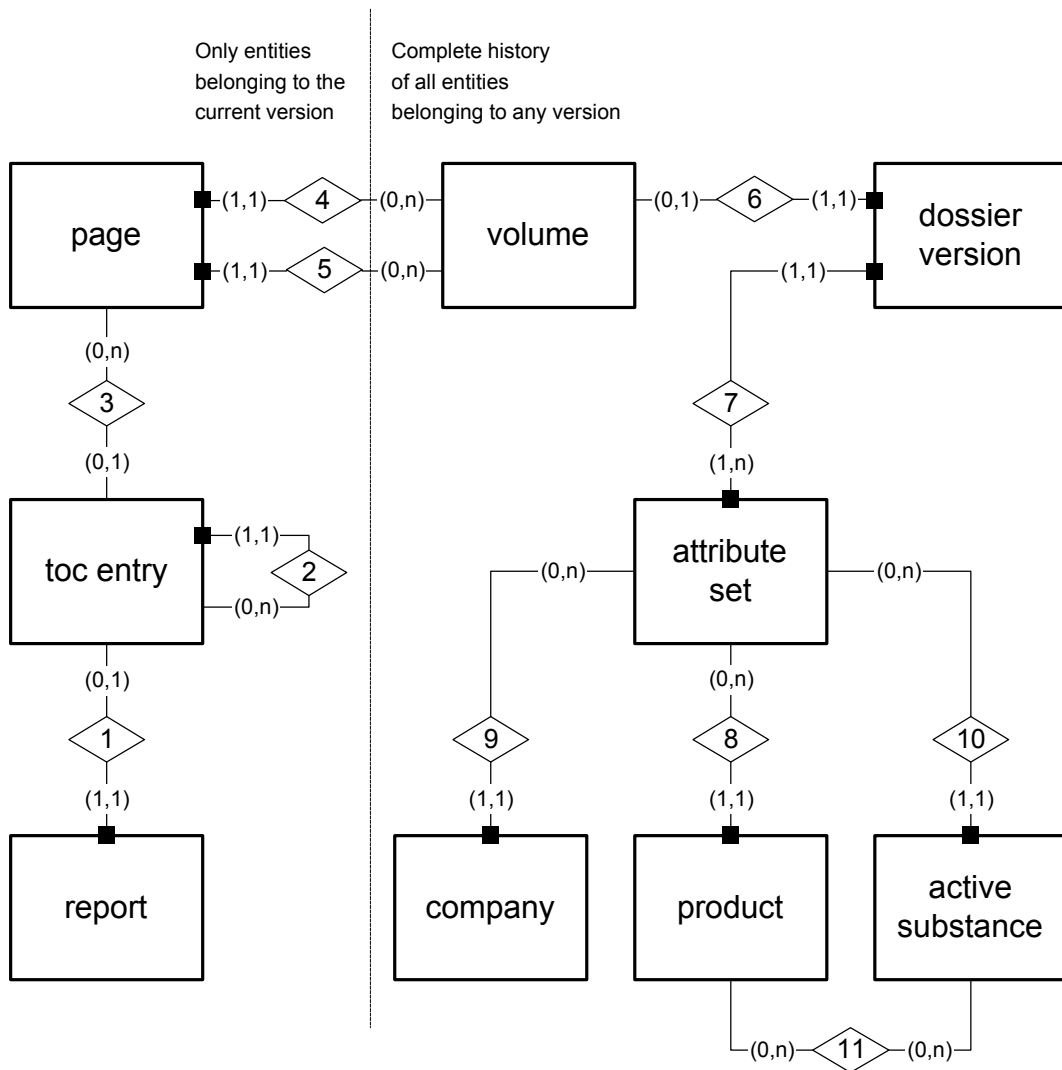


Figure 4 Entity Relationship Model

- (1) Every report has a corresponding TOC entry. (mandatory)
- (2) All nodes, with exception of the root node, will have exactly one parent. (mandatory)
The TOC is a hierarchical structure. Every TOC node can have n subnodes.
- (3) Every TOC entry can have a link to one page entry. (optional)

- (4) Every page entry has a link to the original volume (volume entry) where the page file was originally stored. (mandatory)
- (5) Every page entry has a link to the current volume (volume entry) where the page file is currently stored. (mandatory)
- (6) Every version entry has a link to a volume entry. The corresponding volume is called the current index volume. (mandatory)
- (7) Every version entry has attached header attributes. (mandatory)
- (8) A header attribute set can contain products. (optional)
- (9) A header attribute set can contain companies. (optional)
- (10) A header attribute set can contain active substances. (optional)
- (11) A product can contain active substances. (optional)
This n to m relation is implemented as a separate table.

The realization of the relationships is shown in the tables in Appendix A.

Database table attributes

A description of all the columns of the database tables is given in Appendix A.

Dossier header

The dossier header is a set of attributes belonging to the whole dossier. A history of dossier attribute sets is contained in the dossier database. Once submitted, the header attribute set will never be changed. It is only possible to submit a new version of the dossier with a new attribute set. The attribute set of the current version is called the current header attribute set.

In the table header attributes set each record describes a set of attributes that were valid for some versions of the dossier. Whenever it is necessary to change an attribute of the dossier it is necessary to add a new record with the new attribute set.

Each record has its own system key, an integer value ranging from 1 to 99. This key is organized as a counter and is 1 for the first set of attributes. The first version of a dossier has to contain exactly one record in this table.

