Technical Report No. 9

The CERA Metadata Model

Heinke Höck • Hannes Thiemann
Michael Lautenschlager • Ingo Jessel
DKRZ - German Climate Compute Centre
Bundesstr. 55, D - 20146 Hamburg

Beate Marx • Manfred Reinke
AWI - Alfred-Wegener Institute
Am Handelshafen 12, D - 27750 Bremerhaven, Germany

Edited by:

Modellbetreuungsgruppe DKRZ
Hamburg, July 1995

ISSN 0940-9327

Table of Contents

General Description

Detailed information about contents and organization within a database system is required. Data description or metadata is often defined as data about data or as information required to make scientific data useful. Following the IEEE reference model, the metadata should support four different interfaces to scientific data management.

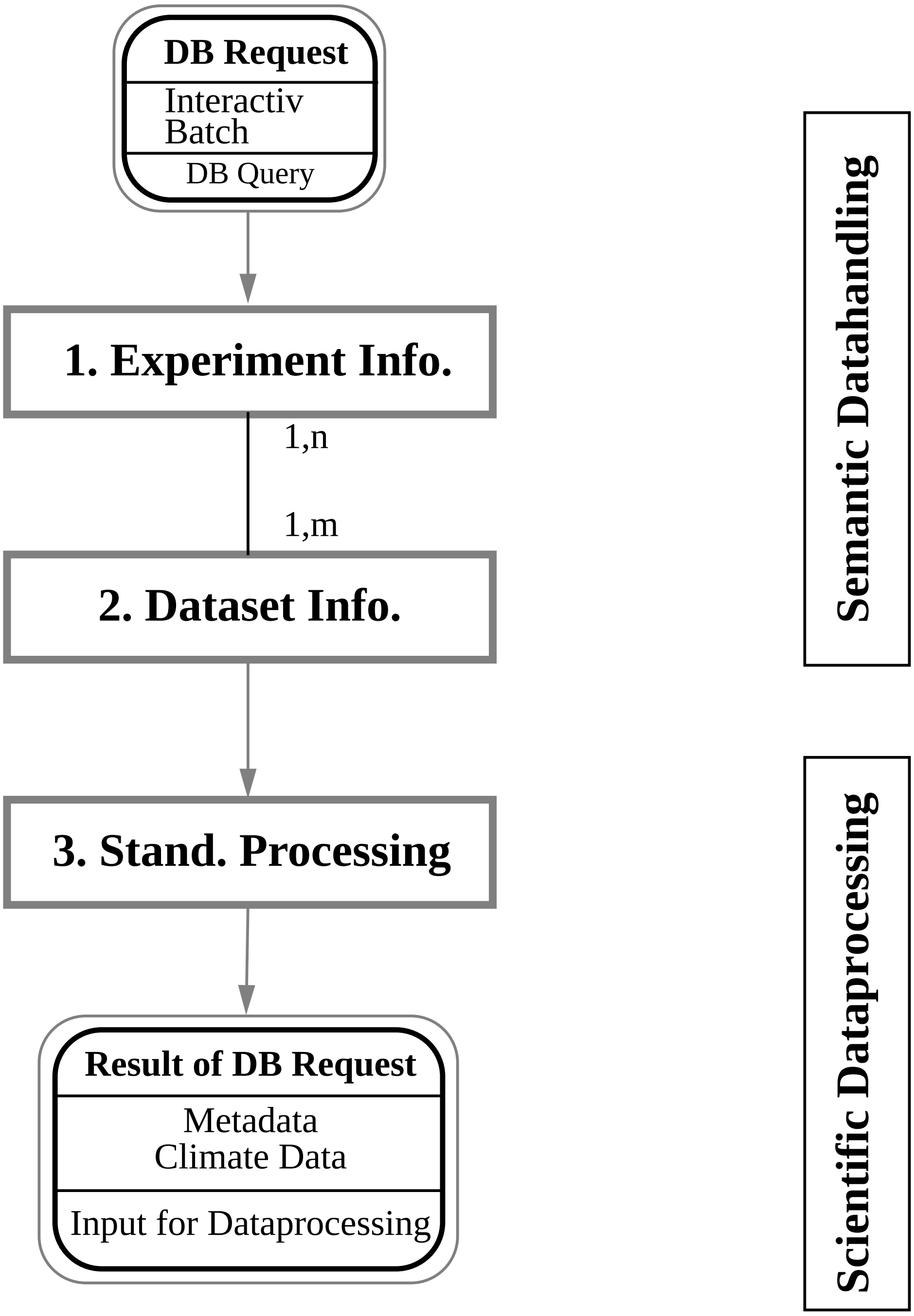
- **Browse, Search and Retrieval**
  This external interface to a database system is related to the needs of users who interact with the database system by, for example, a graphical interface. The usage will mainly be interactive which requires a response time that keeps the user engaged. Typical questions to a database system are: what data exist, is it likely to be of use to me, is it really what I want or how do I get it? The underlying datamodel is mainly dictated by the user requirements.
- **Ingest, Quality Assurance, Reprocessing**
  This interface should ensure the logical and scientific integrity of the database. The reprocessing contributes to the data quality which should be accessible within the metadata.
- **Application to Application Transfer**
  The interface places many demands on the robustness and completeness of the descriptions of the data structure. It requires standardization in terms of the datamodel and the storage formats. The interface contributes to scientific dataprocessing within the database system and to realizations of geographically distributed databases
- **Storage and Archive**
  The interface is driven by the need for efficient search and retrieval implementation with the overall goal of total cost minimization. The user requirements to access the data must be used as a guideline. Efficient climate data access minimizes the requests to slow sequential mass storage systems.

The CERA metadata model tries to incorporate the IEEE reference model ideas. The development was basically guided by the intention to keep the metadata system as simple as possible, but as flexible as necessary to match the user requirements for a climate database system and to incorporate international data description standards like DIF from NASA and INFOCLIMA from the WMO. The mapping between CERA and DIF is realized, the mapping between CERA and INFOKLIMA and CERA and FGDC metadata standard will be incorporated. DIF required information is marked with an asterisk in brackets. The DIF required information is used in CERA as minimum information as well.

Simplicity for the metadata system is required in order to make progress in a complete semantic datahandling. Not only the data retrieval will be organized by the semantic within the metadata system, but also the data production and storage will then be organized based on the defined semantic context. The data retrieval based on the metadata system, i.e. the description of the physical data, is the classical application of climate database systems. The data storage along the semantic information means that climate data are produced within a specified context and the database system itself will organize the file structure for the physical storage in a way suitable for the future data access.

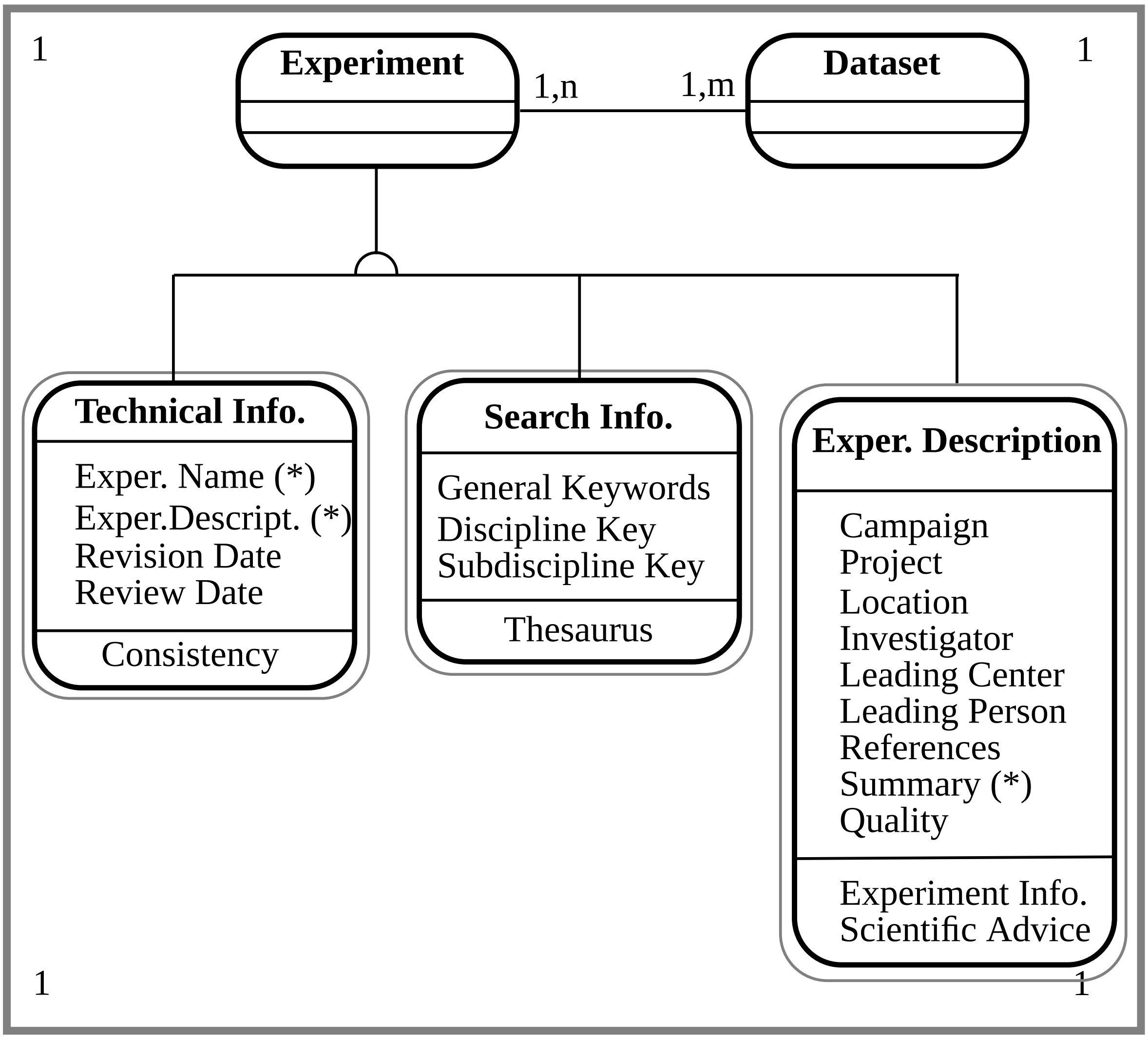
The usage of international standards in (climate) database systems is required for future communication with other database or information systems which are based on these standards. Standardization of the metadata systems and additionally of the data storage formats will in future become more important in terms of geographically distributed databases.

