



Joint CADDY Steering Group

CADDY 2.0

Format Specification

A Document Interchange Format For Pesticides Registration Applications

FINAL DRAFT
June 19, 2002



Copyright

© 1995 - 2002 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.

Trademarks are only used to identify products. They may be trademarks or registered trademarks of the respective companies. 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
HP Consulting
Hewlett-Packard-Straße
D-61352 Bad Homburg
Germany
email: michael_gross@hp.com

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

Wolf-Martin Maier

Directorate General Health and Consumer
Protection
Sector E1 Plant Health
B-1049 Brussels
Belgium
email: wolf-martin.maier@cec.eu.int

Hans Mattaar

Aventis CropScience
P.O. Box 147
4906 AB Oosterhout
The Netherlands
email: hans.mattaar@aventis.com



Table of contents

Introduction	6
About this format specification	6
CADDY version 1.0	6
CADDY version 1.1	6
CADDY version 2.0	6
Future CADDY versions	6
Electronic submissions of plant protection products	7
What is a CADDY dossier?	7
Standardization objectives	8
Application of the standard for pesticides registrations	9
Explanation of document specific terms	10
CADDY format version vs. CADDY dossier version	10
CADDY compilation date vs. issue date	10
Media	11
Information Filing	12
Volume sets	13
Index volumes	13
Confidential volumes	13
Distribution of page files to the volumes (recommendation)	14
Distribution of controlled files to the volumes (recommendation)	14
Volume set ID	14
Volume names	14
Label files	15
Creating new dossier versions by adding supplements	16
Index files	17
Basic index files	17
Changes history	17
Data dictionary	18
Page files	18
Unique page ID	18
Storage Location	18
Controlled files	19
Unique file ID	19
Storage Location	19
Copies of page files and controlled files	19
Unsupported additional files (optional)	20
Frequently asked questions	20



Dossier database	21
Entity Relationship Model	21
Entities	22
Relationships.....	22
Database table descriptions	23
Tables covering all dossier versions	23
Tables covering current dossier version	25
Confidentiality of pages, controlled files, reports and TOC nodes.....	27
Tracking of changes between dossier versions.....	27
Frequently asked questions	27
Index File Format	28
The basic data types	28
The data type [boolean]	30
The data type [integer]	30
The data type [date]	30
The data type [text]	30
Storing the tables to the index files	30
General index file format.....	31
Frequently asked questions	31
Image File Format	32
TIFF 6.0 baseline bilevel images	33
Fields to omit.....	34
Informational fields	34
Compression schemes.....	34
Paper formats.....	35
Conformity Testing	36
CADDY compliance	36
Media	36
Page Files	36
Index Files.....	36
Other Files.....	36
Conformity Test Software	37
References.....	38
Glossary.....	39
Index.....	40



Appendix A	Structure definition of CADDY index files	41
Changes to CADDY version 1.1.....		42
Index structure of CADDY 2.0.....		43
Legend		43
Table dh_ver: version history		44
Table dh_vol: volume list.....		44
Table dc_rep: report list		45
Table dc_toc: table of contents		46
Table dc_pag: page table		46
Table dc_lnk: hyperlinks		47
Table dc_ccf: controlled files.....		47
Table dh_has: header attribute set		48
Table dh_asu: header active substances.....		48
Table dh_rps: relation product substance		48
Table dh_pro: header products		49
Table dh_com: header companies.....		49
Table da_chg: changes history		49
Table dd_tab: data dictionary tables		50
Table dd_col: data dictionary columns.....		50



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 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, January 31, 1996) was published after being accepted by EC, EU Member States and ECPA. CADDY 1.0 was not used for compiling dossiers and is not supported any longer.

CADDY version 1.1

The second version, called CADDY 1.1, was released September 23, 1997. 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. There was an addendum to the format specification CADDY 1.1 released on February 24, 1998. CADDY 1.1 was used to submit European dossiers and must therefore be supported by all current CADDY viewers.

CADDY version 2.0

The current version is called CADDY 2.0. This enhanced version was created from CADDY 1.1 by adding some enhancements that were requested by European authorities. The format specification (CADDY 2.0) was enhanced by adding support for CADDY controlled files, CADDY hyperlinks and other features that were agreed upon by the Joint CADDY Steering Group.

Changes to CADDY 1.1

- Additional concept for CADDY file attachments (CADDY controlled files).
CADDY file attachments are additional files that are linked to the CADDY pages. These files are controlled in terms of dossier versioning. Their content or format is not limited by CADDY.
- Additional concept for CADDY hyperlinks.
CADDY Hyperlinks are dossier internal hyperlinks from an area (x-Coordinate, y-Coordinate, width, height) on a source page to either a location (x, y) on a target page, a table-of-contents entry or a controlled file. A submitter of a CADDY dossier will provide these links.
- Enhancements to include the experience that was gained with CADDY 1.1.
This includes some minor changes in the TIFF tag limitations and some minor changes in the table Dossier Version.
- Improvement of readability and inclusion of CADDY 1.1 Addendum^[17] into format specification 2.0 document.

Future CADDY 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 2.0. The readability of CADDY 2.0 dossiers by future CADDY viewers has to be achieved.



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 included in Format Specification 1.1.

In 1999 the international situation changed. US and Canada started to evaluate alternatives to CADDY and the DTWG started a survey on CADDY experience with European authorities in order to improve the CADDY system. The strategic goal was kept but it was amended by a statement to request enhanced support for the reviewing process for future CADDY versions. Results of the CADDY experience from the last four years were built into the CADDY 2.0 format specification. New CADDY Retrieval Software is developed in parallel.

What is a CADDY dossier?

A CADDY dossier is an electronic representation of a complete pesticide registration dossier as a series of pages together with structural access information.

CADDY Features

- Table of contents and report list
- Hyperlinks
- Attached source files (like tables in spreadsheet format, documents in native word processor formats, ...)
- Dossier version control mechanism
- Confidentiality support
- Suitable as replacement for paper copies of the dossier
- Suitable for long term archiving
- Suitable as data delivery system for global dossier databases
- Suitable as basis for review support systems (appropriate retrieval/communication software required)
- Creatable and readable on all common operating system platforms like MS-Windows, Unix,... (Appropriate retrieval/compilation software required)
- Pages can be displayed by a wide variety of standard imaging software
- Index information can be read by a wide variety of standard database applications



Standardization objectives

The following basic decisions for CADDY were made by the Data Transfer Steering Group in 1995 (CADDY 1.0) after discussion with authorities and industry.