Warning: The system key is for internal use only. It is to be generated by the compilation software. There should be no possibility for users to change the value of this key. This key should not be displayed by the retrieval software to avoid confusion with other data.

The following attributes of a dossier are stored in the table header attributes set:

- dossier title
- dossier sub title
- authority (e.g. EEC, EPA or PMRA)
- requirement / official guideline (in Europe: e.g. 1663/VI/94 Rev 7.4)
- regulation (in Europe: e.g. 3600/92)
- rapporteur member state (for Europe only)

and (stored in an extra table called header active substances)

- active substance / CIPAC code / CAS number
In most cases the extra table will contain only one row. Detailed information on the active substances can be found on the dossier pages. In this table each record describes an active substance that belongs to a header attribute set. Since the active substances are stored in an extra table, the number of active substances, belonging to a header attribute set is not limited.

and (stored in an extra table called header products)

- product name / formulation type
Detailed information on the products can be found on the dossier pages. In this table each record describes a product that belongs to a header attribute set. Since the product names are stored in an extra table, the number of products, belonging to a header attribute set is not limited.

and (stored in an extra table called header companies)

- company name / company country name / company code / company country code
In this table each record describes an dossier owner that belongs to a header attribute set. Since the company names are stored in an extra table, the number of companies, belonging to a header attribute set is not limited.

and (stored in an extra table called relation product - substance)

- relation between products and active substances / total concentration
In this table each record describes the total concentration of one active substance in one product.

Report attributes

The table report list contains a list of reports for the current dossier.

Each record has its own system key, an integer value ranging from 1 to 9999. This key, called URID, is a system key for internal use only. The URID is to be generated by the compilation software. It can never be used again (for other reports in supplements), even if the original report was deleted. There should be no possibility for users to change the value of this key. The URID should not be displayed by the retrieval software to avoid confusion with other data.

The following user attributes are stored in the report list:

- dossier file number (for EU) / MRID (for US) / (***) (for Canada)
- company file number
- date of document
- title of document
- authors of document
- source of document
- owners of report
- test facility
- is report confidential ?
- is report GLP conform ?
- was report published ?
- test on vertebrates ?
- data protection claimed ?

Table of contents

This table contains a table of contents for the current dossier version.

- abbreviated TOC entry string
- hierarchical number
 - guideline reference number (within US and Canada)
 - annex point number (within EU)

Page table

This table contains a list of all pages of the current dossier version.

- pagination string

Confidentiality of pages, reports and TOC nodes

If a report contains a page that is flagged confidential, the report itself has to be flagged confidential and all other pages of this report too.

In general: A page that is directly referenced by a TOC node is called a referenced page. The pages between two referenced pages, including the first referenced page (excluding the second referenced page), are called a page sequence. The TOC node referencing the first page is called the TOC node of the sequence.

If a page within a sequence is confidential, all pages of the sequence have to be confidential and the TOC node of the sequence is confidential too. If a TOC node is confidential all sub nodes have to be confidential and the referenced page (if one exists).

The confidentiality of TOC nodes is not stored explicitly because it can be easily determined by any software using the page confidential flag. This is also possible for the report confidential flag. Nevertheless this flag was stored to help end users to produce report lists with confidential flags.

Frequently asked questions

Is it possible to have TOC entries that are not reports ?

Yes

Which tables must have at least one entry ?

volume list / version history / table of contents / page table / header attributes set

Is it allowed to store empty index files in the directory „INDEX“ ?

No

Why is it necessary to duplicate reports which are used more than once in a CADDY dossier ?

In most of today's paper dossiers such reports are contained as copies. To represent the paper dossier as close as possible and to simplify the production of a CADDY dossier when scanning an existing dossier, this approach was chosen.

Is it possible to reference pages from TOC nodes that have sub nodes ?

Yes

Is it possible to have sub nodes of a TOC entry that is a report ?

Yes

Index file format

In the following chapter we describe the syntax of the data representation inside the index files.

The basic data types

There are four basic data types that are used in the index files:

- text
all character strings containing up to 256 graphic characters
and no leading or trailing blanks
- date
all valid dates between January 1, 1900 and December 31, 2099
- integer
all integers between zero and 99 999 999
- boolean
the boolean values true and false

In the following part a syntax description for the data types is given. The notation used for this description will be described in the next draft.

- represents the empty word (no character)
- [char] is a character of a 8-bit character set

The function `asc([char])` returns the byte value which represents the character.

The inverse function `chr(byte)` returns the character encoded by byte.

$$\text{chr}(\text{asc}([\text{char}])) = [\text{char}] \quad \text{and} \quad \text{asc}(\text{chr}(\text{byte})) = \text{byte}$$

The character set used is : ECMA-94 Latin-1 (ISO 8859-1)

This character set consists of 191 graphic characters. Three of them have special meanings:

SPACE	[space] ::= chr(32)
NO-BREAK SPACE	[NBSP] ::= chr(160)
SOFT HYPHEN	[SHY] ::= chr(173)

For CADDY 1.0 all graphic characters of the ECMA 94 Latin 1 character set except of [NBSP] and [SHY] are allowed. This subset is called [pchar].