The CERA datamodel was developed following the outlined constraints. The CERA metadata contain information about experiments and additionally about the datasets themselves. In CERA, the experiments are defined as a compilation of datasets. Datasets are defined as compilation of two-dimensional data fields, for example, a time series of global surface temperatures as calculated by a climate model. Time series of other physical quantities like wind or precipitation form different datasets belonging to the same experiment.



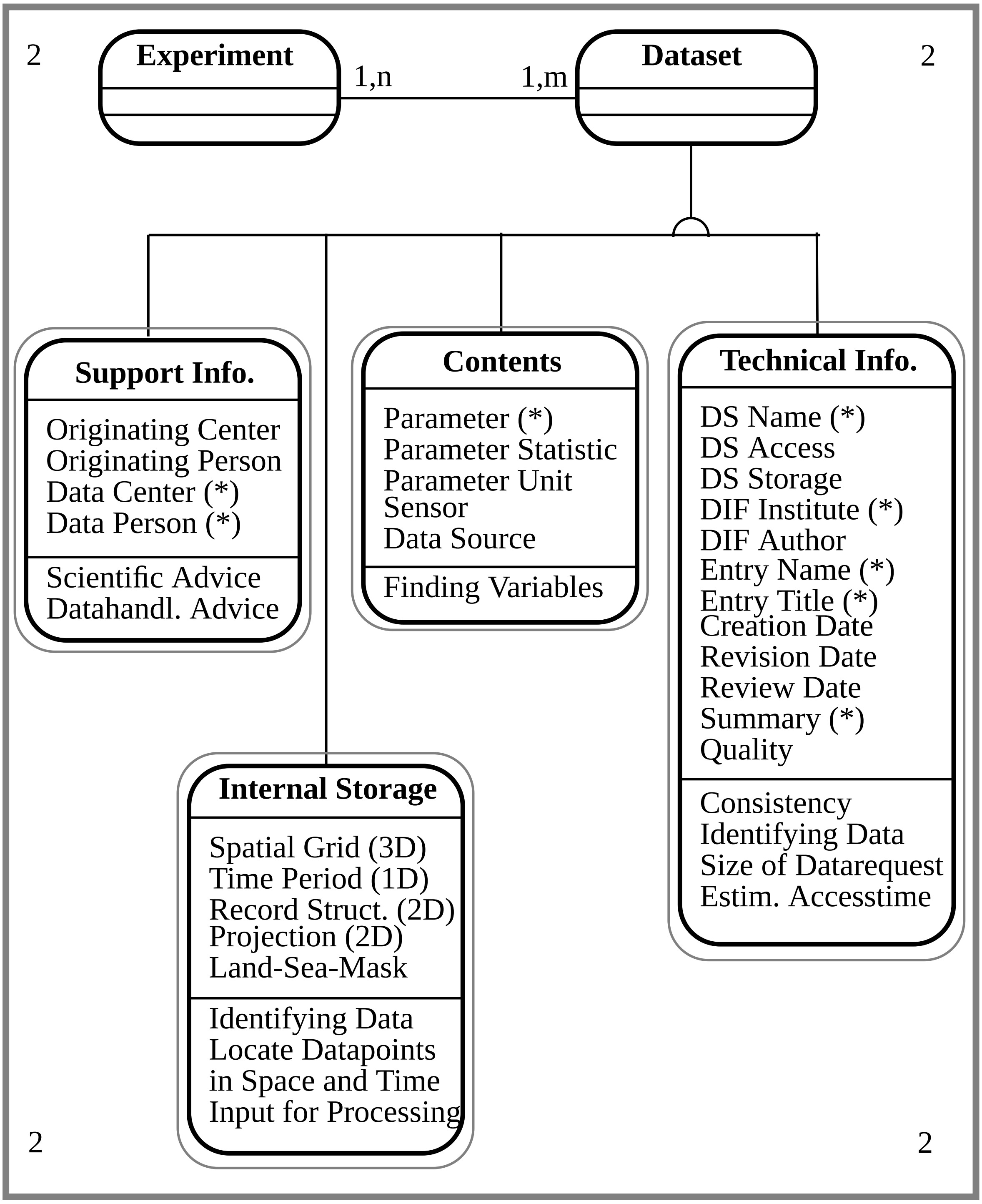
The **Experiment Information** is subdivided into three sections.

- The **Technical Information** contains the name of the experiment, the description of the
- experiment, the revision date, and the review date. This technical information contributes to the consistency of the database.
- The **Search Information** is based on general keywords, a discipline key and a subdiscipline key. This keyword based search allows database users who are not familiar with the overall subject of the stored data to search for terms of interest.
- The **Experiment Description** contains a designation of the campaign, the geographical location of the experiment, the investigator, the centre and person who enforced the experiment, references to get further information, summary and quality containing other details.This description should provide detailed information about the experiment. Additionally the contact for scientific advice on the experiment and the related data is established.



The experiment information contains the overall description of the related data. A detailed description of the climate data can be obtained from the **Dataset Information**. The overall information is split into four categories.

- The **Support Information** contains the originating centre and person who produces the data and the data centre and person who is responsible for the storage and access. The support information establishes the contact for scientific advice, how to use the data and for technical advice on how to handle them.
- The **Contents** information contains the name of the physical parameter stored, the parameter statistics and the physical unit, the sensor used for production, and the data source.The dataset contents information allow for searching for specific physical quantities.
- The **Technical Information** gives the dataset name, access method and storage device, the DIF institute and author who is responsible for the database entries, a special entry name and title which is required by the DIF, a creation date, a revision date and a review date, Additionally text fields containing data quality information and a brief summary are added. These information blocks support database consistency and help identify data. Additionally the, size of requested data and the probable access time can be deduced.
- The **Internal Storage** defines in detail the underlying implied three-dimensional spatial grid, the time period or series which is represented by the data, the coordinates of the stored two-dimensional data records, the definition of the missing two coordinates, and if necessary, the identification of land and sea points. In in a horizontal cross section of a global data field the latitude and longitude define the geographical location of the data points. The location in the four-dimensional dataspace is completed by the vertical level and time. The internal storage information locates the data in the physical space and time. This information is necessary as input for scientific dataprocessing.



The CERA metadata model is suitable to describe climate and environmental data, numerical model data as well as observational data. The metadata system can be extended to incorporate new types of data like satellite data. It matches international standards namely DIF and INFOKLIMA. The CERA system was designed as simply as possible to match the concept of storing data according to the semantic context and as flexibly as necessary to fulfil the requirements of semantic data retrieval and scientific dataprocessing.

Conditions Required for CERA Climate Database

- **Relation between CERA and DIF**

    1. Minimum CERA description is identical to DIF required items.

    2. Complete CERA Description includes complete DIF.

- **Relation between CERA data processing and data description**

    1. No climate data are available, only metadata:
       Only catalogue information can be extracted, no data and no data processing is available.

    2. No BLOB entries in database tables are available, only pointers to (UNIX) data files are realized:
       No data processing is available, only file transport is realized.

    3. BLOB table entries are missing the entire storage description and/or GRIB storage format:
       No data processing is available, only file transport is realized.

    4. BLOB table entries possess the complete storage description and the GRIB storage format:
       CERA dataprocessing is available.

Table Description and Data Examples

This section provides a detailed description of CERA database tables containing the table number consistent with the flow chart, the data element, the data type and the description. The relation to the DIF is indicated printed in italic type in the description part. An example for each table is given.

Table 1. DATASET Description

Data Element	Data_Type	Description
dataset_id	integer*8	Dataset number uniquely identifying the dataset within CERA.
dataset_name	*char*80*	Dataset name uniquely identifying the dataset within CERA. Dataset definition page 55. *DIF required Data_Center (DATASET ID)*
access_id	integer*1	Current status of dataset. Range: 1: Database/permanent 2: Database/Unitree 3: Tape 4: External access with net, host and path 5: Personal contact 1 ,2 ,3: This dataset is stored in the database as blobs in a table. Blob definition page 55. The Name of the table containing the dataset is found in the table table_name identified with dataset_id. 4: This dataset is stored outside of the Database. Name of host and net in table NET, name of path in table PATH, identified with dataset_id in table EXTERNAL_ACCESS.
entry_id	integer*4	Identifier for Entry_name and entry_title found in table ENTRY. *DIF required Entry_name and Entry_title.*
format	char*32	Dataset format e.g. Grib.
creation_date	date	Date of dataset creation.
revision_date	date	Date of last update of dataset. *DIF optional Revision_Date.*
review_date	date	Date of the latest review for accuracy of scientific content. *DIF optional Science_Review_Date.*
future_review_date	date	Future time at which the DIF should be reviewed for accuracy of scientific or technical content. *DIF optional Future_Review_Date.*
person_id	integer*4	Person identifier found in table PERSON writing the CERA and is responsible for the accuracy of the information. *DIF optional Author.*
institute_id	integer*2	Identifier for the data center found in table INSTITUTE that provided the information contained in the DIF. *DIF required Originating_Center.*
quality_id	integer*4	Quality identifier found in table QUALITY. *DIF optional Quality.*
summary_id	integer*4	Summary identifier found in table SUMMARY. *DIF required Summary.*

Table 1. DATASET Content Example

<u>Data Element</u>	
dataset_id	2
dataset_name	CTRLMONMEA.108
access_id	4
entry_id	1
format	grib
creation_date	20.01.1992
revision_date	20.01.1992
review_date	20.01.1992
future_review_date	NULL
person_id	5
institute_id	1
quality_id	1
summary	1

Table 2. EXTERNAL_ACCESS Description

Data Element	Data_Type	Description
dataset_id	integer*8	Dataset identifier found in table DATASET if the dataset is not stored in CERA. access_id=4 in table DATASET.
path_id	integer*4	Path identifier found in table PATH.
net_id	integer*2	Identifier found in table NET for net type and host name.
file_id	integer*4	Filename identifier found in table FILENA-MES.
access_time	integer*4	Estimated access time in seconds for the data-set depending on access_id in table DATASET.