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.

List of objectives

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

- The first release of CADDY should be very simple and is not intended to cover more than absolutely necessary.
- The user of a dossier (reviewer) 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 for CADDY were made by the Data Transfer Steering Group:

- 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 in the Data Transfer Steering Group 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.

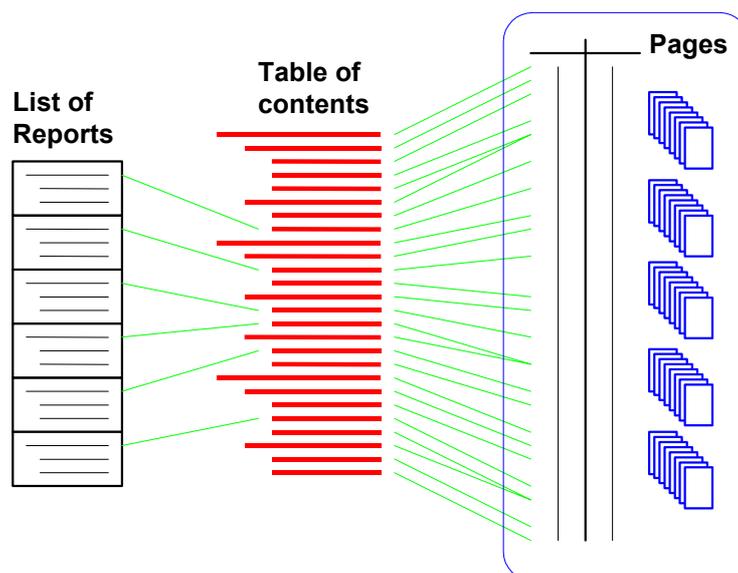


Figure 1 *Retrieval possibilities*

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

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 can be found in the Application Guide.

CADDY Retrieval Software

Every organization that is using CADDY may create its own software for accessing CADDY dossiers. However, in order to enhance the acceptance of CADDY a common standard retrieval software was developed by the ECPA. This software includes a lot of useful features, like annotations etc., to support the review process for a pesticides dossier. The CADDY Retrieval Software (RSW) will be enhanced in order to provide support for CADDY 2.0.

Next steps

The following documents and software should be updated, before CADDY 2.0 is used as a new standard to submit pesticides registration dossiers:

- Example application on CD-ROM
(Short example for the application of the CADDY standard)
- CADDY Application Guide^[18]: How to use the format specification to file a pesticides registration dossier on electronic media
- CADDY Retrieval Software (RSW)
- CADDY Conformity Test Software (CTSW)



Explanation of document specific terms

CADDY format version vs. CADDY dossier version

There are two different "versions" referred to in the format specification: The CADDY format version and the CADDY dossier version.

The CADDY format version is the version number of the CADDY format specification. For the current format specification the format version number is "2.0".

The CADDY dossier version is the version number assigned to a version of a CADDY dossier. The dossier version number starts with 1 for the first version of each dossier and is increased by one with each supplement.

CADDY compilation date vs. issue date

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" ^[18].



Media

There was an urgent need to choose a medium for transfer and storage of pesticides registration dossiers that 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 that 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

<p style="text-align: center;">Yellow Book, Mode 1 (and in case of CD-R: Orange Book, DAO)</p>
--

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

<p style="text-align: center;">ECMA-119, Interchange Level 1</p>

In the future it might be necessary to choose additional media types for new CADDY format 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 five 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 Files**
A label file is stored on every volume to allow an easy identification of the volume.
- **Controlled Files**
These files are controlled attachments to the dossier that can have any format. The CADDY controlled files (CCF) are controlled regarding the version control mechanism of CADDY. The content and the format is not controlled. CADDY controlled files allow to add renditions (alternative electronic format) of reports or to add data elements like tables or pictures in native format.
- **Unsupported Files**
These files are not covered by the CADDY format. They can be stored on the CD-ROMs but they are not controlled in terms of content or version control. These files are stored in specific "FILES"-directories.

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 and the controlled files.

The file system on the volumes has to be readable by a wide variety of different operating systems. To avoid possible difficulties with Non-Microsoft operating systems all file names and directory names have to be handled case sensitive.

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.

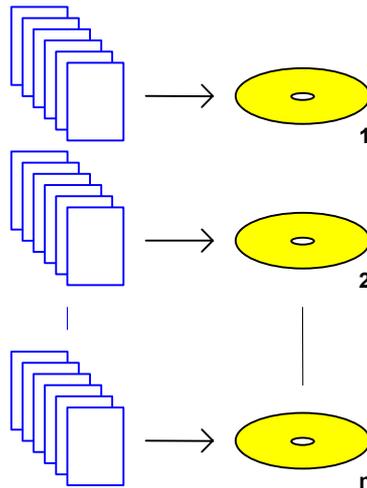


Figure 2 *Spread pages on volumes*

A single page file cannot be split across two volumes.

Index 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 and controlled files that may be contained on index volumes too.

Confidential volumes

If there are any confidential pages or confidential controlled files 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 or confidential controlled files, 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 files are always stored on non confidential volumes.



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.

Distribution of controlled files to the volumes (recommendation)

It is recommended that the controlled files are stored on the same volume as the corresponding page files. Nevertheless, the retrieval software cannot rely on any special distribution of the controlled 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, only capital letters are allowed.)

[country shortcode] ::= Country code of the notifying company (short form with 2 characters)
(ISO 3166 - Alpha-2-code, only capital letters are allowed.)

[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 name]
	[format string]	[format string]
	[volume string]	[volume string]
	[version string]	[version string]
	[notifying company]	[notifying company]
	[country name]	[country name]
	[active substance]	[active substance]
	[CIPAC code]	[CIPAC code]
	[CAS code]	[CAS code]
[volume name]		name of the volume as described above
[format string]		CADDY format version string ("CADDY 2.0") (see example below and structure definition in Appendix A)
[volume string]		either "confidential" or "index volume" or "." (see example below and structure definition in Appendix A)
[version string]		CADDY dossier version number (see field ver_UVER) followed by the CADDY compilation date (see field ver_master_date) in English format both parts are separated by a comma and a space (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 the 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
CADDY 2.0
.
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 dossier 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.