[pchar] ::= char(x) | char(y) | char(z) & 32 ≤ x ≤ 126 & 161 ≤ y ≤ 172 & 174 ≤ z ≤ 255
allowed graphic characters

The data type [boolean]

[boolean] ::= [true] | [false]

The data type [integer]

[digit/0] ::= "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"
 remark: all single digits from 1 to 9 without zero

[digit] ::= "0" | [digit/0]
 remark: all single digits including zero

[integer/0] ::= [digit/0] | [integer/0] [digit]
 remark: all positive integers without zero (no leading zeros possible)

[integer] ::= "0" | [integer/0] & $0 \leq \text{val}([\text{integer}]) \leq 99\,999\,999$
 remark: all positive integers including zero (no leading zeros allowed)

The function `val([integer])` returns the numerical value of the item.

The data type [date]

[day] ::= [integer] & $1 \leq \text{val}([\text{day}]) \leq 31$

[month] ::= [integer] & $1 \leq \text{val}([\text{month}]) \leq 12$

[year] ::= [integer] & $1900 \leq \text{val}([\text{year}]) \leq 2099$

[date] ::= [month] "/" [day] "/" [year] & [date] represents a valid date
 remark: all valid dates between January 1, 1900 and December 31, 2099

The data type [text]

[text] ::= [text] [pchar] | □ & leading and trailing blanks are not allowed
 & $0 \leq \text{len}([\text{text}]) \leq 256$

The function `len([text])` returns the number of characters in [text]

Storing the tables to the index files

Before storing a text item to an index file the item has to be encoded to be sure that the delimiter characters contained in the text item cannot cause misleading.

The following replacements are made:

- every escape character "\" is replaced by "\\\""
- every delimiter character "|" is replaced by "\\|\""

The function `enc([text])` is the function which converts a [text] item to a [dbtext] item.

The inverse function is called `dec([dbtext])`. This functions inverts the replacements performed by `enc`.

`dec(enc([text])) = [text]` and `enc(dec([dbtext])) = [dbtext]`

General index file format

The general index file format is as follows:

[file] ::= [file] [line] | □

[line] ::= [field] [delimiter] [line] | [field] [eol]

[field] ::= [boolean] | [integer] | [date] | [dbtext]

There is a couple of standard database software available on UNIX and MS-DOS that can read directly the specified format without writing any application specific code.

The field order and the field types for the basic index files are given in Appendix B.

Frequently asked questions

Does the general index file format allow for empty files ?

Yes

Image file format

The Tag Image File Format (TIFF) is an accepted industry standard file format that allows to describe the storage format for raster images. TIFF was designed to promote the interchange of digital image data. It is intended to be independent of specific operating systems, filing systems, compilers and processors. On the market exists a wide variety of software supporting TIFF. Because of the flexibility and richness of TIFF, there is no "standard" TIFF file. Different applications can support different tag fields. To avoid problems that can arise out of the TIFF variety, we will clearly define a subset of TIFF and give a description of the used tag fields in this chapter.

The image files are stored as TIFF Revision 6.0 / baseline bilevel images.

All the following descriptions are to be seen in context to the TIFF 6.0 specification.

The selection of tag fields was made regarding the following two objectives:

- Existing image viewers on the market should be able to read the image files.
- In-house systems should be able to produce these files as simple as possible.

The special CADDY requirements are:

- ⇒ only bilevel images
(SamplesPerPixel=1, BitsPerSample=1, PhotometricInterpretation)
- ⇒ only one image per file
(NewSubfileType=0)
- ⇒ only one strip per file
(StripByteCounts, StripOffsets, RowsPerStrip)
- ⇒ fixed resolution (300 dpi)
(ResolutionUnit=2, Xresolution=300, Yresolution=300)
- ⇒ size limited to A4 and US Letter
(ImageLength, ImageWidth)

TIFF 6.0 baseline bilevel images

Only fields that are described in the chapters "Baseline Field Reference Guide" and "CCITT Bilevel encoding" of the TIFF 6.0 specification are allowed inside CADDY image files. The following fields are used within CADDY image files. Other fields, not listed here, are not recommended.

Field	Type	Default	Valid Values	Remarks
SamplesPerPixel	S	1	1	only bilevel images
BitsPerSample	S	1	1	only bilevel images
PhotometricInterpretation	S	0	0, 1	only bilevel images
StripByteCounts	L or S	none	im	mandatory
StripOffsets	L or S	none	im	mandatory
RowsPerStrip	L or S	maxL	im	mandatory
NewSubfileType	L	0	0	only one image per TIFF file
Compression	S	1	1, 2, 3, 4	see "Compression schemes" below
T4Options	L	0	0,1	mandatory if compression = 3
T6Options	L	0	0	mandatory if compression = 4
ResolutionUnit	S	2	2	resolution in inch
Xresolution, Yresolution	R	none	300	only 300 dpi images
ImageLength, ImageWidth	L or S	none	im	A4 or Letter (portrait or landscape)
Orientation	S	1	1	
Thresholding	S	1	1	
FillOrder	S	1	1	

Type: S = Short
 L = Long
 R = Rational

Default: none no default, field mandatory
 maxL maximal long value = $2^{32} - 1$

Valid values: im image dependent

The following fields described in chapter "Baseline Field Reference Guide" of the TIFF 6.0 specification must not be used inside CADDY image files:

CellLength, CellWidth, ColorMap, ExtraSamples, FreeByteCounts,
FreeOffsets, GrayResponseCurve, GrayResponseUnit, MaxSampleValue,
MinSampleValue, PlanarConfiguration, SubfileType

Informational fields

Some image viewers on the market are able to display informational fields to the user. Others simply ignore these fields. Therefore it is not recommended to use the following informational fields inside CADDY image files:

Artist, Copyright, DateTime, HostComputer, ImageDescription, Make,
Model, Software

If there are informational fields inside your CADDY image files, a CADDY viewer should ignore these fields.