Table 3. FILENAMES Description

Data Element	Data_Type	Description
file_id	integer*4	Filename number uniquely identifying the file_name.
file_name	char*80	File name.

Table 4. NET Description

Data Element	Data_Type	Description
net_id	integer*2	Net number uniquely identifying net_type and host_name found in table EXTERNAL_ ACCESS for dataset found in table DATASET identified with dataset_id in table EXTERNAL_ACCESS.
net_type	char*64	Nettype of the dataset identified in table EXTERNAL ACCESS i.e. internet/ftp
host_name	char*80	Internet address ... of the dataset identified in table EXTERNAL_ACCESS.

Table 5. PATH Description

Data Element	Data_Type	Description
path_id	integer*4	Path number uniquely identifying the path found in EXTERNAL_ACCESS.
path	char*256	Path name of the dataset identified in table EXTERNAL_ACCESS.

Table 2. EXTERNAL_ACCESS Content Example

Data Element	
dataset_id	2
path_id	1
net_id	1
file_id	1
access_time	60

Table 3. FILENAMES Content Example

Data Element	
file_id	1
file_name	FILE1

Table 4. NET Content Example

Data Element	
net_id	1
net_type	internet/ftp
host_name	136.172.38.10

Table 5. PATH Content Example

Data Element	
path_id	1
path	/model/ctrl/atmosphere

Table 6. INTERNAL_ACCESS Description

Data Element	Data_Type	Description
dataset_id	integer*8	Dataset identifier found in table DATASET if the dataset is stored in CERA. access_id=1,2 in table DATASET.
table_name	char*128	Name of table containing the blob identification.
access_time	integer*4	Estimated access time in seconds for the dataset depending on access_id in table DATASET.

Table table_name Description

Data Element	Data_Type	Description
dataset_id	integer*8	Identifier of dataset found in table DATASET
start_year	integer*4	YYYYYY... see PERIOD
start_month	integer*1	MM
start_day	integer*1	DD
start_hour	integer*1	HH
start_second	integer*2	SSSS
blob_id	integer*4	Blob Identifier found in table blob_table_name.
blob_size	integer*8	Size of blob in bytes.
blob_table_name	char*128	Name of table containing the blobs identified with blob_id and of the dataset identified with dataset_id. One dataset can be splitted over more than one table.

Table blob_table_name Description

Data Element	Data_Type	Description
blob_id	integer*4	Blob number uniquely identifying the blob
blob	blob	Grid values

Table 6. INTERNAL_ACCESS Content Example

Data Element	
dataset_id	1
table_name	DATA_1
access_time	40

Table DATA_1 Content Example

Data Element				
dataset_id	1	1	1	1
start_year	1935	1935	1935	1935
start_month	1	2	3	4
start_day	1	1	1	1
start_hour	0	0	0	0
start_second	0	0	0	0
blob_id	1	2	3	4
blob_size	65536	65536	65536	65536
blob_table_name	DATA_1_1	DATA_1_1	DATA_1_2	DATA_1_2

Table DATA_1_1 Content Example

Data Element		
blob_id	1	2
blob	....	....

Table DATA_1_2 Content Example

Data Element		
blob_id	3	4
blob	....	....

Table 7. ORIGINATING_CENTER Description

Data Element	Data_Type	Description
institute_id	integer*2	Institute identifier found in table INSTITUTE where the dataset is created.
dataset_id	integer*8	Dataset identifier found in table DATASET.
person_id	integer*4	Person identifier found in table PERSON creating the datasets on a computer or with measurements.

Table 8. DATA_CENTER Description

Data Element	Data_Type	Description
institute_id	integer*2	Institute identifier found in table INSTITUTE holding the data, data set identification. *DIF required Data_Center (DATA CENTER NAME).*
dataset_id	integer*8	Dataset identifier found in table DATASET
person_id	integer*4	Person identifier found in table PERSON. This Person should be able to supply information about data storage, access, contents, quality, availability, costs of the data. *DIF required Data_Center(DATA CENTRE_CONTACT).*

Table 9. TECHNICAL_CONTACT Description

Data Element	Data_Type	Description
person_id	integer*4	Identifier for the person who is knowledgeable about the technical content of the data (quality, processing methods, units, available software for further processing)*DIF optional TECHNICAL CONTACT.*
dataset_id	integer*8	Dataset identifier found in table DATASET.

Table 7. ORIGINATING_CENTER Content Example

Data Element	
institute_id	1
dataset_id	2
person_id	3

Table 8. DATA_CENTER Content Example

Data Element	
institute_id	1
dataset_id	2
person_id	3

Table 9. TECHNICAL_CONTACT Content Example

Data Element	
person_id	3
dataset_id	2

Table 10. STORAGE Description

Data Element	Data_Type	Description
dataset_id	integer*8	Dataset identifier found in table DATASET.
medium_id	integer*4	Medium identifier found in table MEDIUM giving the type of storage medium of the data-set. *DIF optional together with size Storage_Medium.*
size	integer*8	Size of dataset in Bytes. *DIF optional together with medium in table MEDIUM Storage_Medium.*

Table 11. MEDIUM Description

Data Element	Data_Type	Description
medium_id	integer*2	Medium number uniquely identifying the medium found in table STORAGE.
medium	*char*80*	Description type of storage medium of the dataset found in table STORAGE. *DIF optional together with size Storage_Medium.*

Table 12. ENTRY Description

Data Element	Data_Type	Description
entry_id	integer*4	Entry number uniquely identifying Entry_name and Entry_title found in table DATASET. *DIF required Entry_ID and Entry_Title.*
entry_name	*char*31*	Unique identifier of the DIF for the Dataset, a collection of Datasets or part of a Dataset. *DIF required. Entry_ID.*
entry_title	*char*160*	Title of the DIF should convey DIF content i.e. instrument, investigator, mission, parameters measured. *DIF required Entry_Title.*

Table 10. STORAGE Content Example

Data Element	
dataset_id	1
medium_id	1
size	20231

Table 11. MEDIUM Content Example

Data Element	
medium_id	1
medium	magnetic tape

Table 12. ENTRY Content Example

Data Element	
entry_id	1
entry_name	PRE_INDUSTRIELLE
entry_title	150 years run of coupled t21 echam1 with lsg ocean with flux correction with increasing CO2 equivalent.

Table 13. PARAMETER_CONNECT Description

Data Element	Data_Type	Description
dataset_id	integer*8	Dataset identifier found in table DATASET
parameter_id	integer*4	Identifier for parameter group, parameter term, parameter type, unit, statistics identifier found in table PARAMETER. *DIF required Parameter(PARAMETER GROUP). DIF required Parameter(PARAMETER TERM).*
sensor_id	integer*4	Sensor_name identifier found in table SENSOR. *DIF optional Sensor_name.*
source_id	integer*4	Source_name identifier found in table SOURCE. *DIF optional Source_name.*
projection_id	integer*4	The projection_id is the identifier for the description of the position, order and position regularity of the values of a dataset. With this description it is possible to identify the position of every value in a blob.
record_structure_id	integer*2	Record_structure identifier found in table RECORD_STRUCTURE.

Table 14. SOURCE Description

Data Element	Data_Type	Description
source_id	integer*4	Source number uniquely identifying source_name, source_acronym, source_type found in table PARAMETER_CONNECT.
source_name	*char*80*	Spacecraft, platform, ship, ground station, telescope, computer center etc. housing the sensor(s) and computer(s) used to acquire the data. *DIF optional Source_name(LONG NAME).*
source_acronym	*char*31*	Abbreviated version of source_name. *DIF optional Source_name(SHORT NAME).*
source_type	char*80	Description of the source type.

Table 15. SENSOR Description

Data Element	Data_Type	Description
sensor_id	integer*4	Sensor number uniquely identifying sensor_name, sensor_acronym, sensor_type found in table PARAMETER_CONNECT.
sensor_name	*char*80*	Instrument or hardware used to acquire the data. *DIF optional Sensor_name(LONG NAME).*
sensor_acronym	*char*31*	Abbreviated version of sensor_name. *DIF optional Sensor_name(SHORT NAME).*
sensor_type	char*80	Description of the sensor type.