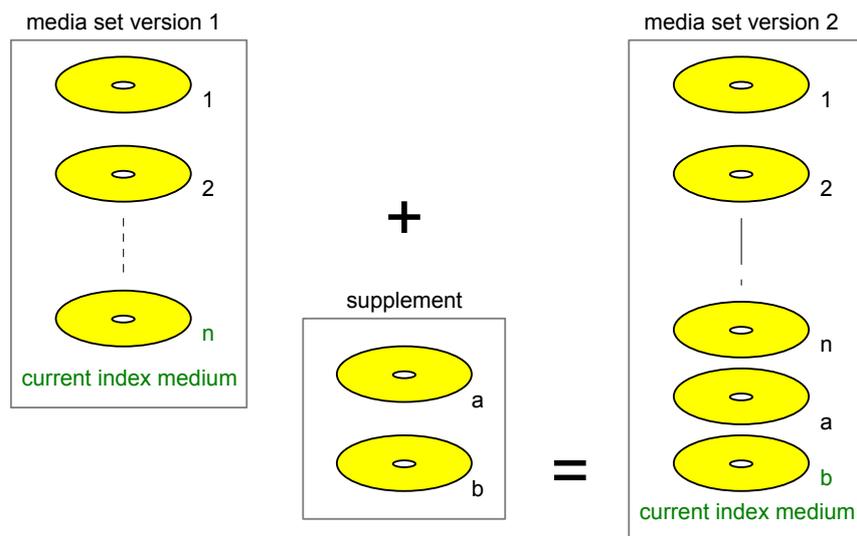


Figure 3 Adding supplements

All volumes of a supplement must have the same CADDY format version string within their label files. Supplements must not use CADDY format versions that are older than the format version used for the previous dossier version. In the example above it is possible that media set version 1 was issued in CADDY format 1.1 and the supplement containing the volumes (a) and (b) is issued in CADDY format 2.0. All further supplements have to be issued at least in CADDY format 2.0.



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.

File names of index files are built by using the corresponding table name without a file name extension. The description of the tables and table names can be found in Appendix A.

There are three groups of index files

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

Basic index files

The following basic index files are defined within CADDY 2.0:

index file	short description
report list	list of all reports contained in the current dossier version
table of contents	hierarchical table of contents of the current dossier version
page table	list of all pages of the current dossier version
controlled files table	list of all controlled files of the current dossier version
hyperlink table	list of all hyperlinks contained in the current dossier version
version history	list of all versions of the dossier
volume list	list of all volumes belonging to the current volume set
header attribute sets	history of dossier attributes for all versions of a dossier
header companies	owner companies of the dossier
header products	product names covered by the dossier
header active substances	active substances covered by the dossier
relation product – substance	relation between products and active substances

Changes history

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

index file	short description
changes history	list of all pages and controlled files that are deleted and of all report attributes that are changed or deleted

This additional file has to be present if pages or controlled files were deleted or if report attributes were deleted or changed. It is useful to determine the differences to a previous dossier version. The file can also be produced by using the page table, the CCF table and the report list of all old dossier versions and the page table, the CCF table and the report list of the current dossier version. Since these tables



can be very big and a dossier can have many dossier versions, 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 appropriate new rows for changes have to be added.

Data dictionary

A data dictionary for all the files belonging to the CADDY index files.

Two index files are used to store the data dictionary:

index file	short description
table list	list of all tables which can be contained in directory "INDEX"
column list	list of all columns of the tables listed in the table list

The data dictionary is a constant for the version 2.0 of CADDY. It may be changed in future CADDY format versions. Only one column within the table list is an exception. This column contains the values appropriate for the current CADDY dossier version.

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.

Storage Location

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 directory PAGES and all of its subdirectories are only allowed to contain page files or subdirectories. They must not be empty.

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 built 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"

Controlled files

Unique file ID

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

Storage Location

The retrieval of controlled files is done by using the controlled files table. This table contains information about the storage location (volume) of the file. All controlled files are stored on the volumes in subdirectories of a directory with the name "CCF". The parent directory of the directory "CCF" has to be the root directory. The directory CCF and all of its subdirectories are only allowed to contain controlled files or subdirectories. They must not be empty.

The **basename** of a controlled file consists of the UFID padded with leading zeros to 8 characters. The file name of a controlled file is formed using its basename and the extension ".CCF".

basename + ".CCF"

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 built 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 controlled 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 + ":" + "CCF" + "\" + **subpath** + "\" + **basename** + ".CCF"
- The path name in a UNIX environment is built as follows: (**pathdelimiter** = "/")
CD-ROM mounting point + "/" + "CCF" + "/" + **subpath** + "/" + **basename** + ".CCF"

Copies of page files and controlled files

It is possible to store additional copies of already submitted page files or controlled 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 page files and controlled files 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 or controlled files any more. Even if copies of page files or controlled files are submitted, the references to the original page files are kept in the dossier database for audit trail reasons.



Unsupported additional files (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. These files are called unsupported files, because they are not supported by the CADDY dossier version mechanisms. The option of storing unsupported files on CADDY CD-ROMs 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 word processor files accessible together with the dossier. The content of the directory "FILES" is not standardized, which means, that there has 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 in the international CADDY discussions, but is not within the scope of this format specification. It is recommended by the CADDY steering group to avoid the usage of unsupported files. Instead, CADDY controlled files should be used. They have the advantage of links to the dossier pages and the dossier version mechanism of CADDY.

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 important to distinguish between uppercase and lowercase letters within file names and directory names?

Yes

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.

Must each volume contain a PAGES directory ?

No. If a volume contains no page files the directory PAGES must not exist.

Is it allowed to use the extension ".tif" for page files ?

No, only uppercase letters are allowed for the extension of page files.

Is it allowed to use the extension ".txt" for index files ?

No, there is no extension for index files allowed.

How many lines of text are contained within the label file ?

Each label file has to contain 8 lines of text.