Compression schemes

The compression types used, are limited to those which are used in common in-house systems and which are supported by a wide variety of image viewers:

- Uncompressed
- TIFF Type 2 (variant of CCITT Group 3)
- CCITT Group 3 1D
- CCITT Group 3 2D
- CCITT Group 4

The following values are valid for the compression field:

- 1 : No compression
- 2 : CCITT Group 3 1-dimensional modified huffman run length encoding
- 3 : Facsimile compatible CCITT Group 3
T4Options = 0 (1-dimensional coding)
T4Options = 1 (2-dimensional coding)
- 4 : Facsimile compatible CCITT Group 4
T6Options = 0

Paper formats

The paper formats supported are limited to avoid handling problems (printing, display) on the user's side. The limitation to the common formats enables the user to reproduce the pages on commonly used printers and screens.

The paper formats supported are:

- A4 210 x 297 mm
- Letter 8,5 x 11 inch

Assuming a resolution of 300 dpi the following values (given in pixel) for the height and width of the images are possible:

Portrait	Landscape	A4	Letter
Height	Width	3508	3300
Width	Height	2480	2550

The values given in the table are the maximum values for the fields ImageLength and ImageWidth.

Conformity Testing

There are three areas in which the conformity to the CADDY specification should be shown:

- 1) media conformance
- 2) index file conformance
- 3) image file conformance

The conformance of the media to the requirements described in chapter "Media" should be claimed by the producer of the CD-ROM. In most cases he will rely on the conformity of the software he has used for the production of the CD-ROM.

The conformance of the image files to the TIFF 6.0 specification has also to be claimed by the submitter of a CADDY dossier. The conformity test system of CADDY should check only the validity of the TIFF-Tags which are described in this specification.

To simplify the whole checking process and to avoid checking entities more than once, the conformity test system will operate based on dossier versions. This means that for a conformity check of version x, the system can rely on the CADDY conformity of version x- 1.

The conformity test system should produce a report with the results of the various tests. This report has to contain on every page the

- test facility name
- test person
- date and time of the test
- page number
- number of in report
- tested version
- dossier name

As a minimal test the conformity test system should carry out the following steps:

1. Check location, name and availability of index files.
2. Check general index file format.
3. Check content of all index files and the conformance to the data dictionary (min. value, max. value, data type).
4. Check dependencies between index files.
5. Check availability, label files and directory structure of all volumes in the volume table.
6. Check availability of all pages in the page table (original files and file copies).
7. Check history by comparing the index files of the previous version with the current version.

The following tests are optional. They can be performed for all new images and copies of images belonging to the current version or on a random sample of images.

- Check TIFF-Flags in page files for CADDY conformity.
- Check readability and quality of page files.

References

- [1] Directive 91/4141/EEC (Article 5.3 and 8.2), Document 1663/VI/94 Rev 5, (August 1994)
- [2] Consultancy report: Open information interchange study on image/graphics standards. Final report. Prepared for Commission of the European Communities by Pira International, 41 TCON 1730, (June 1993)
- [3] ECMA-119 Volume and File Structure of CDROM for Information Interchange, 2nd Edition ECMA, Geneva, (December 1993)
- [4] ECMA-168 Volume and File Structure for Read-Only and Write-Once Compact Disc, ECMA, (September 1993)
- [5] ISO 9660: Information Processing - Volume and file structure of CD-ROM for information interchange, ISO, (1988)
- [6] DAMOS: Core module: Structure Specification for Dossier Exchange. Version 2.01, DAMOS Working Group, (May 1994)
- [7] ECMA-94 8-Bit Single-Byte Coded Graphic Character Sets - Latin Alphabets No. 1 to No. 4, 2nd Edition, ECMA, Geneva, (June 1986)
- [8] TIFF 6.0 specification. Final, Aldus Corporation, (June 1992)
- [9] Conformity testing and certification for information processing systems, DIN, 1. Aufl. Berlin; Köln: Beuth, 1985
- [10] The meaning of conformance to standards. Technical Report TR/18, ECMA, Geneva, (September 1983)
- [11] Compact Disc Digital Audio System Description ("Red Book"), N.V. Phillips, Sony Corp., (May 1985) (corresponds to CEI / IEC 908)
- [12] CD-ROM: Compact Disc Digital Read Only Memory, System Description ("Yellow Book"), N.V. Phillips, Sony Corp., (May 1985) (corresponds to ISO / IEC 10149)
- [13] Recordable Compact Disc Systems ("Orange Book"), N.V. Phillips, Sony Corp., (November 1990)
- [14] ISO 3166: Codes for the representation of names of countries, Fourth edition, ISO, (December 1993)
- [15] Document Interchange Format for PPP registration applications
Format Specification CADDY 1.0 / Version 0.7 / January 31, 1997
Joint EU Member States / ECPA Data Transfer Steering Group, (January 1996)

Appendix A

Structure definition of CADDY 1.1 index files

The structure definition given on the following pages describes the content of the index files. Columns in tables correspond to fields in the index files.