Table 13. PARAMETER_CONNECT Content Example

Data Element	
dataset_id	1
parameter_id	1
sensor_id	1
source_id	1
projection_id	1
record_structure_id	1

Table 14. SOURCE Content Example

Data Element	
source_id	1
source_name	Deutsches Klimarechenzentrum GmbH
source_acronym	DKRZ
source_type	Computing Center

Table 15. SENSOR Content Example

Data Element	
sensor_id	1
sensor_name	Cray II
sensor_acronym	C2
sensor_type	Computer

Table 16. PARAMETER Description

Data Element	Data_Type	Description
parameter_id	integer*4	Parameter number uniquely identifying the table records of table PARAMETER found in PARAMETER CONNECT.
group_id	integer*4	Group_name identifier found in table PARAMETER_GROUP. *DIF required Parameter(PARAMETER GROUP).*
term_id	integer*4	Term_name identifier found in table PARAMETER_TERM. *DIF required Parameter(PARAMETER).*
type_id	integer*4	If the Parameter is DIF independent.
unit_id	integer*4	Unit_name identifier found in table PARAMETER_UNIT.
statistics_id	integer*4	Statistics_name identifier found in table PARAMETER_STATISTICS.

Table 17. PARAMETER_GROUP Description

Data Element	Data_Type	Description
group_id	integer*4	Group number uniquely identifying the group_name.
group_name	char*80	Kinds of measurements represented by the data. *A DIF keyword list exists (e.g. ocean dynamics). DIF required Parameter(PARAMETER_GROUP).*

Table 18. PARAMETER_TERM Description

Data Element	Data_Type	Description
term_id	integer*4	Term number uniquely identifying the term_name together with the group_id.
term_name	char*80	Kinds of measurements represented by the data. *A DIF keyword list exists (e.g. sea level). DIF required Parameter(PARAMETER).*
term_acronym	char*31	Abbreviated version of term_name.
group_id	integer*4	Parameter group identifier found in table PARAMETER_GROUP. One term_name can belong to different parameter groups.

Table 16. PARAMETER Content Example

Data Element	
parameter_id	1
group_id	1
term_id	1
type_id	NULL
unit_id	1
statistics_id	1

Table 17. PARAMETER_GROUP Content Example

Data Element	
group_id	1
group_name	Atmospheric Dynamics

Table 18. PARAMETER_TERM Content Example

Data Element	
term_id	1
term_name	Atmospheric Temperature
term_acronym	NULL
group_id	1

Table 19. PARAMETER_TYPE Description

Data Element	Data_Type	Description
type_id	integer*4	Type number uniquely identifying the type_name.
type_name	char*80	DIF independent description of the parameter.
type_acronym	char*31	Abbreviated version of type_name.

Table 20. PARAMETER_UNIT Description

Data Element	Data_Type	Description
unit_id	integer*4	Unit number uniquely identifying the unit_name
unit_name	char*80	Unit of parameter found in table PARAMETER.
unit_acronym	char*31	Abbreviated version of unit_name
unit_description	char*250	Definition of units an measured physical quantities.

Table 21. PARAMETER_STATISTICS Description

Data Element	Data_Type	Description
statistics_id	integer*4	Statistics number uniquely identifying the statistics_name.
statisctics_name	char*80	Statistics description.i.e. Annual mean, monthly mean, global average annual mean, 10 years mean...
statistics_acronym	char*31	Abbreviated version of statistics_name.

Table 19. PARAMETER_TYPE Content Example

Data Element	
type_id	1
type_name	ECHAM code 130
type_acronym	130

Table 20. PARAMETER_UNIT Content Example

Data Element	
unit_id	1
unit_name	Kelvin
unit_acronym	K
unit_description	Absolute temperature

Table 21. PARAMETER_STATISTICS Content Example

Data Element	
statistics_id	1
statisctics_name	monthly mean
statistics_acronym	NULL

Table 22. PERIOD_CONNECT Description

Data Element	Data_Type	Description
dataset_id	integer*8	Dataset identifier found in table DATASET. With this table an attachment between n not connected periods and m datasets is possible. i.e. missing data. One period can exist of one point of time. It is possible to describe irregular points of time in a timeseries.
period_id	integer*4	Period identifier for start_<time>, timestep, period_len found in table PERIOD. *DIF optional Start_Date, Stop_Date.*

Table 23. PERIOD Description

Data Element	Data_Type	Description
period_id	integer*4	Period number uniquely identifying the period found in PERIOD_CONNECT. No period_order_id necessary, assuming ascending order. But accelerate the access. The start_<time> gives the date the data begin and the timestep with period_len gives the date the data end. For simulation data the data type date is not sufficient because the timeperiod can be thousands of years and i.e. the early jurassic starts 180 mio. years before present.
start_year	integer*4	YYYYYY... *DIF optional part of Start_Date.*
start_month	integer*1	MM *DIF optional part of Start_Date.*
start_day	integer*1	DD *DIF optional part of Start_Date.*
start_hour	integer*1	HH *DIF optional part of Start_Date.*
start_second	integer*2	SSSS *DIF optional part of Start_Date.*
step_year	integer*4	YYYYYY...
step_month	integer*1	MM
step_day	integer*1	DD
step_hour	integer*1	HH
step_second	integer*2	SSSS
period_len	integer*4	Number of points of time. Gives together with <time>_step *DIF optional Stop_Date.*
time_id	integer*4	Identifier for time reference and time coordinate definition.

Table 22. PERIOD_CONNECT Content Example

Data Element	
dataset_id	1
period_id	1

Table 23. PERIOD Content Example

Data Element	
period_id	1
start_year	1935
start_month	1
start_day	1
start_hour	0
start_seconds	0
step_year	0
step_month	1
step_day	0
step_hour	0
step_seconds	0
period_len	1800
time_id	1

Table 24. UTC_REFERENCE_TIME Description

Data Element	Data_Type	Description
time_id	integer*4	Time identifier uniquely identifying the reference time and the time coodinate system.
reference_time_id	integer*4	Reference time identifier
coordinate_def_id	integer*2	Time coordinate system identifier. If coordinate_def_id=1 then real time else look in table TIME_COORDINATE.
description	char*256	Description of time.

Table 25. REFERENCE_TIME Description

Data Element	Data_Type	Description
reference_time_id	integer*4	Reference time identifier uniquely identifying the reference time.
dif_year	integer*4	Difference to UTC reference time in years.
dif_month	integer*1	Difference to UTC reference time in months.
dif_day	integer*1	Difference to UTC reference time in days.
dif_hour	integer*1	Difference to UTC reference time in hours.
dif_second	integer*2	Difference to UTC reference time in seconds.

Table 24. UTC_REFERENCE_TIME Content Example

Data Element	
time_id	1
reference_time_id	1
coordinate_def_id	1
description	

Table 25. REFERENCE_TIME Content Example

Data Element	
reference_time_id	1
dif_year	0
dif_month	0
dif_day	0
dif_hour	0
dif_seconds	0

Table 26. TIME_COORDINATE Description

Data Element	Data_Type	Description
coordinate_def_id	integer*2	Time coordinate number uniquely identifying the time coordinate system. Description of time coordinate system.
months_per_year	integer*2	Modell time...
days_per_january	integer*1	
days_per_february	integer*1	
days_per_march	integer*1	
days_per_april	integer*1	
days_per_may	integer*1	
days_per_june	integer*1	
days_per_july	integer*1	
days_per_august	integer*1	
days_per_september	integer*1	
days_per_october	integer*1	
days_per_november	integer*1	
days_per_december	integer*1	
hours_per_day	integer*1	
seconds_per_hour	integer*2	
description	char*256	Description of modell time.

Table 26. TIME_COORDINATE Content Example

Data Element	
coordinate_def_id	1
months_per_year	12
days_per_january	30
days_per_february	30
days_per_march	30
days_per_april	30
days_per_may	30
days_per_june	30
days_per_july	30
days_per_august	30
days_per_september	30
days_per_october	30
days_per_november	30
days_per_december	30
hours_per_day	24
seconds_per_hour	3600
description	ideal earth time (360*86400s days per year)

Table 27. RECORD_STRUCTURE Description

Data Element	Data_Type	Description
record_structure_id	integer*2	record_structure number uniquely identifying the record structure found in table PARAMETER_CONNECT.
coordinate1_id	integer*1	Coordinate identifier for the first coordinate found in table COORDINATE_SPECIFICATION. Range: 1: no specification 2: z 3: x 4: y
coordinate2_id	integer*1	Coordinate identifier for the second coordinate found in table COORDINATE_SPECIFICATION. Range: 1: no specification 2: z 3: x 4: y
dependence_id	integer*2	Dependence identifier found in table COORDINATE_DEPENDENCE. Dependence of the third dimension in space (x,y,z) if coordinate 1 and coordinate 2 are x, y or z.
structure_description	char*80	These are examples of the record_structure_id and the structure_description: 1: horizontal field first vary x then y 2: horizontal field first vary y then x 3: hovmoeller first vary t then y 4: hovmoeller first vary y then t 5: section cut first vary y then z 6: section cut first vary z then y 7: (reserved for future use) 8: section cut first vary x then z 9: section cut first vary z then x 10: timeseries. 11: cross grid section first vary y then z then x. 12: cross grid section first vary z then y then x.

Table 27. RECORD_STRUCTURE Content Example

Data Element	
record_structure_id	1
coordinate1_id	3
coordinate2_id	4
dependence_id	0
structure_description	horizontal field first vary x then y

Table 28. COORDINATE_SPECIFICATION Description

Data Element	Data_Type	Description
coordinate_id	integer*1	Coordinate number uniquely identifying the coordinate for one direction
coordinate_name	char*31	coordinate name. Range: 1: no specification 2: z 3: x 4: y

Table 29. COORDINATE_DEPENDENCE Description

Data Element	Data_Type	Description
dependence_id	integer*2	Dependence number uniquely identifying the dependence of the third dimension in space (x,y,z) if coordinate 1 and coordinate 2 x,y or z.
dependence_definition	char*80	Description of dependence. Range: 0: no dependence 1: dependence to coordinate 1 2: dependence to coordinate 2 3: dependence to coordinate 1 and 2.