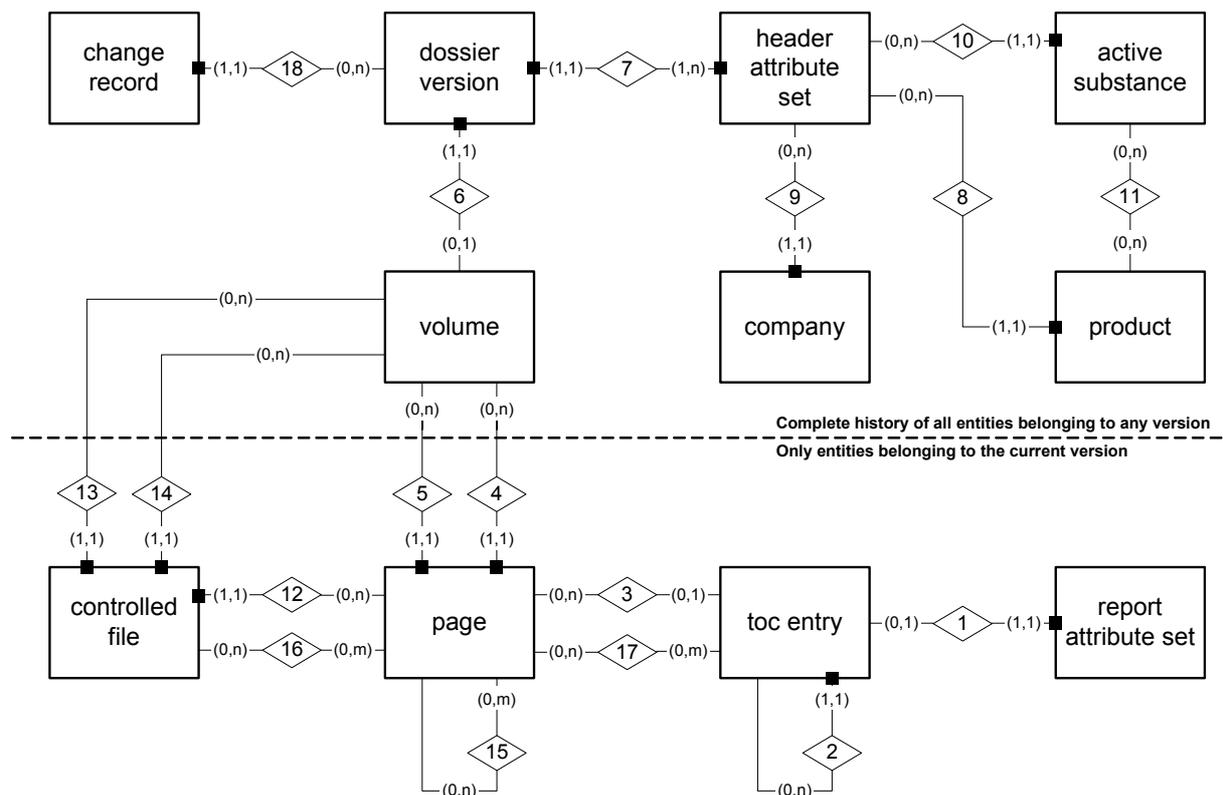
Dossier database

The index information of a CADDY dossier is organized as a relational database. The tables are stored in the basic index files and the changes history. The index files for the data dictionary do not belong to the dossier database.

First an overview on entities and relationships is given in an entity relationship model. This model is refined further by explaining the tables belonging to the dossier database. The complete database design is shown in "Appendix A".

Entity Relationship Model

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



Legend

This is a logical ERM model, independent of the concrete database design. The item x is a relation between entity 1 and entity 2, that might be implemented in a separate table and can have attributes attached. The tuples (x, y) are called „weights“. Example: i is the minimum number of occurrences of entity 1 in relation x , n is the maximum number of occurrences of entity 1 in relation x . The black square signals a mandatory occurrence of entity 1 in relation x , which means $i > 0$.

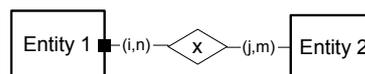


Figure 4 Entity Relationship Model



Entities

dossier version	version of a CADDY dossier
header attribute set	set of header attributes belonging to dossier versions
active substance	active substance belonging to a header attribute set
company	a company name belonging to a header attribute set
product	a product name belonging to a header attribute set
volume	volume (CD-ROM) of a CADDY media set
page	single page of a dossier
controlled file	controlled file that is attached to a page
toc entry	entry within the hierarchical table of contents
report attribute set	set of report attributes representing an instance of a report
change record	log entry representing "deleted" pages, controlled files or report attribute sets

Relationships

- (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 sub nodes.
- (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.
- (12) Every controlled file has a link to a page. (mandatory)
- (13) Every controlled file has a link to the original volume (volume entry) where the controlled file was originally stored. (mandatory)
- (14) Every controlled file has a link to the current volume (volume entry) where the controlled file is currently stored. (mandatory)
- (15) Every page can have one or more hyperlinks to pages. (optional)
- (16) Every page can have one or more hyperlinks to controlled files. (optional)
- (17) Every page can have one or more hyperlinks to TOC entries. (optional)



- (18) Every deleted item (or changed report attribute set) was available last in the referenced dossier version (mandatory)

The realization of the relationships that are not stored in separate tables is shown in the tables in Appendix A.

Database table descriptions

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

Tables covering all dossier versions

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.

The names of all tables within the dossier database will start with the prefix "dh_".

table	table name	content
version history	dh_ver	entity dossier version
volume list	dh_vol	entity volume
header attribute set	dh_has	entity header attribute set
header active substances	dh_asu	entity active substance
header products	dh_pro	entity product
header companies	dh_com	entity company
relation substance - product	dh_rps	relation <11>

Table version history

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

Table volume list

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

Table header attribute set

The dossier header is a set of attributes belonging to the whole dossier. A history of header 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 attribute 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 attribute 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)

Table header active substances

This table contains a history of all active substance names that were attached to the attribute sets of the dossier. For European dossiers the annex II substances are stored in this table.

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.

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

- active substance
- CIPAC code
- CAS number

Table header products

This table contains a history of all product names that were attached to the attribute sets of the dossier. For European dossiers the annex III products are stored in this table.

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.

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

- product name
- formulation type

Table header companies

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

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.

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

- company name
- company country name
- company code
- company country code



Table relation product - substance

This table contains the relation between products and active substances. Each record describes the total concentration of one active substance in one product. The referenced product and the referenced active substance have to belong to the same header attribute set.

The following attributes of a dossier are stored in the table relation product - substance:

- total concentration of substance within product

Tables covering current dossier version

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

The names of all tables within the dossier database will start with the prefix “dc_”.

table	table name	content
report list	dc_rep	entity report attribute set
table of contents	dc_toc	entity TOC entry
pages table	dc_pag	entity page
controlled files table	dc_ccf	entity controlled file
hyperlink table	dc_lnk	relations <15>, <16>, <17>

Table report list

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).

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 belong to reports:

- dossier file number (for EU) / MRID (for US and 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 table-of-contents

The TOC table represents a hierarchical table of contents (TOC tree) for the current dossier version. Every entry in this table represents a node in the TOC tree. One entry represents the root of the tree. All other entries are sub nodes. 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).