The following index files / tables are described:

CADDY-Table	File/Table Name
version history	dh_ver
volume list	dh_vol
report list	dc_rep
table of contents	dc_toc
page table	dc_pag
changes history	da_chg
header attribute set	dh_has
header active substances	dh_asu
relation product-substance	dh_rps
header products	dh_pro
header companies	dh_com
dd tables	dd_tab
dd columns	dd_col

Additional remarks on fields

field	value	remark
ver_stdv	2	CADDY 1.1 conform
ver_max_UPID		every dossier has at least one page
ver_max_UNID		every dossier has at least one toc entry
vol_name		same string as in label.txt file
vol_confidential	TRUE	This volume contains only pages flagged as confidential (p_confidential = TRUE)
	FALSE	This volume contains only non-confidential pages (p_confidential = FALSE)
vol_necessary	FALSE	This volume is not the current index volume and does not contain any of the pages referenced by p_curvol UVOL in the current page table

r_date, r_valid_day, r_valid_month		<p>If r_valid_month and rep_valid_day is TRUE, then r_date contains the effective date of the report.</p> <p>If r_valid_day = FALSE and r_valid_month = TRUE, then the day of the month is unknown and r_date should only be displayed by the software without a day. For sorting and consistency reasons the r_date field will contain the first day of the month.</p> <p>If r_valid_day and r_valid_month = FALSE, then the month of the year is unknown and r_date should only be displayed by the software as a year. For sorting and consistency reasons the r_date field will contain the first january of the year.</p> <p>The combination r_valid_day = TRUE and r_valid_month = FALSE is forbidden.</p>
r_confidential	TRUE	All pages of the report are to be marked as confidential
r_added	x	This value can never be changed in new versions. If x > 1 then the retrieval software should display a hint for the user.
r_changed	0	The entry of the report was not changed since first added. (==> when a report entry is new then r_changed = 0)
UNID n_parent_UNID n_level	1 0 1	Root node => UNID = 1 and n_parent_UNID = 0 and n_level = 1
n_report_URID	0	There are no report attributes attached (NIL-Pointer)
n_page_UPID	0 > 0	No page is attached (NIL-Pointer) The order of toc entries and pages has to be the same. That means a pointer is not allowed to point to a page which is located before the page where the previous pages pointer points at.
p_curvol_UVOL		If no copy exists, then p_curvol_UVOL = p_orgvol_UVOL
p_added	x	This value can never be changed in new versions. If x > 1 then the retrieval software should display a hint for the user.
tab_rowcount	0	The table contains no rows and the index file is not stored on the volume.
tab_version	2	The value of this field is constantly 2 for all tables in CADDY 1.1 format. In future CADDY versions this value will be incremented when the structure of a table is changed.

Index structure changes from CADDY 1.0 to CADDY 1.1

Renamed fields

Most columns were renamed to make the Data Dictionary better readable. Only the names were changed, not the data types, the column order or the content. Therefore these name changes have no effect on the index file contents.

table	newname	oldname
dh_ver	ver_master_date	h_date
dh_ver	ver_idxvol_UVOL	h_vol_ptr
dh_ver	ver_stdv	h_stdv
dh_ver	ver_attrib_UHID	h_dhead_ptr
dh_ver	ver_issue_date	h_issue_date
dh_ver	ver_max_UPID	h_max_UPID
dh_ver	ver_max_UNID	h_max_UNID
dh_ver	ver_max_URID	h_max_URID
dh_ver	ver_chg_rep	h_chg_rep_list
dh_ver	ver_chg_has	h_chg_dhead
dh_ver	ver_chg_toc	h_chg_toc
dh_ver	ver_chg_pag	h_chg_pag
dh_vol	vol_name	v_vol_name
dh_vol	vol_pages	v_pages
dh_vol	vol_index	v_index
dh_vol	vol_confidential	v_confidential
dh_vol	vol_necessary	v_necessary
dc_rep	r_tocnode_UNID	rep_toc_ptr
dc_rep	r_doss_fn	rep_doss_fn
dc_rep	r_comp_fn	rep_comp_fn
dc_rep	r_date	rep_date
dc_rep	r_valid_day	rep_valid_day
dc_rep	r_valid_month	rep_valid_month
dc_rep	r_title	rep_title
dc_rep	r_authors	rep_authors
dc_rep	r_source	rep_source
dc_rep	r_owners	rep_owners
dc_rep	r_test_facility	rep_test_fac
dc_rep	r_confidential	rep_confidential
dc_rep	r_glp	rep_glp
dc_rep	r_published	rep_published
dc_rep	r_vertibrates	rep_vertibrates
dc_rep	r_protect	rep_protect
dc_rep	r_added	rep_added

dc_rep	r_changed	rep_changed
dc_toc	n_report_URID	n_rep_ptr
dc_toc	n_parent_UNID	n_father_ptr
dc_toc	n_page_UPID	n_page_ptr
dc_pag	p_curvol_UVOL	p_vol_current
dc_pag	p_orgvol_UVOL	p_vol_original
da_chg	chg_type	ch_type
da_chg	chg_object_ID	ch_ID
da_chg	chg_ver_UVER	ch_ver
dh_has	UHID	ASID
dh_has	has_title	dh_title
dh_has	has_sub_title	dh_sub_title
dh_has	has_authority	dh_authority
dh_has	has_guideline	dh_guideline
dh_has	has_regulation	dh_regulation
dh_has	has_rapporteur	dh_rapporteur
dh_com	com_UHID	dc_ASID
dh_com	com_name	dc_company
dh_com	com_code	dc_comp_code
dh_com	com_country	dc_country
dh_com	com_country_code	dc_country_code
dh_pro	pro_UHID	dp_ASID
dh_pro	pro_name	dp_product
dh_pro	pro_formulation	dp_formulation
dd_tab	tab_name	t_tab_name
dd_tab	tab_version	t_tab_version
dd_tab	tab_rowcount	t_rowcount
dd_col	col_tab_name	c_tab_name
dd_col	col_name	c_col_name
dd_col	col_type	c_col_type
dd_col	col_value_min	c_value_min
dd_col	col_value_max	c_value_max

New tables

- header active substances dh_asu
- relation product-substance dh_rps

They are necessary to allow for more than one active substance to be covered by one dossier. See Appendix A for a description of these tables.