Table 30. GRID Description

Data Element	Data_Type	Description
projection_id	integer*4	Grid description number uniquely identifying the Grid for the datasets identified in table PARAMETER_CONNECT.
grid_id1	integer*4	Identifier for the description of position, order and position regularity of the first coordinate.
grid_id2	integer*4	Identifier for the description of position, order and position regularity of the second coordinate.
grid_id3	integer*4	Identifier for the description of position, order and position regularity of the third coordinate.
mask_id	integer*4	Identifier for the land-sea mask found in table MASK.

Table 30. GRID Content Example

Data Element	
projection_id	1
grid_id1	1
grid_id2	2
grid_id3	1

Table 31. MASK Description

Data Element	Data_Type	Description
mask_id	integer*4	Mask number uniquley identifying the mask for the member of a projection group.
position1_id	integer*8	Position of the first coordinate.
position2_id	integer*8	Position of the second coordinate.
land_sea_flag	integer*1	Flag for land or sea. Range: 1: sea 0: land

Table 33. GRID_VALUES Description

Data Element	Data_Type	Description
grid_id	integer*4	Identifier for grid found in table GRID.
position	integer*8	Position of value in grid. Together with grid_id uniquely identification of grid coordinate with value found in this table.
value	real*4 IEEE	Grid coordinate value. Unit of value found in table GRID_UNITS. *DIF optional Coverage(Southernmost_Latitude Northermost_Latitude Westernmost_Longitude Easternmost_Longitude) found with position=1 and postion=grid_len from table GRID_UNITS.*

Table 33. GRID_DESCRIPTION Description

Data Element	Data_Type	Description
grid_id	integer*4	Identifier for grid found in table GRID.
increment	real*4 IEEE	If the grid is regular value of increment.
grid_unit	char*80	Unit of grid value found in table GRID_VALUES identified with grid_id.
grid_len	integer*8	Number of points needed for grid description for one coordinate. i.e. if regular number of latitudes.
grid_type_name	character*80	Description of grid type. Range: 1: regular 2: staggered regular rec 3: staggered regular sal 4: staggered irregular rec 5: staggered irregular sal 6: irregular rectangular 7: irregular not rectangular 8: spectral

Table 32. GRID_VALUES Content Example

Data Element	
grid_id	1
position	1
value	0

Table 33. GRID_DESCRIPTION Content Example

Data Element	
grid_id	2
increment	5
grid_unit	degrees
grid_len	72
grid_type_name	regular

Table 34. DATASET_QUALITY Description

Data Element	Data_Type	Description
quality_id	integer*8	Quality number uniquely identifying the quality.
quality	char*2000	Information about any quality procedures followed in producing the data described in the DIF. *DIF optional Quality.*

Table 35. DATASET_SUMMARY Description

Data Element	Data_Type	Description
summary_id	integer*8	Summary number uniquely identifying the summary.
summary	char*2000 *500 words*	Contains an abstract and information not contained in other fields. Access routines if the format is not grib. Order of 5-dim room if not grib. Quality. Description of time i.e.real earth time, ideal earth time. *DIF required Summary.*

Table 34. DATASET_QUALITY Content Example

Data Element	
quality_id	1
quality	NULL

Table 35. DATASET_SUMMARY Content Example

Data Element	
summary_id	1
summary	Coupled ocean atmosphere run with flux correction. Ocean lsg and atmosphere echam1 t21 resolution with constant CO2 equivalent from 1985.

Table 36. EXPERIMENT_CONNECT Description

Data Element	Data_Type	Description
experiment_id	integer*4	Experiment identifier found in table EXPERIMENT. A lot of datasets can belong to one experiment and in one dataset data from a lot of experments can be collected together i.e. global fields for observation data.
dataset_id	integer*8	Dataset identifier found in table DATASET.

Table 37. EXPERIMENT Description

Data Element	Data_Type	Description
experiment_id	integer*4	Experiment number uniquely identifying the experiment i.e. Szenario A. If the datasets of this experiment have no ENTRY, experiment_id is identifying *DIF required Entry_ID, Entry_Title, Revision_Date, Science_Review_Date, Summary.*
experiment_name	*char*31*	An experiment name or unique identifier of the DIF for the collection of Datasets. *DIF required Entry_ID.*
experiment_description	*char*160*	Experiment description or title of the DIF should convey DIF content i.e. instrument, investigator, mission, parameters measured. *DIF required Entry_Title.*
revision_date	date	Date of last update of datasets. *DIF optional Revision_Date.*
review_date	date	Date of the latest review for accuracy of scientific content. *DIF optional Science_Review_Date.*
future_review_date	date	Future time at which the DIF should be reviewed for accuracy of scientific or technical content. *DIF optional Future_Review_Date.*
quality_id	integer*4	Identifier for quality found in table EXPERIMENT_QUALITY. *DIF optional Quality.*
summary_id	integer*4	Identifier for summary found in table EXPERIMENT_SUMMARY. *DIF required Summary.*

Table 39. GENERAL_KEY Description

Data Element	Data_Type	Description
general_keywords	*char*31*	Any words or phrases used to further describe data sets. *DIF optional General_Keywords.*
experiment_id	integer*4	Identifier for experiment_name found in table EXPERIMENT.

Table 37. EXPERIMENT Content Example

Data Element	
experiment_id	1
experiment_name	ControlT21
experiment_description	lsg+echam t21 control
revision_date	10.02.1992
review_date	10.02.1992
future_review_date	NULL
quality	1
summary	1

Table 38. GENERAL_KEY Content Example

Data Element	
general_keywords	Greenhouse Warming
experiment_id	1

Table 39. DISCIPLINE_CONNECT Description

Data Element	Data_Type	Description
experiment_id	integer*4	Experiment identifier found in table EXPERIMENT.
discipline_key_id	integer*2	Discipline identifier found in table DISCIPLINE_KEY. *DIF optional Discipline.*

Table 40. DISCIPLINE_KEY Description

Data Element	Data_Type	Description
discipline_key_id	integer*2	Discipline number uniquely identifying the discipline in DIF.
discipline_name	char*80	Describes the science discipline in which the data are normally used. *DIF optional Discipline must be selected from a valid keyword list.*
subdiscipline_id	integer*2	Subdiscipline identifier found in table SUBDISCPLINE_KEY. *DIF optional Discipline(SUBDISCIPLINE) must be selected from a valid keyword list.*

Table 41. SUBDISCIPLINE_KEY Description

Data Element	Data_Type	Description
subdiscipline_id	integer*2	Subdiscipline number uniquely identifying the subdiscipline in DIF.
subdiscipline_name	char*80	Describes the science subdiscipline in which the data are normally used. *DIF optional Discipline(SUBDISCIPLINE) must be selected from a valid keyword list.*
add_term_id	integer*2	Add_term identifier found in table DISCPLINE_ADD_TERM_KEY. *DIF optional Discipline(Finer Breakdown).*

Table 42. DESCIPLINE_ADD_TERM_KEY Description

Data Element	Data_Type	Description
add_term_id	integer*2	Add_term number uniquely identifying the finer breakdowns of *DIF optional Discipline(Finer Breakdown).*
term_text	char*80	Describes the science finer breakdown in which the data are normally used. *DIF optional Discipline(Finer Breakdown) must be selected from a valid keyword list.*

Table 39. DISCIPLINE_CONNECT Content Example

Data Element	
experiment_id	1
discipline_key_id	1

Table 40. DISCIPLINE_KEY Content Example

Data Element	
discipline_key_id	1
discipline_name	Earth Science
subdiscipline_id	1

Table 41. SUBDISCIPLINE_KEY Content Example

Data Element	
subdiscipline_id	1
subdiscipline_name	Atmosphere
add_term_id	1

Table 42. DISCIPLINE_ADD_TERM_KEY Content Example

Data Element	
add_term_id	1
term_text	NULL

Table 43. CAMPAIGN Description

Data Element	Data_Type	Description
project_id	integer*4	Identifier for project found in table PROJECT. *DIF optional Project.*
experiment_id	integer*4	Identifier for experiment found in table EXPERIMENT.

Table 44. PROJECT Description

Data Element	Data_Type	Description
project_id	integer*4	Project number uniquely identifying the project. *DIF optional Project.*
project_name	*char*80*	Scientific endeavour encompassing data from a number of data sources. *DIF optional Project(LONG_NAME).*
project_acronym	*char*31*	Abbreviated version of project_name. *DIF optional Project(SHORT_NAME).*
project_description	char*2000	Description of the Project.

Table 45. LOCATION Description

Data Element	Data_Type	Description
location_id	integer*4	Identifier for experiment found in table LOCATION_KEY. *DIF optional Location.*
experiment_id	integer*4	Identifier for experiment found in table EXPERIMENT.

Table 46. LOCATION_KEY Description

Data Element	Data_Type	Description
location_id	integer*4	Location number uniquely identifying the location. *DIF optional Location.*
location_name	char*80	Names of places which may be used as search parameters. *Must be selected from a DIF keywords list. DIF optional Location.*

Table 43. CAMPAIGN Content Example

Data Element	
project_id	1
experiment_id	1

Table 44. PROJECT Content Example

Data Element	
project_id	1
project_name	Greenhouse warming
project_acronym	NULL
project_description	NULL

Table 45. LOCATION Content Example

Data Element	
location_id	1
experiment_id	1

Table 46. LOCATION_KEY Content Example

Data Element	
location_id	1
location_name	global

Table 47. INVESTIGATOR Description

Data Element	Data_Type	Description
experiment_id	integer*4	Identifier for experiment found in table EXPERIMENT.
person_id	integer*4	Person who headed the investigation or experiment that resulted in the acquisition of the data described. In table EMPLOYMENT and INSTITUTE the investigator address is accessible. *DIF optional Investigator.*

Table 48. LEADING_CENTER Description

Data Element	Data_Type	Description
institute_id	integer*2	Identifier for the enforcing center found in table INSTITUTE.
experiment_id	integer*4	Identifier for experiment found in table EXPERIMENT.
person_id	integer*4	Identifier for the enforcing person found in table PERSON. This person is responsible for the scientific company of the experiment.