The following user attributes belong to table of contents entries:

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

Table page table

This table stores the access information for the pages. It contains a list of all pages of the current dossier version.

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 user attributes belong to pages:

- pagination string

Table controlled files

This table stores the access information for the controlled files. It contains a list of all controlled files of the current dossier version.

Every entry corresponds to a controlled file on the volume set. The table is ordered by the order of the controlled files in the dossier. This order is not explicitly stored on the volume. The primary key of this table is the unique file ID (UFID).

The following user attributes belong to controlled files:

- file title
- file extension

Table hyperlinks

This table stores the hyperlinks inside the dossier. It contains a list of all hyperlinks within the current dossier version. Every entry corresponds to a hyperlink between one area on a page and either a table-of-contents entry, a controlled file or a position on another page.

The coordinates system (X, Y) is derived from the TIFF coordinate scheme: X defines the offset in [1/10 mm] of the left side of the image; Y defines the offset in [1/10 mm] of the top of the image. In the TIFF coordinate scheme, the positive X direction is right and the positive Y direction is down, so that both values are always positive.

The table is ordered by the order of the source pages in the dossier. This order is not explicitly stored on the volume. The primary key of this table is the unique link ID (ULID).

The following user attributes belong to hyperlinks:

- hyperlink title



Confidentiality of pages, controlled files, 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 any software using the page confidential flag can easily determine it. 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.

Controlled files are marked confidential if and only if they are linked to a confidential page.

Tracking of changes between dossier versions

The changes history mechanism applies to controlled files, pages and report attribute sets. All three entities are stored in tables, where only the current entities are contained. These entities have an attribute (relation to dossier version) which indicates in which version they were added to the dossier. In addition there is an attribute for report attributes which indicates in which version they were changed. Entities which were removed in a specific dossier version are not contained in these tables any longer. Instead they are added to the table changes history.

Frequently asked questions

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

Yes

Which tables of the dossier database must have at least one entry ?

volume list / version history / table of contents / page table / header attribute 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**.

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

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].

$$[\text{pchar}] ::= \text{char}(x) \mid \text{char}(y) \mid \text{char}(z) \quad \& \ 32 \leq x \leq 126 \ \& \ 161 \leq y \leq 172 \ \& \ 174 \leq z \leq 255$$

allowed graphic characters



	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00																
10																
20		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
80																
90																
A0		ı	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	¯
B0	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D0	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F0	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Figure 5 ECMA 94 Latin 1 character set

Two additional control characters are used:

- [cr] ::= chr(13) carriage return character
- [lf] ::= chr(10) line feed character

The following characters and tags have special meanings:

- [eol] ::= [cr] | [cr] [lf] end of line tag
- [delimiter] ::= "|" delimiter character for general index file format
- [escape] ::= "\" escape character for general index file format
- [false] ::= "0" logical value representing "FALSE"
- [true] ::= "1" logical value representing "TRUE"



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.

$\text{dec}(\text{enc}([\text{text}])) = [\text{text}]$ and $\text{enc}(\text{dec}([\text{dbtext}])) = [\text{dbtext}]$



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 some standard database software available on UNIX and MS-DOS that can read the specified format directly 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 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. There is a wide variety of software on the market 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	Mandatory Fields	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	mandatory	only 300 dpi images
ImageLength, ImageWidth	L or S	none	im	mandatory	A4 or Letter (portrait or landscape) see "Paper formats" below
Orientation	S	1	1		
FillOrder	S	1	1		

Type: S = Short
 L = Long
 R = Rational

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

Valid values: im image dependent

Mandatory fields: mandatory mandatory tag (has to be contained in each CADDY tif-file)



Fields to omit

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

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
Thresholding, CellWidth, CellLength
DocumentName, PageName, PageNumber

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

Compression schemes

The compression types allowed 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 (ImageLength and ImageWidth) are possible:

Portrait	A4	Letter
ImageLength	3508	3300
ImageWidth	2480	2550

Landscape	A4	Letter
ImageWidth	3508	3300
ImageLength	2480	2550

The values given in the tables are the standard values for the fields ImageLength and ImageWidth.

The maximum values of ImageLength and ImageWidth are 10 pixel more than the standard values. (This rule should allow for more flexibility to ignore unprecise calculations of scanning software.)

The minimum values of ImageLength and ImageWidth are 35 pixel less than the standard values. (This rule should avoid to the submission of much smaller images than A4 or Letter.)



Conformity Testing

CADDY compliance

A CADDY dossier is CADDY compliant if all the requirements, described in this format specification, are met. Beside the format requirements it must also be noted that a dossier needs to be carefully prepared in order to be evaluated and kept for the future. For example, this includes not only the technical preparation of the CD-ROM but also the thorough preparation of a meaningful index and good quality images of the dossier pages.

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

1. media conformance
2. page file conformance
3. index file conformance
4. other files

Media

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

Remark: There are a lot of different formats that can be written onto CD-ROM, but if the wrong format is used it can be not expected that the CD-ROM is readable on future operating systems.

Page Files

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.

Readability and quality of page files will not be checked by the conformity test system of CADDY. The submitter of a dossier must ensure them by applying appropriate quality measures.

Index Files

The following checks on the index files will be performed by the conformity test system of CADDY:

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.

Other Files

- Controlled files will be checked only for their existence. The content is not accessed by the conformity test system of CADDY.
- Label files will be checked for conformity of their syntax.
- Unsupported files will not be checked.



Conformity Test Software

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 the

- test facility name
- test person
- date and time of the test
- dossier name and version
- version of conformity test software used for testing

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

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



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, 1996
Joint EU Member States / ECPA Data Transfer Steering Group, (January 1996)
- [16] Document Interchange Format for PPP registration applications
Format Specification CADDY 1.1 / September 23, 1997
Joint EU Member States / ECPA Data Transfer Steering Group, (September 1997)
- [17] Addendum to Format Specification CADDY 1.1 / February 24, 1998
Joint EU Member States / ECPA Data Transfer Steering Group, (February 1998)
- [18] CADDY Application Guide / April 22, 1998
Joint EU Member States / ECPA Data Transfer Steering Group, (April 1998)