Removed fields

- dc_toc: remove field n_subnodes
turned out to be not necessary
- dh_vol: remove field v_supplement
turned out to be not necessary
- dh_has: remove field substance
the information is now stored in table dh_asu
- dh_pro: remove dp_concentration
the information is now stored in table dh_rap

Modified and added fields

- dh_pro: add integer field UPRO (unique product ID)
this text field was added to identify products system internal
- dh_com: add integer field UCOM (unique company ID)
this text field was added to identify dossier owners system internal
- dc_toc: add text field n_h_number (hierarchical number)
this text field was added to allow more precise description of TOC nodes
- dc_toc: modify field n_text (max. length was set to 70)
the length of this text field was adjusted to current needs
- dh_com: modify field dc_comp_code (max. length was set to 6)
the length of this text field was adjusted to allow storing of US company IDs
- dh_has: modify field dh_guideline (max. length was set to 20)
the length of this text field was adjusted to allow storing of the full European guideline name ("1663/VI/94 Rev 7.4")

Frequently asked questions

Which file name extension should be used for index files ?

For the index files there is no file name extension.

Where is the order of rows stored ?

This information is implicitly contained in the index files. The order of rows has to be the same as the order of lines in the index file. Each CADDY application should define an explicit order field that will be filled during import of the index files with the line number.

Index structure of CADDY 1.1

Legend

The column type contains the field type, one of the types defined in the chapter "Index file format".

The columns min and max are depending on the field type:

- If the type is integer, then min and max are the minimum and maximum values which are allowed for the field value.
- If the type is text, then the min value is the minimum number of characters in the text and the max value gives the maximum number of characters. In this case a min value greater than 0 means that a field value is mandatory.

Remarks

[unique]	The value has to be unique throughout the table
[index]	It is recommended to use this field in the database as index.
[not visible]	The retrieval software should not display the value of this field to the reviewer
[line number]	This field is not stored in the index file. Its value is the line number of the row inside the index file.
[not used]	This field contains no valuable information. Some [line number] fields are of this type.

dh_ver Version History

Column name	Type	Min	Max	Remark
ver_order	integer	1	999	order of rows on CDROM = UVER [not used] [unique] [line number] [not visible]
UVER	integer	1	999	unique version number [unique] [index]
ver_master_date	date			day of mastering the electronic CADDY version of the dossier (CADDY compilation date) [unique]
ver_idxvol_UVOL	integer	1	999	pointer to current index volume (UVOL) [unique] [not visible]
ver_stdv	integer	1	99	version number of standard [not visible]
ver_attrib_UHID	integer	1	99	pointer to header attribute set (UHID) [not visible]
ver_issue_date	date			issue date of the original dossier (submission date)
ver_max_UPID	integer	1	99999999	maximum UPID used in entire dossier (all versions) [not visible]
ver_max_UNID	integer	1	99999999	maximum UNID used in entire dossier (all versions) [not visible]
ver_max_URID	integer	0	9999	maximum URID used in entire dossier (all versions) [not visible]
ver_chg_rep	boolean			report list changed ?
ver_chg_has	boolean			header attribute set changed ?
ver_chg_toc	boolean			table of contents changed ?
ver_chg_pag	boolean			page table changed ?

dh_vol Volume List

Column name	Type	Min	Max	Remark
vol_order	integer	1	999	order of rows on CDROM = UVOL [not used] [unique] [line number] [not visible]
UVOL	integer	1	999	unique volume number [unique] [index]
vol_name	text	1	12	unique volume name [unique] [not visible]
vol_pages	integer	0	99999999	number of pages on this volume
vol_index	boolean			contains index files ?
vol_confidential	boolean			is confidential volume ?
vol_necessary	boolean			contains current page files ?

dc_rep		Report List		
Column name	Type	Min	Max	Remark
rep_order	integer	1	9999	order of rows on CDROM = order of reports in dossier [unique] [line number] [index] [not visible]
URID	integer	1	9999	unique report ID [unique] [index] [not visible]
r_tocnode_UNID	integer	1	99999999	pointer to toc entry (UNID) [unique] [not visible]
r_doss_fn	text	0	20	dossier file number (for EU) / MRID (for US) / (***) (for Canada)
r_comp_fn	text	0	20	company file number
r_date	date			date of document [index]
r_valid_day	boolean			is day of report date applicable ? [not visible]
r_valid_month	boolean			is month of report date applicable ? [not visible]
r_title	text	0	240	title of document
r_authors	text	0	120	authors of document [index]
r_source	text	0	180	source of document
r_owners	text	0	120	owners of report at submission date
r_test_facility	text	0	60	test facility
r_confidential	boolean			is report confidential ? (Documents J)
r_glp	boolean			is report GLP conform ?
r_published	boolean			was report published ?
r_vertebrates	boolean			test on vertebrates ?
r_protect	boolean			data protection claimed ?
r_added	integer	1	999	pointer to version (UVER) in which report was added first [not visible]
r_changed	integer	0	999	pointer to version (UVER) in which the report attributes where changed last [not visible]

