



BSI Standards Publication

Information technology database languages — SQL technical reports

Part 8: Multi-dimensional arrays (SQL/MDA)

National foreword

This Published Document is the UK implementation of ISO/IEC TR 19075-8:2019.

The UK participation in its preparation was entrusted to Technical Committee IST/40, Data management and interchange.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2019
Published by BSI Standards Limited 2019

ISBN 978 0 580 99519 4

ICS 35.060

Compliance with a British Standard cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 May 2019.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------

**Information technology database
languages — SQL technical reports —
Part 8:
Multi-dimensional arrays (SQL/MDA)**

*Langages de base de données IT — SQL rapport techniques —
Partie 8: Tableaux multidimensionnels*





COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword.....	vii
Introduction.....	viii
1 Scope.....	1
2 Normative references.....	3
3 Terms and definitions.....	5
4 Multidimensional Arrays (MDA) concepts.....	7
4.1 Concept.....	7
4.2 Why consider support for MDA in SQL?.....	7
4.3 Array representations.....	9
4.4 Use cases for MDA support in SQL.....	9
4.4.1 The use cases.....	9
4.4.2 Array data ingestion and storage.....	9
4.4.3 Integrated querying of array and relational data.....	10
4.4.4 Updating stored array data.....	10
4.4.5 Exporting arrays.....	10
4.5 Non-Use cases: Direct access to external array data.....	10
5 SQL/MDA data model.....	11
5.1 Data model concepts.....	11
5.2 MD-array.....	11
5.3 MD-array type definition.....	12
5.3.1 Type definition concepts.....	12
5.3.2 Element type.....	12
5.3.3 MD-dimension.....	13
5.3.4 MD-axis names.....	13
5.3.5 MD-axis lower and upper limits.....	13
5.3.6 Putting it all together.....	14
5.4 MD-array creation.....	16
5.4.1 MD-array creation concepts.....	16
5.4.2 Explicit element enumeration.....	17
5.4.3 From SQL table query result.....	18
5.4.4 Construction by implicit iteration.....	19
5.4.5 Decoding a format-encoded array.....	20
5.5 MD-array updating.....	21
5.5.1 MD-array updating introduction.....	21
5.5.2 Updating MD-arrays of equal MD-dimension.....	22
5.5.3 Updating MD-arrays of greater MD-dimension.....	23
5.5.4 Updating a single element of an MD-array.....	24
5.6 Exporting MD-arrays.....	24

5.6.1	Encoding to a data format.	24
5.6.2	Converting to an SQL table.	26
6	SQL/MDA operations.	29
6.1	Introduction to SQL/MDA operations.	29
6.2	MD-extent probing operators.	29
6.3	MD-array element reference.	31
6.4	MD-extent modifying operations.	32
6.4.1	Introduction to MDE-extent modifying operations.	32
6.4.2	Subsetting.	32
6.4.3	Reshaping.	34
6.4.4	Shifting.	36
6.4.5	MD-axis renaming.	36
6.5	MD-array deriving operators.	37
6.5.1	Introduction to MD-array deriving operators.	37
6.5.2	Scaling.	37
6.5.3	Concatenation.	39
6.5.4	Induced operations.	39
6.5.5	Join MD-arrays on their coordinates.	46
6.6	MD-array aggregation.	47
6.6.1	General aggregation expression.	47
6.6.2	Shorthand aggregation functions.	48
7	Remote sensing example.	51
7.1	Introduction to remote sensing example.	51
7.2	Data setup.	51
7.3	Band math.	53
7.3.1	Introduction to band math.	53
7.3.2	NDVI.	53
7.3.3	Band Swapping.	56
7.4	Histograms.	57
7.5	Change detection.	58
7.6	Extracting features.	59
7.7	Data search and filtering.	60
	Bibliography.	63
	Index.	65