Glossary

ACPA	American Crop Protection Association
CADDY	Computer Aided Dossier and Data Supply
CAS	The CAS code is a chemical identification code used for identification of an active substance
CCITT	Consultative Committee for International Telegraphy and Telephony
CD-R	CD-Recordable
CD-ROM	Compact Disc Digital Read Only Memory
CIPAC	Collaborative International Pesticides Analytical Council The CIPAC number is a chemical identification code used for identification of an active substance
DAMOS	Drug Application Methodology on Optical Storage
DAO	disk-at-once (method for burning CD-Rs)
DIN	Deutsches Institut für Normung (German institute for standardization)
DTSG	Data Transfer Steering Group (CADDY work group of EC and ECPA)
DTWG	Data Transfer Working Group (CADDY work group of ECPA)
EC	European Commission
ECMA	European Computer Manufacturers Association
ECPA	European Crop Protection Association
EEC	European Economic Community
EPA	Environmental Protection Agency
EU	European Union
GLP	Good Laboratory Practice
ISO	International Standardization Organization
JCSG	Joint CADDY Steering Group (International CADDY working group)
MRID	Master Record ID
MS	Microsoft
PMRA	Pesticides Management Regulatory Agency
TIFF	Tagged Image File Format
US	United States

Document specific abbreviations

A4	Paper format DIN A4
CCF	CADDY controlled file
CR	Carriage return character
CTSW	Conformity Test Software
EOL	End of line tag
LF	Line feed character
NBSP	No-break space character
RSW	Retrieval Software
SHY	Soft hyphen character
TOC	Table of contents
UFID	Unique file ID
UHID	Unique header ID
ULID	Unique link ID
UNID	Unique node ID
UPID	Unique page ID
URID	Unique report ID
UVER	Unique version ID
UVOL	Unique volume ID



Index

A

active substance · 15, 17, 22,
23, 24, 25, 39, 41
additional copies · 19
additional files · 20
annotations · 9
Application Guide · 9, 10, 38

C

CADDY 1.0 · 6, 8, 28, 38, 41
CADDY 1.1 · 6, 38, 41
CADDY dossier · 6, 7, 9, 10, 15,
18, 20, 21, 22, 27, 36
CAS code · 15, 24, 39
CD-ROM · 8, 9, 11, 12, 15, 18,
19, 20, 22, 36, 38, 39
changes history · 17, 18, 21, 27,
41
character set · 28, 29
CIPAC code · 15, 24, 39
compatibility · 6, 11, 17
compilation software · 7, 8, 18,
19, 24, 25
confidential volumes · 13
confidentiality · 9, 13, 15, 16, 20,
25, 27
conformity test · 36, 37
controlled files · 6, 12, 13, 14,
17, 19, 20, 22, 25, 26, 27, 39,
41
current dossier number · 14
current index volume · 13, 16,
17, 20, 22

D

data protection · 25
data security · 9
data type · 28, 30, 36, 37
database applications · 7, 8, 12
database software · 31
dossier database · 7, 12, 17, 19,
21, 23, 25, 27, 43
dossier header · 23
dossier owners · 24

dossier version · 10, 16, 17, 18,
20, 22, 23, 25, 26, 27, 37

E

entity relationship · 21

F

format specification · 6, 7, 8, 9,
10, 20, 36
format version · 10, 11, 15, 16,
17, 18, 43

H

header attribute set · 17, 22, 23,
24, 25, 27, 41
hyperlinks · 6, 17, 22, 25, 26, 41

I

image · 32, 33, 35, 36
image file · 12, 32, 33, 34, 36
image file format · 12
imaging software · 7
index file format · 8, 12, 15, 29,
31, 36, 37
index files · 12, 13, 15, 17, 18,
20, 21, 27, 28, 30, 31, 36, 37,
41, 43
index volume · 13, 15, 16, 17,
20

L

label file · 12, 15, 16, 20, 37
long term archiving · 7

M

medium · 6, 7, 8, 9, 11, 12, 14,
16, 18, 19, 22, 36
modular design · 8
MS-DOS · 2, 31

N

notifying company · 14, 15

O

operating system · 7, 11, 32, 36

P

page table · 17, 18, 26, 27, 37,
41
pagination string · 26

R

referenced page · 27
report list · 7, 8, 17, 25, 27, 41
retrieval software · 7, 8, 9, 14,
18, 19, 24, 25, 39, 43

S

storage location · 18, 19
supplements · 16, 18, 19, 25

T

table of contents · 8, 12, 17, 22,
25, 26, 27, 39, 41
TIFF · 2, 8, 32, 33, 34, 36, 37,
38, 39
compression types · 34
paper formats · 35
resolution · 32, 33, 35

U

UNIX · 2, 19, 31
unsupported files · 20, 36

V

version history · 17, 23, 27, 41
version string · 15
volume list · 17, 19, 23, 27, 41
volume number · 14
volume set · 12, 13, 14, 16, 17,
26
volume set ID · 14



Appendix A

Structure definition of CADDY 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	table name index file name	table version	index file required
version history	dh_ver	3	yes
volume list	dh_vol	2	yes
report list	dc_rep	2	
table of contents	dc_toc	2	yes
page table	dc_pag	2	yes
hyperlinks	dc_lnk	3	
controlled files	dc_ccf	3	
header attribute set	dh_has	2	yes
header active substances	dh_asu	2	
relation product-substance	dh_rps	2	
header products	dh_pro	2	
header companies	dh_com	2	
changes history	da_chg	2	
data dictionary tables	dd_tab	2	yes
data dictionary columns	dd_col	2	yes

Table version: 1 = Table in CADDY 1.0 format (not used)
2 = Table in CADDY 1.1 format (not changed in CADDY 2.0)
3 = Table in CADDY 2.0 format

Index file required: If “yes”, this file is mandatory in the INDEX directory,
otherwise there are situations, were this file can be omitted.



Changes to CADDY version 1.1

Table	Field	Changes
dc_ccf		table was added
dc_Ink		table was added
dh_ver	ver_max_UNID	field was removed
dh_ver	ver_max_URID	field was removed
dh_ver	ver_max_UPID	field was removed
dh_ver	ver_chg_ccf	field was added
dh_ver	ver_chg_Ink	field was added

The name of some fields were changed to harmonize the naming mechanisms for field names. Because the field names are not contained within the index files, this will only affect the content of the table dd_col.



Index structure of CADDY 2.0

Legend

Field name

The column *field name* gives a name for each field. These names are used within the current format version only. Field names have a table specific prefix like "ver_" or "pag_".

There is one special field in each database table called *order*, that is not explicitly contained within the index files. Instead the *order* field contains the order of lines within an index file. The order of rows in a table 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.

Type

The column *type* contains the field type, one of the types defined in the chapter "Index File Format".

Min and Max

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.