dc_toc **Table of Contents**

Column name	Type	Min	Max	Remark
toc_order	integer	1	99999999	order of rows on CDROM = order of toc nodes [unique] [line number] [index] [not visible]
UNID	integer	1	99999999	unique node ID [unique] [index] [not visible]
n_report_URID	integer	0	9999	pointer to report attributes (URID) [unique] [index] [not visible]
n_parent_UNID	integer	1	99999999	pointer to parent node (UNID) [not visible]
n_level	integer	1	99	level in toc hierarchy [not visible]
n_page_UPID	integer	0	99999999	pointer to first page (UPID) [not visible]
n_h_number	text	1	20	hierarchical number (guideline reference number / MRID within US and Canada and annex point number / dossier file number within EU) [index]
n_text	text	1	70	abbreviated toc entry string [index]

dc_pag **Page Table**

Column name	Type	Min	Max	Remark
pag_order	integer	1	99999999	order of rows on CDROM = order of pages in dossier [unique] [line number] [index] [not visible]
UPID	integer	1	99999999	unique page ID [unique] [index] [not visible]
p_curvol_UVOL	integer	1	999	current storage location (UVOL) [not visible]
p_orgvol_UVOL	integer	1	999	original storage location (UVOL) [not visible]
p_confidential	boolean			is page confidential ? (Documents J)
p_pagination	text	1	20	pagination string which is shown (printed) on the page [index]
p_added	integer	1	999	pointer to version (UVER) in which page was added [not visible]

da_chg **Changes History**

Column name	Type	Min	Max	Remark
chg_order	integer	1	99999999	order of rows on CDROM = [not used] [unique] [line number] [not visible]
chg_type	text	1	1	P (Page deleted) / R (Report entry deleted) / A (Attributes of report changed) [not visible]
chg_object_ID	integer	1	99999999	UPID or URID of deleted/changed item [not visible]
chg_ver_UVER	integer	1	999	last version (UVER) in which record of the item can be found [not visible]

dh_has Header Attribute Set

Column name	Type	Min	Max	Remark
has_order	integer	1	99	order of rows on CDROM = UHID [not used] [unique] [line number] [not visible]
UHID	integer	1	99	unique header ID (key) [unique] [index] [not visible]
has_title	text	1	60	dossier title
has_sub_title	text	0	60	dossier sub title
has_authority	text	1	10	authority (e.g. EEC)
has_guideline	text	1	20	requirement / official guideline
has_regulation	text	1	10	regulation (e.g. 3600/92)
has_rapporteur	text	2	2	European rapporteur member state (2-letter ISO code) or code of US or Canada

dh_com Header Companies

Column name	Type	Min	Max	Remark
com_order	integer	1	99	order of rows on CDROM = order of companies [unique] [line number] [index] [not visible]
UCOM	integer	1	99	unique company ID [not used] [unique] [index] [not visible]
com_UHID	integer	1	99	header ID (foreign key of dh_has) [index] [not visible]
com_name	text	1	30	company name
com_code	text	3	6	3-digit EU company code or 6-letter US company ID
com_country	text	1	30	country name of the companies location
com_country_code	text	2	2	2-letter ISO code of the country

dh_asu Header Active Substances

Column name	Type	Min	Max	Remark
asu_order	integer	1	99	order of rows on CDROM = order of active substances [unique] [line number] [index] [not visible]
UASU	integer	1	99	unique active substance ID [unique] [index] [not visible]
asu_UHID	integer	1	99	header ID (foreign key of dh_has) [index] [not visible]
asu_substance	text	1	60	active substance
asu_cipac	text	0	4	CIPAC number
asu_cas	text	0	11	CAS code
asu_annex	boolean			TRUE, if active substance is covered by European annex II dossier, else FALSE

dh_pro Header Products

Column name	Type	Min	Max	Remark
pro_order	integer	1	99	order of rows on CDROM = order of products [unique] [line number] [index] [not visible]
UPRO	integer	1	99	unique product ID [unique] [index] [not visible]
pro_UHID	integer	1	99	header ID (foreign key of dh_has) [index] [not visible]
pro_name	text	1	60	product name or product code
pro_formulation	text	1	3	formulation type according to GIFAP rules
pro_annex	boolean			TRUE, if product is covered by European annex III dossier, else FALSE [not visible]

dh_rps Relation Product - Substance

Column name	Type	Min	Max	Remark
rps_order	integer	1	99	order of rows on CDROM = [not used] [unique] [line number] [not visible]
rps_UPRO	integer	1	99	product ID (foreign key of dh_pro) [index] [not visible]
rps_UASU	integer	1	99	active substance ID (foreign key of dh_asu) [index] [not visible]
rps_concentration	text	1	10	total concentration (with units: e.g. 15 ml/l)

dd_tab DD Tables

Column name	Type	Min	Max	Remark
tab_order	integer	1	99	order of rows on CDROM = [not used] [unique] [line number] [not visible]
tab_name	text	1	8	table/file name (key) [unique] [index] [not visible]
tab_version	integer	1	99	version of table [not visible]
tab_rowcount	integer	1	99999999	number of rows in table/file [not visible]

dd_col DD Columns

Column name	Type	Min	Max	Remark
col_order	integer	1	9999	order of rows on CDROM = order of columns [unique] [line number] [index] [not visible]
col_tab_name	text	1	8	table/file name (foreign key to dd_tab) [not visible]
col_name	text	1	20	name of column [not visible]
col_type	text	1	7	type of column (boolean/integer/date/text) [not visible]
col_value_min	integer	0	99999999	text: min. length / integer: min. value [not visible]
col_value_max	integer	1	99999999	text: max. length / integer: max. value [not visible]