Tables

Table	Page
1 Examples of MD-array type definitions.	15
2 Examples of MD-arrays constructed by element enumeration.	18
3 Examples of MD-arrays created with the constructor by iteration.	20
4 Examples of MD-arrays created from JSON-encoded arrays.	21
5 Examples of MD-arrays encoded to JSON arrays.	25
6 Result of example UNNEST query.	26
7 Result of example UNNEST query specifying WITH ORDINALITY.	27
8 Examples with MD-extent probing functions.	30
9 Result of MDEXTENT(kernel).	30
10 Result of MDEXTENT_MAX(kernel).	30
11 Examples of referencing a single element in an MD-array.	31
12 Examples of MD-array subsetting.	34
13 Examples of MD-extent reshaping.	35
14 Examples of MD-extent shifting.	36
15 Examples of MD-axis renaming.	37
16 Interpolation methods defined in ISO 19123:2005	38
17 Examples of MD-array concatenation.	39
18 Examples of induced function application to MD-arrays.	42
19 Operations corresponding to the <md-array value expression> grammar rules.	44
20 Examples of induced MD-array expressions.	44
21 Example of induced MD-array casting.	45
22 Examples of induced CASE expression.	45
23 Examples of MDJOIN.	47
24 Identity elements for the <md-array aggregation operator>s.	47
25 Examples of general MD-array aggregation.	48
26 Predefined aggregation operators.	49
27 Landsat TM bands.	51

Figures

Figure	Page
1 Aerial greyscale image of size 1024x1024 (San Diego).	8
2 Relationships between MDA and SQL/MDA.	11
3 The structure of an MD-array value illustrated on a sample 3x3 array.	12
4 Placement of satellite images of each country on a world map (from Geographic Bounding Boxes).	14
5 Example of an SQL table that corresponds to a 3x3 MD-array.	19
6 Example of an SQL table converted to a 3x3 MD-array with MD-extent [i(-1:1), j(-1:1)].	19
7 Example of array update.	23
8 Updating a 3-D MD-array with a 2-D source MD-array.	24
9 MD-array subsetting examples.	32
10 MD-array reshaping example.	35
11 MD-array shifting example.	36
12 MD-array scaling example.	37
13 Concatenation examples.	39
14 Example of summing two MD-arrays.	40
15 Colorized array.	46
16 Visible color (RGB) bands of a Landsat TM scene.	52
17 NDVI result stretched to the range (0,255).	54
18 NDVI values between 0.2 and 0.4 shown in white, while everything else is black.	55
19 Color-mapped NDVI result, from dark blue, through grey, to dark green.	56
20 False color image constructed from the near IR, red and green bands.	57
21 Histogram of the NDVI index of a Landsat TM scene.	58
22 A composite image with an NDVI index from different years in each channel.	59
23 Natural RGB color of barrier islands area.	60
24 Binary image showing isolated islands.	60

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents), or the IEC list of patent declarations received (see <http://patents.iec.ch>).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

A list of all parts in the ISO/IEC 19075 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

NOTE 1 — The individual parts of multi-part standards are not necessarily published together. New editions of one or more parts may be published without publication of new editions of other parts.

Introduction

This Technical Report describes the definition and use of multi-dimensional arrays in SQL. Multidimensional arrays represent a core underlying structure of manifold science and engineering data. It is generally recognized today, therefore, that arrays have an essential role in Big Data and should become an integral part of the overall data type orchestration in information systems. This Technical Report discusses the syntax and semantics of operations on the MD-array data type defined in [ISO/IEC 9075-15](#).

The organization of this Technical Report is as follows:

- 1) [Clause 1, “Scope”](#), specifies the scope of this Technical Report.
- 2) [Clause 2, “Normative references”](#), identifies standards that are referenced by this Technical Report.
- 3) [Clause 4, “Multidimensional Arrays \(MDA\) concepts”](#), introduces the concept of Multidimensional Arrays.
- 4) [Clause 5, “SQL/MDA data model”](#), introduces the data model.
- 5) [Clause 6, “SQL/MDA operations”](#), covers the supported operations on MD-arrays..
- 6) [Clause 7, “Remote sensing example”](#), illustrates the supported functionality through realistic examples.

Information technology database languages — SQL technical reports —

Part 8:

Multi-dimensional arrays (SQL/MDA)

1 Scope

This Technical Report describes the support in SQL for Multi-Dimensional Arrays (MDA) as defined in [ISO/IEC 9075-15](#).

(Blank page)

2 Normative references

There are no normative references in this document.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>.
- IEC Electropedia: available at <http://www.electropedia.org/>.