For the field types *boolean* and *date* the values of *min* and *max* are not used.

Field description

The column *field description* contains a description of the field and can contain the following remark items:

[key]	This field is a key for the table.
[FK: UxID]	Foreign key to table with key UxID.
[line number]	This field is not stored in the index file. Its value is the line number of the row inside the index file. It is recommended to create this field in the dossier database.
[unique]	The value has to be unique throughout the table. E.g. all [key] fields are [unique].
[index]	It is recommended to use this field in the dossier database as index. E.g. all [key] fields are [index].
[not visible]	The retrieval software should not display the value of this field to the reviewer. E.g. all [FK: UxID] and [key] fields are [not visible].
[not used]	This field contains no valuable information. E.g. some [line number] fields are [not used].



Table dh_ver: version history

field name	type	min	max	field description
ver_order	integer	1	999	order of rows on CDROM = UVER [not used] [unique] [not visible]
ver_UVER	integer	1	999	[key] unique version number [unique] [index]
ver_master_date	date			day of mastering the electronic CADDY version of the dossier (CADDY compilation date) [unique] see also [version string] in file label.txt
ver_idxvol_UVOL	integer	1	999	[FK: UVOL] current index volume [unique] [not visible]
ver_stdv	integer	1	99	format version number of CADDY standard [not visible] ver_stdv = 2: version is CADDY 1.1 conform ver_stdv = 3: version is CADDY 2.0 conform
ver_attrib_UHID	integer	1	99	[FK: UHID] header attribute set [not visible]
ver_issue_date	date			issue date of the original dossier
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 ?
ver_chg_ccf	boolean			controlled files changed?
ver_chg_ink	boolean			hyperlinks changed?

Table dh_vol: volume list

field name	type	min	max	field description
vol_order	integer	1	999	order of rows on CDROM = UVOL [not used] [unique] [not visible]
vol_UVOL	integer	1	999	[key] unique volume number [unique] [index]
vol_name	text	1	12	unique volume name [unique] [not visible] same string as in label.txt file: The volume name is described in chapter Volume Sets of the Format Specification.
vol_pages	integer	0	99999999	number of pages on this volume
vol_index	boolean			contains index files ?
vol_confidential	boolean			contains confidential files (confidential volume) ? vol_confidential = TRUE: This volume contains only pages flagged as confidential (pag_confidential = TRUE) and controlled files flagged as confidential (ccf_confidential = TRUE) vol_confidential = FALSE: This volume contains only non-confidential pages (pag_confidential = FALSE) and non-confidential controlled files (ccf_confidential = FALSE)
vol_necessary	boolean			contains current index files, page files or controlled files ? vol_necessary = FALSE: This volume is not the current index volume and does not contain any of the pages referenced by pag_curvol_UVOL in the current page table and does not contain any of the controlled files referenced by ccf_curvol_UVOL in the current controlled files table



Table dc_rep: report list

field name	type	min	max	field description
rep_order	integer	1	9999	order of rows on CDROM = order of reports in dossier [unique] [index] [not visible]
rep_URIID	integer	1	9999	[key] unique report ID [unique] [index] [not visible]
rep_tocnode_UNID	integer	1	99999999	[FK: UNID] toc entry representing the report [unique] [not visible]
rep_doss_fn	text	0	20	dossier file number (for EU) / MRID (for US and Canada)
rep_comp_fn	text	0	20	company file number
rep_date	date			date of document [index] If rep_valid_month and rep_valid_day is TRUE, then rep_date contains the effective date of the report. If rep_valid_day = FALSE and rep_valid_month = TRUE, then the day of the month is unknown and rep_date should only be displayed by the software without a day. For sorting and consistency reasons the rep_date field will contain the first day of the month. If rep_valid_day and rep_valid_month = FALSE, then the month of the year is unknown and rep_date should only be displayed by the software as a year. For sorting and consistency reasons the rep_date field will contain the first January of the year. The combination rep_valid_day = TRUE and rep_valid_month = FALSE is forbidden.
rep_valid_day	boolean			is day of report date applicable ? [not visible] for explanation see field rep_date
rep_valid_month	boolean			is month of report date applicable ? [not visible] for explanation see field rep_date
rep_title	text	0	240	title of document
rep_authors	text	0	120	authors of document [index]
rep_source	text	0	180	source of document
rep_owners	text	0	120	owners of report at submission date
rep_test_facility	text	0	60	test facility
rep_confidential	boolean			is report confidential ? (Documents J) rep_confidential = TRUE: All pages of the report are to be marked as confidential
rep_glp	boolean			is report GLP conform ?
rep_published	boolean			was report published ?
rep_vertebrates	boolean			test on vertebrates ?
rep_protect	boolean			data protection claimed ?
rep_added_UVER	integer	1	999	[FK: UVER] version in which report was added first [not visible] This value can never be changed in new versions. If value > 1 then the retrieval software should display a hint for the user.
rep_changed_UVER	integer	0	999	[FK: UVER] version in which the report attributes where changed last [not visible] rep_changed_UVER = 0: The entry of the report was not changed since first added. (==> when a report entry is new then value = 0)



Table dc_toc: table of contents

field name	type	min	max	field description
toc_order	integer	1	99999999	order of rows on CDROM = order of toc nodes [unique] [index] [not visible]
toc_UNID	integer	1	99999999	[key] unique node ID [unique] [index] [not visible] Root node: toc_UNID = 1 and toc_parent_UNID = 0 and toc_level = 1
toc_report_URID	integer	0	9999	[FK: URID] report attributes [index] [not visible] toc_report_URID = 0: There are no report attributes attached (NIL-Pointer)
toc_parent_UNID	integer	0	99999999	[FK: UNID] parent node [not visible] Root node: toc_UNID = 1 and toc_parent_UNID = 0 and toc_level = 1
toc_level	integer	1	99	level in toc hierarchy [not visible] Root node: toc_UNID = 1 and toc_parent_UNID = 0 and toc_level = 1
toc_page_UPID	integer	0	99999999	[FK: UPID] first page [not visible] toc_page_UPID = 0: No page is attached (NIL-Pointer) toc_page_UPID > 0: 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.
toc_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]
toc_text	text	1	70	abbreviated toc entry string [index]

Table dc_pag: page table

field name	type	min	max	field description
pag_order	integer	1	99999999	order of rows on CDROM = order of pages in dossier [unique] [index] [not visible]
pag_UPID	integer	1	99999999	[key] unique page ID [unique] [index] [not visible]
pag_curvol_UVOL	integer	1	999	[FK: UVOL] current storage location [not visible] If no page copy exists, then pag_curvol_UVOL = pag_orgvol_UVOL
pag_orgvol_UVOL	integer	1	999	[FK: UVOL] original storage location [not visible]
pag_confidential	boolean			is page confidential ? (Documents J)
pag_pagination	text	1	20	pagination string which is shown (printed) on the page [index]
pag_added_UVER	integer	1	999	[FK: UVER] version in which page was added [not visible] This value can never be changed in new versions. If value > 1 then the retrieval software should display a hint for the user.