Table 49. REFERENCES Description

Data Element	Data_Type	Description
experiment_id	integer*4	Identifier for experiment found in table EXPERIMENT.
reference_id	integer*4	Identifier for bibliographic reference found in table TITLES and PUBLICATIONS. *DIF optional References.*

Table 50. TITLES Description

Data Element	Data_Type	Description
reference_id	integer*4	Reference number uniquely identifying the title.
title	char*2000	Title and optional key bibliographic reference. *DIF optional References.*

Table 51. PUBLICTIONS Description

Data Element	Data_Type	Description
reference_id	integer*4	Identifier for reference found in table TITLES
person_id	integer*4	Author of the referenced title. *DIF optional References.*

Table 47. INVESTIGATOR Content Example

Data Element	
experiment_id	1
person_id	1

Table 48. LEADING_CENTER Content Example

Data Element	
institute_id	2
experiment_id	1
person_id	3

Table 49. REFERENCE Content Example

Data Element	
experiment_id	1
reference_id	1

Table 50. TITLES Content Example

Data Element	
reference_id	1
title	Time-depending greenhouse Warming Computations with a Coupled Ocean-Atmosphere Model. Climate Dynamics 8, 55-69

Table 52. PUBLICATION Content Example

Data Element	
reference_id	1
person_id	3

Table 52. INSTITUTE Description

Data Element	Data_Type	Description
institute_id	integer*2	Institute number uniquely identifying the institute in CERA.
institute	*char*80*	Name of the institute. *DIF.*
institute_acronym	*char*31*	Abbreviated version of the institute name. *DIF.*
country	*char*80*	Name of the country (address). *DIF.*
place	*char*80*	Name of the place (address). *DIF.*
street	*char*80*	Name of the street (address). *DIF.*
pobox	*char*80*	Pobox (address). *DIF.*
additional_info	char*400	Additional address informtion.

Table 53. PERSON Description

Data Element	Data_Type	Description
person_id	integer*4	Person number uniquely identifying the person in CERA.
first_name	*char*31*+49	First name. *DIF.*
second_name	*char*31*+49	Second name. *DIF.*
last_name	*char*31*+49	Last name. *DIF.*

Table 54. EMPLOYMENT Description

Data Element	Data_Type	Description
institute_id	integer*2	Institute identifier of the current institute of the person identified with person_id. This is important if the person has changed the institute after he was for example the enforcing person of an experiment where the old institute is the enforcing institute.
person_id	integer*4	Person identifier found in table PERSON.
email	*char*31*+49	email address. *DIF.*
email_type	*char*31*	Network name. *DIF.*
telephone	*char*80*	Phone. *DIF.*
FAX	*char*80*	FAX. *DIF.*

Table 52. INSTITUTE Content Example

Data Element	
institute_id	1
institute	Deutsches Klimarechenzentrum GmbH
institute_acronym	DKRZ
country	Germany
place	20146 Hamburg
street	Bundesstr. 55
pobox	NULL

Table 53. PERSON Content Example

Data Element	
person_id	3
first_name	Ulrich
second_name	NULL
last_name	Cubasch

Table 54. EMPLOYMENT Content Example

Data Element	
institute_id	1
person_id	3
email	cubasch@dkrz.d400.de
email_type	NULL
telephone	NULL
FAX	NULL

Table 55. EXPERIMENT_QUALITY Description

Data Element	Data_Type	Description
quality_id	integer*8	Qualtiy number uniquely identifying the quality of an experiment.
quality	char*2000	Information about any quality procedures followed in producing the data described in the DIF. *DIF optional Quality.*

Table 56. EXPERIMENT_SUMMARY Description

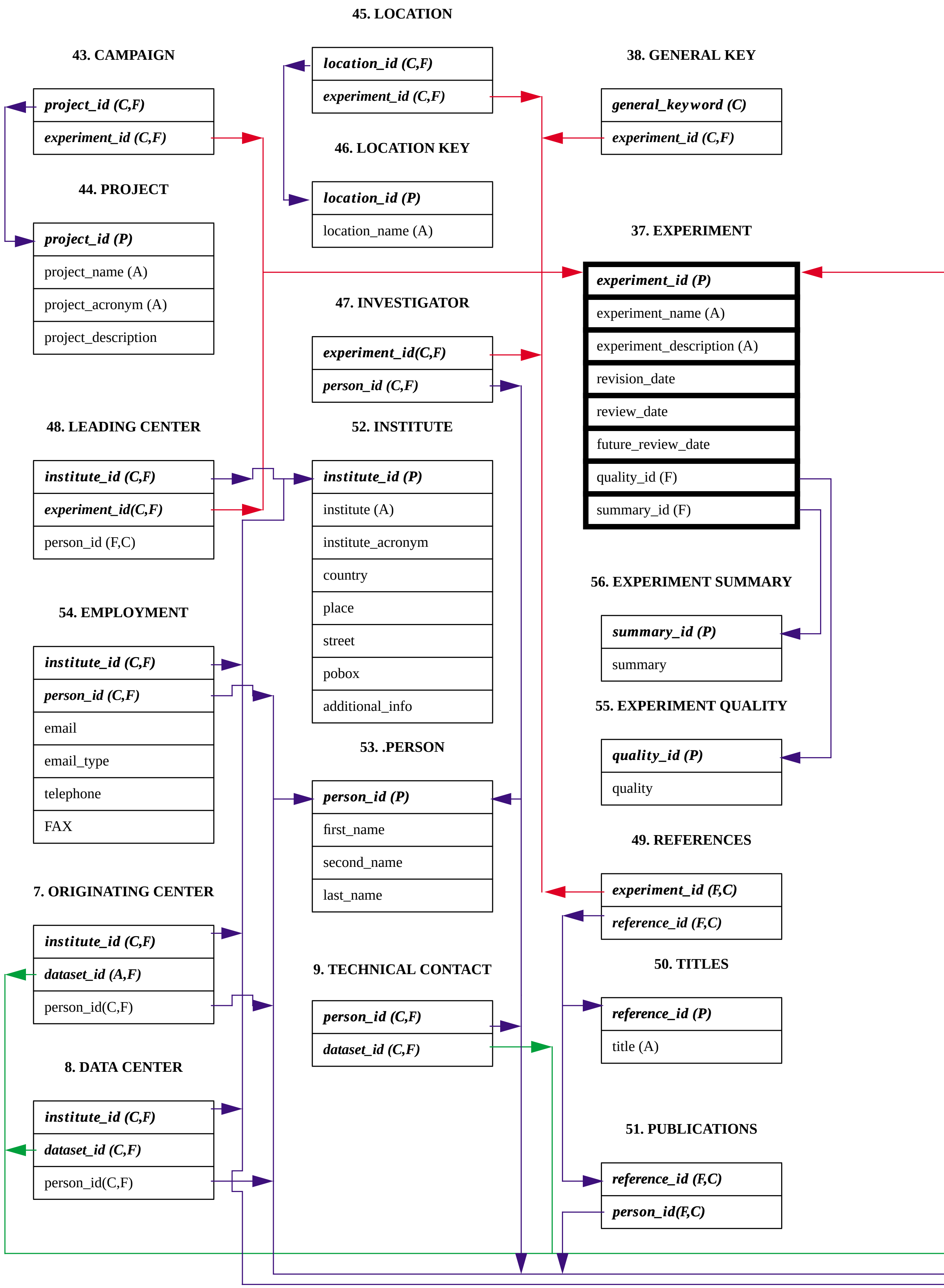
Data Element	Data_Type	Description
summary_id	integer*8	Summary number uniquely identifying the summary. of an experiment
summary	char*2000 *500 words*	Contains an abstract and information not contained in other fields. Access routines if the format is not grib. Order of 5-dim room if not grib. Quality.. Description of time i.e.real earth time, ideal earth time.*DIF required Summary.*