Addendum to CADDY 1.1 Format Specification/ September 23, 1997

Draft February 24, 1998



= additional definition



= editorial correction







Info = explanation / clarification






= content correction

Item	Type	Chapter	Pages	Remark
1	Info			<p>Explanation: standard version vs. dossier version</p> <p>There are two different "versions" referred to in the format specification: The CADDY standard version and the CADDY dossier version.</p> <p>The CADDY standard version is the version number of the CADDY format specification. For the current format specification the version number is 1.1.</p> <p>The CADDY dossier version is the version number assigned to a version of a CADDY dossier. This version number starts with 1 for the first version and is increased with each supplement.</p>
2	Info			<p>Explanation: CADDY compilation date vs. issue date</p> <p>There are two different "version dates" referred to in the format specification: The CADDY compilation date and the dossier version issue date. Both dates will normally be changed from dossier version to dossier version. For additional information about these dates see "CADDY Application Guide".</p>
3		Volume sets	11	<p>Volume Set ID: [notifying companyshortcode]:</p> <p>Only capital letters are allowed.</p>
4		Volume sets	11	<p>Volume Set ID: [countryshortcode]:</p> <p>Only capital letters are allowed.</p>
5		Volume sets	12	<p>Label Files: [version string]:</p> <p>Replace text "CADDY version number followed by the version date in English format" by "CADDY dossier version number followed by the CADDY compilation date in English format".</p> <p>Add text "(see also fields UVER and ver_master_date in table dh_ver)".</p>
6		Index files	14	<p>Second bullet: Data Dictionary</p> <p>Replace text "CADDY versions" by "CADDY standard versions".</p>

Addendum to CADDY 1.1 Format Specification/ September 23, 1997

Item	Type	Chapter	Pages	Remark
7		Index files	14	Data dictionary: Replace text "version 1.0 of CADDY. It may be changed in future versions." by "standard version 1.1 of CADDY. It may be changed in future CADDY standard version\$.
8	<input checked="" type="checkbox"/>	Index files	14	Data dictionary: The column tab_rowcount within table dd_tab is not constant for CADDY 1.1 but contains the values appropriate for the current CADDY dossier version.
9		Page files	15	The directory PAGES and all of its subdirectories are only allowed to contain page files or subdirectories. They must not be empty.
10	Info	Frequently asked questions	16	Must each volume contain a PAGES directory ? No. If a volume contains no page files the directory PAGES must not exist. This is allowed for index volumes only.
11	Info	Frequently asked questions	16	How many lines of text are contained within the label file ? Each label file has to contain 8 lines of text.
12	Info	TIFF 6.0 baseline bilevel images	27	XResolution and YResolution are mandatory fields.
13	Info	TIFF 6.0 baseline bilevel images	27	ImageLength and ImageWidth are mandatory fields.
14		TIFF 6.0 baseline bilevel images	29	Paper formats: Delete text: "The values given in the table are ..". The values of height and width may be 1 pixel more than stated in the table. (This rule should allow for more flexibility to ignore unprecise calculations of scanning software) A maximum deviation of 1% less than the values in the table is allowed. (This rule should avoid to submit much smaller images than A4 or Letter.)
15		Conformity Testing	30	The 5th bullet: "number of in report" has to be deleted.
16		Database table attributes	20	Dossier header: The header ID of the referenced product and the referenced active substance must be the same.
		Index structure of CADDY 1.1	42	Table dh_rps; column rps_UPRO Table dh_rps; column rps_UASU
17		Index structure of CADDY 1.1	37 42	Legend: For the field types boolean and date the values of min and max are not used. Within the table dd_col the column col_value_min has to contain the value 0 and the column col_value_max has to contain the value 1 for both field types. Table dd_col, column col_value_max Table dd_col, column col_value_min

**Addendum to
CADDY 1.1 Format Specification/ September 23, 1997**

Item	Type	Chapter	Pages	Remark
18		Index structure of CADDY 1.1	38	Table dh_ver; column ver_master_date: Add text "(see also [version string] in file LABEL.TXT)".
19		Index structure of CADDY 1.1	38	Table dh_ver; column ver_issuse_date: Delete text "(submission date)".
20	<input checked="" type="checkbox"/>	Index structure of CADDY 1.1	38	Table dh_vol; column vol_name: Add text "The volume name is described in chapter Volume sets (see also volume names on page 11)".
21	<input checked="" type="checkbox"/>	Index structure of CADDY 1.1	40	Table dc_toc; column n_report_URID: Delet text "[unique]". Explanation: This column is allowed to contain 0 values. Because multiple rows with 0 values are allowed the column is not unique.
22	<input checked="" type="checkbox"/>	Index structure of CADDY 1.1	40	Table dc_toc; column n_parent_UNID: The min value has to be changed to Min = 0. Explanation: For the root node the value of n_parent_UNID is defined to be 0.
23	<input checked="" type="checkbox"/>	Index structure of CADDY 1.1	42	Table dd_tab; column tab_rowcount: The min value has to be changed to Min = 0. Explanation: Tables that are not mandatory can be empty.
24		Index structure of CADDY 1.1	42	Table dd_tab; column tab_version: Replace text "version of table" by "CADDY standard version of table"