Table dc_Ink: hyperlinks

field name	type	min	max	field description
Ink_order	integer	1	99999999	order of rows on CDROM = order of hyperlinks in dossier [unique] [index] [not visible]
Ink_ULID	integer	1	99999999	[key] unique link ID [unique] [index] [not visible]
Ink_link_title	text	1	50	title for display purposes
Ink_source_UPID	integer	1	99999999	[FK: UPID] source page [index] [not visible]
Ink_source_x	integer	0	9999	x-coordinate of source on source page [1/10 mm] [not visible]
Ink_source_y	integer	0	9999	y-coordinate of source on source page [1/10 mm] [not visible]
Ink_source_w	integer	0	9999	width of source on source page [1/10 mm] [not visible]
Ink_source_h	integer	0	9999	height of source on source page [1/10 mm] [not visible]
Ink_target_type	text	1	1	"P" = page, "N" = TOC node, "F" = controlled file [not visible]
Ink_target_UxID	integer	1	99999999	[FK: UxID] unique ID of target (UPID, UNID or UFID) [index] [not visible]
Ink_target_x	integer	0	9999	x-coordinate on target page [1/10 mm] [not visible]
Ink_target_y	integer	0	9999	y-coordinate on target page [1/10 mm] [not visible]

Table dc_ccf: controlled files

field name	type	min	max	field description
ccf_order	integer	1	99999999	order of rows on CDROM = order of CCF's in dossier [unique] [index] [not visible]
ccf_UFID	integer	1	99999999	[key] unique file ID [unique] [index] [not visible]
ccf_curvol_UVOL	integer	1	999	[FK: UVOL] current storage location [not visible] If no copy exists, then ccf_curvol_UVOL = ccf_orgvol_UVOL
ccf_orgvol_UVOL	integer	1	999	[FK: UVOL] original storage location [not visible]
ccf_page_UPID	integer	1	99999999	[FK: UPID] page where CCF is attached [index] [not visible]
ccf_confidential	boolean			is CCF confidential ?
ccf_title	text	1	50	title string which is shown (printed) on the virtual directory [index]
ccf_extension	text	1	20	original file extension
ccf_added_UVER	integer	1	999	[FK: UVER] version in which CCF was added [not visible] This value can never be changed in new versions. ccf_added_UVER > 1: The retrieval software should display a hint for the user.



Table dh_has: header attribute set

field name	type	min	max	field description
has_order	integer	1	99	order of rows on CDROM = UHID [not used] [unique] [not visible]
has_UHID	integer	1	99	[key] unique header ID [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

Table dh_asu: header active substances

field name	type	min	max	field description
asu_order	integer	1	99	order of rows on CDROM = order of active substances [unique] [index] [not visible]
asu_UASU	integer	1	99	[key] unique active substance ID [unique] [index] [not visible]
asu_UHID	integer	1	99	[FK: UHID] header attribute set [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

Table dh_rps: relation product substance

field name	type	min	max	field description
rps_order	integer	1	99	order of rows on CDROM = [not used] [unique] [not visible]
rps_UPRO	integer	1	99	[FK: UPRO] product [index] [not visible] The header ID (UHID) of the referenced product (UPRO) and the referenced active substance (UASU) must be the same.
rps_UASU	integer	1	99	[FK: UASU] active substance [index] [not visible] The header ID (UHID) of the referenced product (UPRO) and the referenced active substance (UASU) must be the same.
rps_concentration	text	1	10	total concentration (with units: e.g. 15 ml/l)



Table dh_pro: header products

field name	type	min	max	field description
pro_order	integer	1	99	order of rows on CDROM = order of products [unique] [index] [not visible]
pro_UPRO	integer	1	99	[key] unique product ID [unique] [index] [not visible]
pro_UHID	integer	1	99	[FK: UHID] header attribute set [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]

Table dh_com: header companies

field name	type	min	max	field description
com_order	integer	1	99	order of rows on CDROM = order of companies [unique] [index] [not visible]
com_UCOM	integer	1	99	[key] unique company ID [not used] [unique] [index] [not visible]
com_UHID	integer	1	99	[FK: UHID] header attribute set [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

Table da_chg: changes history

field name	type	min	max	field description
chg_order	integer	1	99999999	order of rows on CDROM [not used] [unique] [not visible]
chg_type	text	1	1	"P" = Page deleted / "C" = Controlled file deleted / "R" = Report entry deleted / "A" = Attributes of report changed [not visible]
chg_UxID	integer	1	99999999	[FK: UxID] UPID, URID or UFID of deleted/changed item [not visible]
chg_ver_UVER	integer	1	999	[FK: UVER] last version in which record of the item can be found [not visible]



Table dd_tab:	data dictionary tables
----------------------	-------------------------------

field name	type	min	max	field description
tab_order	integer	1	99	order of rows on CDROM = [not used] [unique] [not visible]
tab_name	text	1	8	[key] table/file name [unique] [index] [not visible]
tab_version	integer	1	99	CADDY standard version of table [not visible] tab_version = 2: Table in CADDY 1.1 format tab_version = 3: Table in CADDY 2.0 format Other values are not allowed. Only the tables dh_ver, dc_ccf and dc_lnk are changed in CADDY 2.0 (value = 3). The other tables have only added semantics or changed field names that will not affect the table structure. In future CADDY versions this value will be incremented when the structure of a table is changed.
tab_rowcount	integer	0	99999999	number of rows in table/file [not visible] tab_rowcount = 0: The table contains no rows and the index file is not stored on the volume.

Table dd_col:	data dictionary columns
----------------------	--------------------------------

field name	type	min	max	field description
col_order	integer	1	9999	order of rows on CDROM = order of columns [unique] [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] For col_type = "boolean" or col_type = "date": col_value_min has to contain the value 0
col_value_max	integer	1	99999999	text: max. length / integer: max. value [not visible] For col_type = "boolean" or col_type = "date": col_value_max has to contain the value 1