CERA Logical Datamodel

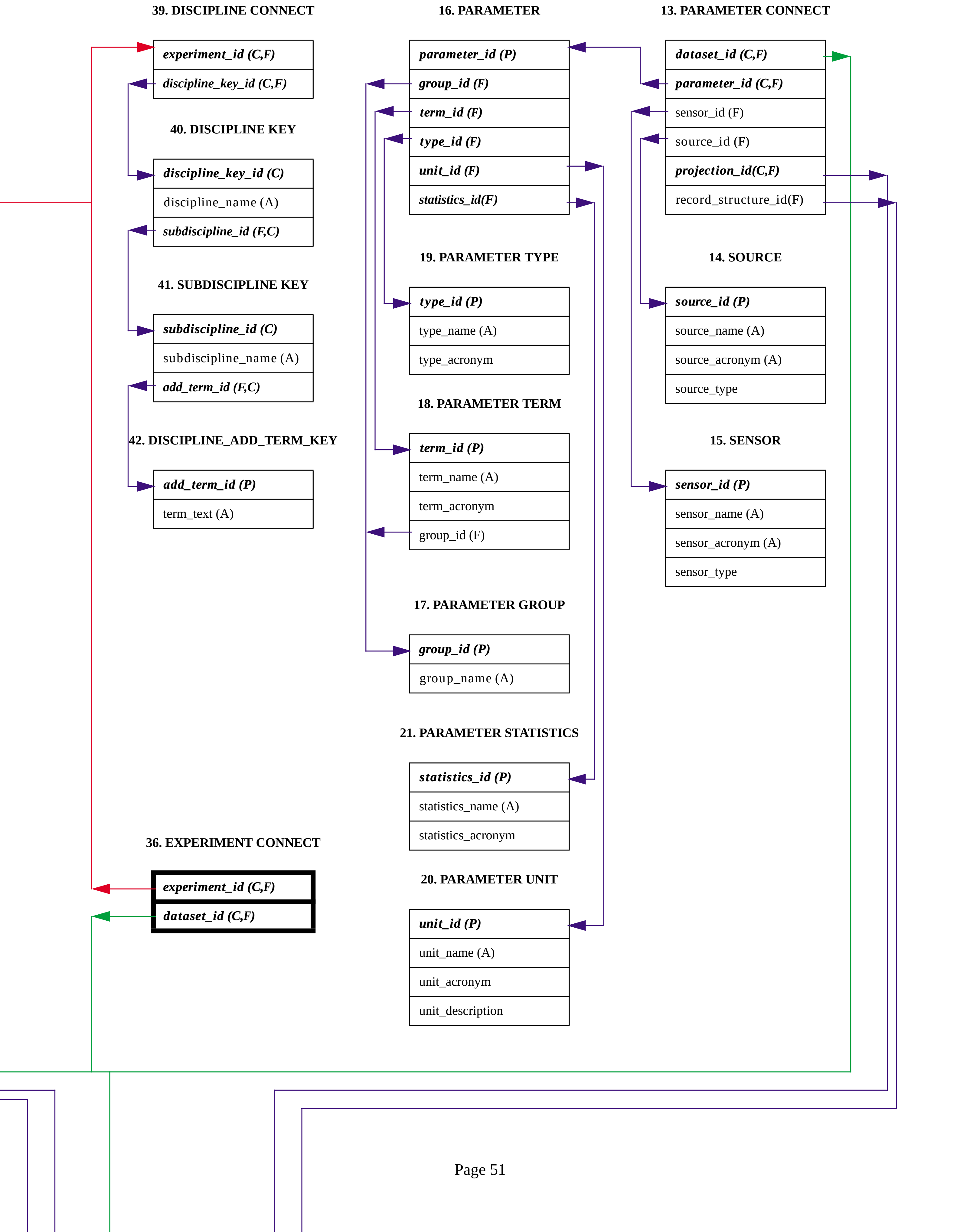
The overall datamodel chart was broken into four pieces to print it on the paper. The correlation and dependencies between the different CERA Tables are indicated.

These symbols are used in the flowchart:

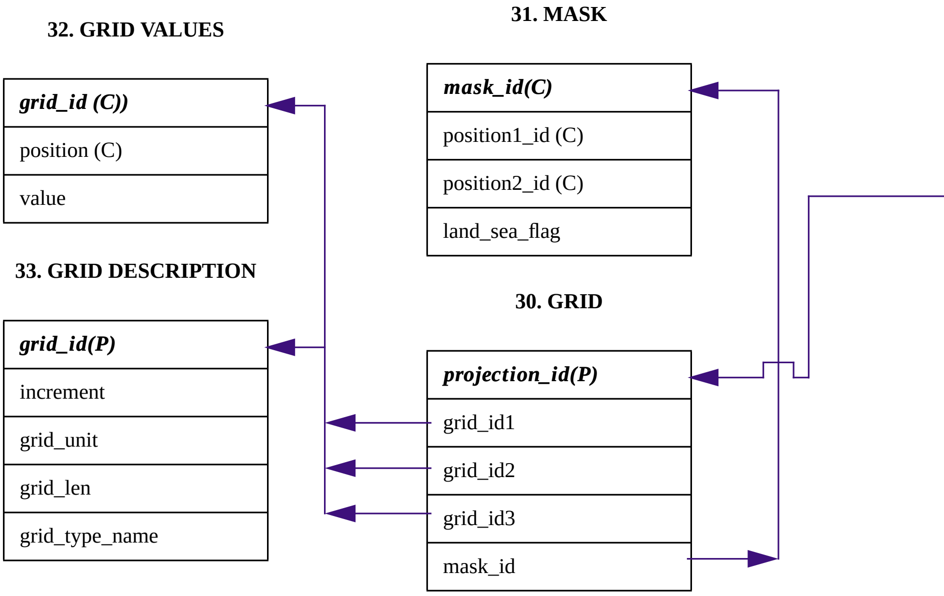
Symbol	Meaning
P	Primary Key
A	Alternate Key
C	Composite Key
F	Foreign Key