(Blank page)

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

- 3.1 coordinate**
non-empty ordered list of integers
- 3.2 cardinality**
number of elements in an MD-array
- 3.3 MD-array**
ordered collection of elements of the same type associated with an MD-extent where each element is 1:1 associated with some coordinate within its MD-extent.
- NOTE 2 — A coordinate is within an MD-extent if every coordinate value from the integer list is greater than or equal to the lower limit, and less than or equal to the upper limit of the MD-interval of the MD-axis at the position in the MD-extent as the coordinate value has within the coordinate
- 3.4 MD-axis**
named MD-interval
- 3.5 MD-dimension**
number of MD-axes in the MD-extent of an MD-array
- NOTE 3 — Also known as “rank” outside of SQL/MDA
- 3.6 MD-extent**
non-empty ordered collection of MD-axes with no duplicate names
- 3.7 MD-interval**
integer interval given by a pair of lower and upper integer limits such that the lower limit is less than or equal to the upper limit; the interval is closed, *i.e.*, both limits are contained in it

(Blank page)

4 Multidimensional Arrays (MDA) concepts

4.1 Concept

The phrase “(Multidimensional) array, raster data” is used to refer to arrays generally, in contrast to the MD-array term confined to the realm of SQL/MDA. It is not to be confused with the term “array” in [ISO/IEC 9075-2:2016](#). This document uses the term ARRAY for the original SQL array collection type.

The array concept is a simple and efficient data representation that finds its use in a wide array of fields, business-related as well as scientific and engineering. Many sensors, images, image timeseries, simulation processes, statistical models, and so on, produce raw data that can immediately be classified as array data. These data may be naturally arranged along more than one axis: position and time, for example.

A *multidimensional array* (MDA) is a set of elements ordered in a multidimensional space. The space considered here is discretized (also called rasterized or gridded), that is, only integer coordinates are admitted as positions of the individual array elements. The number of integers needed to refer a particular position in this space is the array's dimension (sometimes also referred to as its dimensionality).

An element can be a single value (such as an intensity value in case of grayscale images) or a composite value (such as integer triples for the red, green, and blue components of a true-color image). All elements of an array share the same structure, referred to as the array's element type.

4.2 Why consider support for MDA in SQL?

Large multidimensional arrays in particular represent a prevalent data type across most scientific domains, with examples including 1-D sensor data, 2-D satellite images and microscope scans, 3-D x/y/t image timeseries and x/y/z voxel models, as well as 4-D and 5-D climate models.



Figure 1 — Aerial greyscale image of size 1024x1024 (San Diego)

In array terms, the image in [Figure 1, “Aerial greyscale image of size 1024x1024 \(San Diego\)”](#), is a 2-dimensional array of unsigned 8-bit integer elements positioned at coordinates in $\{0, 1, \dots, 1023\}^2$ space.

Arrays rarely occur isolated in practice and are typically ornamented with metadata and embedded in larger overall information structures. Supporting them in narrowly specialized *ad hoc* tools or dedicated array DBMS is thus insufficient when it comes to building modern, complex services and applications. This suggests that integration of array querying into a standardized framework like SQL is a logical next step that will benefit the communities dealing with multidimensional array data in one way or the other.

SQL has had basic support for 1-dimensional arrays since 1999. Instead of attempting to extend the existing 1-dimensional array model to address the needs of multidimensional array manipulation, SQL/MDA addresses those needs with a new feature set integrated into SQL.

4.3 Array representations

The encoding and decoding function semantics for other external representations are implementation-defined. Examples may include data in such representations as PDF, JPEG, PNG, and XML.

4.4 Use cases for MDA support in SQL

4.4.1 The use cases

The question posed by this use case is “How is array data acquired using SQL?”

Following are the primary use cases that support for multidimensional arrays in the SQL-environment is required to satisfy.

- Array data ingestion and storage.
- Integrated querying of array and relational data,
- Updating stored array data.
- Exporting arrays.

The following Subclauses discuss these use cases in greater detail, and how SQL/MDA addresses them.

4.4.2 Array data ingestion and storage

The question posed by this use case is “How is array data acquired using SQL?”