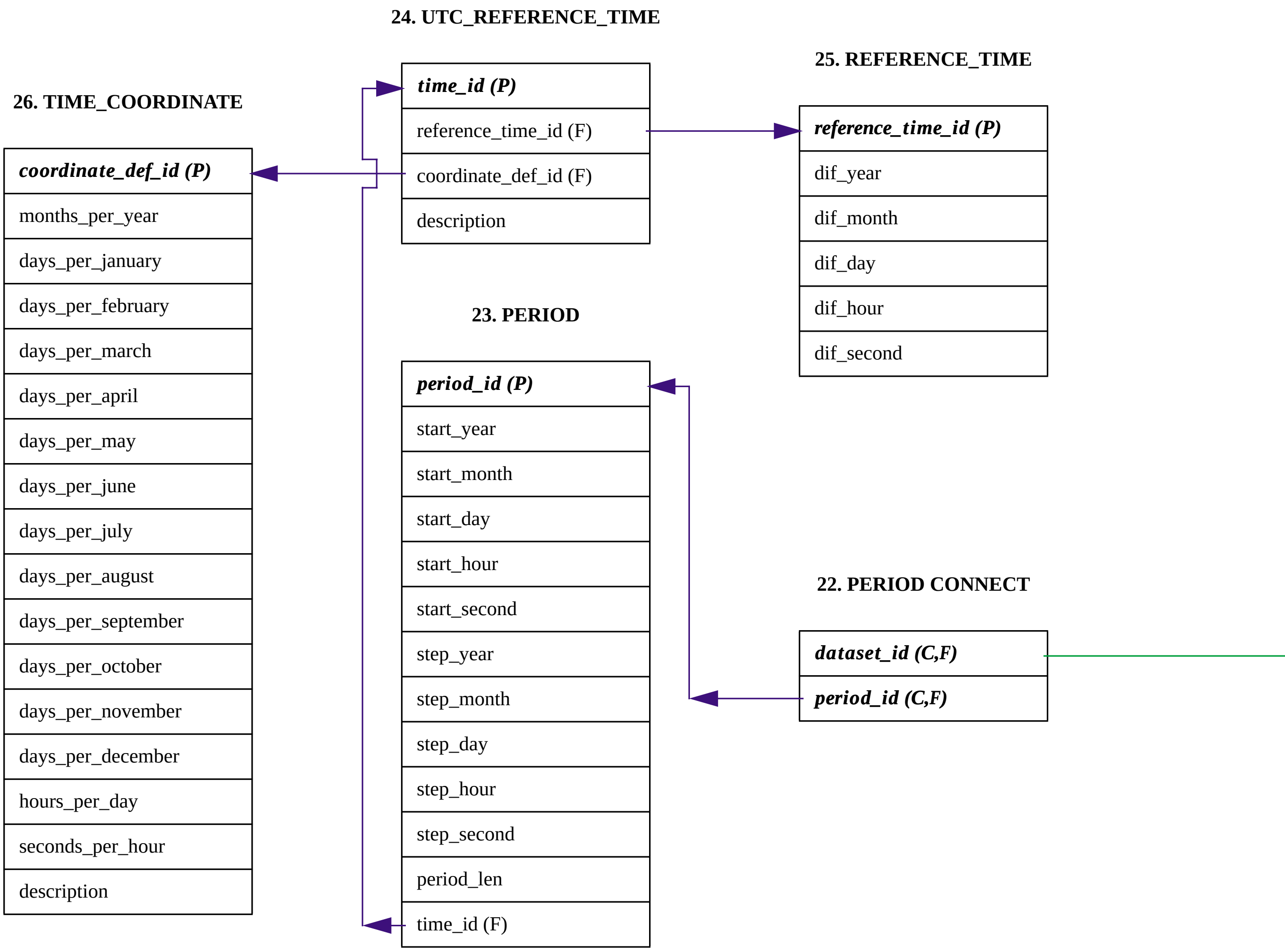


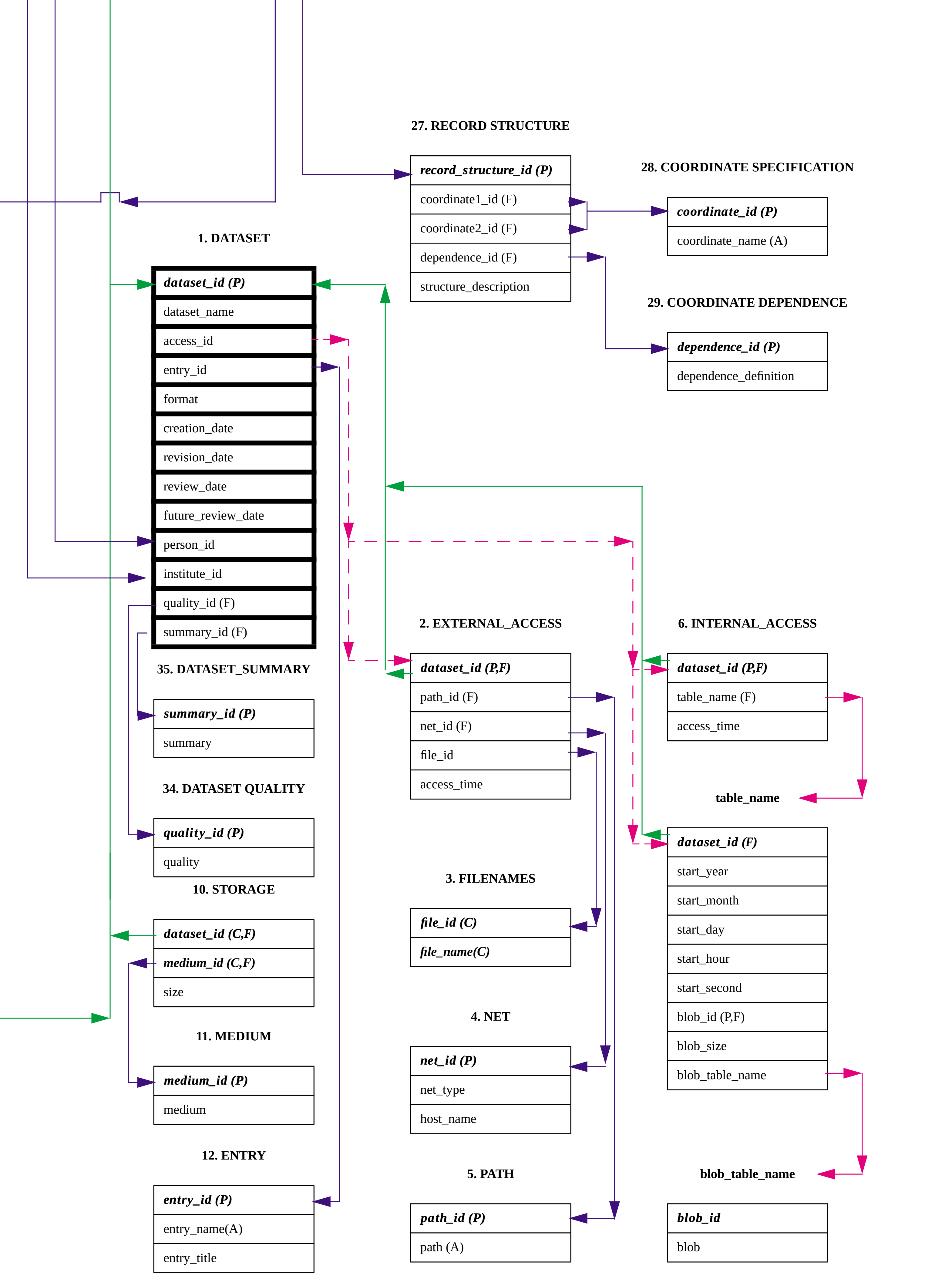
45. LOCATION
location_id (C,F)
experiment_id (C,F)
43. CAMPAIGN
project_id (C,F)
experiment_id (C,F)
38. GENERAL KEY
general_keyword (C)
experiment_id (C,F)
46. LOCATION KEY
location_id (P)
location_name (A)
44. PROJECT
project_id (P)
project_name (A)
project_acronym (A)
project_description
37. EXPERIMENT
experiment_id (P)
experiment_name (A)
experiment_description (A)
revision_date
review_date
future_review_date
quality_id (F)
summary_id (F)
47. INVESTIGATOR
experiment_id(C,F)
person_id (C,F)
48. LEADING CENTER
institute_id (C,F)
experiment_id(C,F)
person_id (F,C)
52. INSTITUTE
institute_id (P)
institute (A)
institute_acronym
country
place
street
pobox
additional_info
56. EXPERIMENT SUMMARY
summary_id (P)
summary
54. EMPLOYMENT
institute_id (C,F)
person_id (C,F)
email
email_type
telephone
FAX
55. EXPERIMENT QUALITY
quality_id (P)
quality
53. .PERSON
person_id (P)
first_name
second_name
last_name
49. REFERENCES
experiment_id (F,C)
reference_id (F,C)
7. ORIGINATING CENTER
institute_id (C,F)
dataset_id (A,F)
person_id(C,F)
9. TECHNICAL CONTACT
person_id (C,F)
dataset_id (C,F)
50. TITLES
reference_id (P)
title (A)
8. DATA CENTER
institute_id (C,F)
dataset_id (C,F)
person_id(C,F)
51. PUBLICATIONS
reference_id (F,C)
person_id(F,C)



39. DISCIPLINE CONNECT
experiment_id (C,F)
discipline_key_id (C,F)
40. DISCIPLINE KEY
discipline_key_id (C)
discipline_name (A)
subdiscipline_id (F,C)
41. SUBDISCIPLINE KEY
subdiscipline_id (C)
subdiscipline_name (A)
add_term_id (F,C)
42. DISCIPLINE_ADD_TERM_KEY
add_term_id (P)
term_text (A)
36. EXPERIMENT CONNECT
experiment_id (C,F)
dataset_id (C,F)
16. PARAMETER
parameter_id (P)
group_id (F)
term_id (F)
type_id (F)
unit_id (F)
statistics_id(F)
19. PARAMETER TYPE
type_id (P)
type_name (A)
type_acronym
18. PARAMETER TERM
term_id (P)
term_name (A)
term_acronym
group_id (F)
17. PARAMETER GROUP
group_id (P)
group_name (A)
21. PARAMETER STATISTICS
statistics_id (P)
statistics_name (A)
statistics_acronym
20. PARAMETER UNIT
unit_id (P)
unit_name (A)
unit_acronym
unit_description
13. PARAMETER CONNECT
dataset_id (C,F)
parameter_id (C,F)
sensor_id (F)
source_id (F)
projection_id(C,F)
record_structure_id(F)
14. SOURCE
source_id (P)
source_name (A)
source_acronym (A)
source_type
15. SENSOR
sensor_id (P)
sensor_name (A)
sensor_acronym (A)
sensor_type


32. GRID VALUES
grid_id (C))
position (C)
value
33. GRID DESCRIPTION
grid_id(P)
increment
grid_unit
grid_len
grid_type_name
31. MASK
mask_id(C)
position1_id (C)
position2_id (C)
land_sea_flag
30. GRID
projection_id(P)
grid_id1
grid_id2
grid_id3
mask_id


24. UTC_REFERENCE_TIME
time_id (P)
reference_time_id (F)
coordinate_def_id (F)
description
26. TIME_COORDINATE
coordinate_def_id (P)
months_per_year
days_per_january
days_per_february
days_per_march
days_per_april
days_per_may
days_per_june
days_per_july
days_per_august
days_per_september
days_per_october
days_per_november
days_per_december
hours_per_day
seconds_per_hour
description
25. REFERENCE_TIME
reference_time_id (P)
dif_year
dif_month
dif_day
dif_hour
dif_second
23. PERIOD
period_id (P)
start_year
start_month
start_day
start_hour
start_second
step_year
step_month
step_day
step_hour
step_second
period_len
time_id (F)
22. PERIOD CONNECT
dataset_id (C,F)
period_id (C,F)


27. RECORD STRUCTURE
record_structure_id (P)
coordinate1_id (F)
coordinate2_id (F)
dependence_id (F)
structure_description
28. COORDINATE SPECIFICATION
coordinate_id (P)
coordinate_name (A)
29. COORDINATE DEPENDENCE
dependence_id (P)
dependence_definition
1. DATASET
dataset_id (P)
dataset_name
access_id
entry_id
format
creation_date
revision_date
review_date
future_review_date
person_id
institute_id
quality_id (F)
summary_id (F)
35. DATASET_SUMMARY
summary_id (P)
summary
34. DATASET QUALITY
quality_id (P)
quality
10. STORAGE
dataset_id (C,F)
medium_id (C,F)
size
11. MEDIUM
medium_id (P)
medium
12. ENTRY
entry_id (P)
entry_name(A)
entry_title
2. EXTERNAL_ACCESS
dataset_id (P,F)
path_id (F)
net_id (F)
file_id
access_time
3. FILENAMES
file_id (C)
file_name(C)
4. NET
net_id (P)
net_type
host_name
5. PATH
path_id (P)
path (A)
6. INTERNAL_ACCESS
dataset_id (P,F)
table_name (F)
access_time
table_name
dataset_id (F)
start_year
start_month
start_day
start_hour
start_second
blob_id (P,F)
blob_size
blob_table_name
blob_table_name
blob_id
blob

Abbreviations and Definitions

- **BLOB** : Binary Large Object, defined as (2dim.) climate data field stored as table entry.
- **Dataset** : compilation of 2dim. data fields for one given parameter (e.g. time series of surface temperature stored in DB table)
- **DIF** : Directory Interchange Format. W3 http://gcmd.gsfc.nasa.gov/difguide/difman.html
- **Experiment** : compilation of datasets
- **FDGC** : Federal Geographic Data Committee
- **GRIB** : Grid in Binary. WMO data format standard. Open gopher from the U.S. National Weather Service. Open:
  gopher://cominfo.nws.noaa.gov/
  and chose: World Meteorological Organization Info and Recent Activities
  and then: Approved Meteorological code changes
  The gopher files
  GRIB Code Documentation
  GRIB Code Tables
  Amendments to GRIB and BUFR Codes
  should contain everything you need.

- **IEEE** : Institute of Electrical and Electronical Engineers
- **INFOCLIMA** : World Climate Data Information Retrieval Service
- **NASA** : National Aeronautics and Space Administration
- **Processing** : standard operations on data (e.g. extracting, averaging, formatting, ...)
- **UNITREE**: Archiving System
- **WMO** : World Meteorological Organization