As discussed earlier in [Subclause 4.3, “Array representations”](#), arrays exist in a wide variety of formats. In order to work with them in a generic way in SQL, it is necessary to build an abstract data model that fits with the SQL philosophy. The MD-array as defined by SQL/MDA provides exactly such a data model, implemented as a new attribute type MDARRAY. Ingestion of array data encoded in an external format into SQL involves transforming it or decoding it into an instance of the internal MD-array data model, which is then inserted into an MDARRAY column of an appropriate type.

What “decode” means in practice depends on many factors, including the data format, the details of physical storage of MD-arrays in a specific DBMS, system architecture, *etc.* This Technical Report and the standard do not dive into these technical details of array data ingestion beyond providing a default specification for JSON encoded arrays and a suitable interface for implementations to attach their ingestion extensions.

It is worth discussing the storage data model here. The several possibilities are:

- MD-array as a first-class object in the same way that SQL tables are.
- Direct mapping of SQL tables into MD-arrays.
- Store within an opaque data type (SQL string or Large Object for example).
- A dedicated column data type with well-defined semantics.

MD-array is a simple data structure defined by a list of MD-axes, each specifying a name, lower and upper limits, paired with an element type. This led to adoption of the last option, following the example of ARRAY and MULTISSET collection data types. Data transformation is handled during ingestion with special functions,

allowing working with values with clearly defined semantics within the SQL-environment. It's minimally intrusive to the SQL standard, while it nevertheless supports all of the requirements identified in this Technical Report.

4.4.3 Integrated querying of array and relational data

As was introduced in the previous Subclause, MD-arrays are stored within a new collection data type MDARRAY that is manipulated through a functional and operational interface described in this Technical Report. This is similar to the existing ARRAY and MULTiset collection data types, except that the operation set is richer. Integration with other data types is seamless (*e.g.*, multiplying the values of all elements of a numeric MD-array column A with the single value of a numeric column C is simply $A * C$), and the general SQL query mechanics are unchanged. In addition, it is possible to generate an SQL table from an MD-array and vice-versa, an MD-array from an SQL table with the appropriate structure.

4.4.4 Updating stored array data

Read-only access to MD-array data is clearly insufficient. Array data is very often continuously and regularly produced, *e.g.*, a temperature sensor taking a reading every hour, or a satellite periodically taking earth-observation images as it orbits around the Earth. In addition, a single array can exceed terabytes in size, and for practical reasons it might be split into multiple smaller arrays; ingesting them all into a single MD-array column requires piece-wise extension and updating of the column. Therefore, SQL/MDA allows updating of entire MD-array values, as well as specific subsets of an MD-array.

4.4.5 Exporting arrays

Frequently the result of operations on MD-arrays will be an MDarray, which needs to be exported using some external representation. This is the counterpart of array data ingestion discussed previously in [Subclause 4.4.2, “Array data ingestion and storage”](#).

4.5 Non-Use cases: Direct access to external array data

All access to array data requires that the array data is first imported into the SQL environment. In order to query external array data using SQL, applications are required to access external arrays themselves, then insert those data into MD-array values, perhaps by using the MDDECODE function.

5 SQL/MDA data model

5.1 Data model concepts

The SQL/MDA model is essentially represented by the concept of MD-array. It is necessary to clearly distinguish between array values “outside” the DBMS, and their analogs “inside” the DBMS. The the following convention is used:

- The terms “array”, “multidimensional array”, and “MDA” refer to array values external to the SQL-environment, encoded in a particular format like TIFF, netCDF, HDF5, JSON, *etc.*
- The terms “MD-array” and “SQL/MDA” refer to constructs within the SQL-environment.

The relationship between “MDA” and “SQL/MDA” is illustrated in Figure 2, “Relationships between MDA and SQL/MDA”.

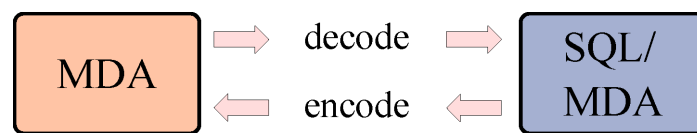


Figure 2 — Relationships between MDA and SQL/MDA

5.2 MD-array

MD-array values are inputs of all SQL/MDA operations, and most often the outputs. Figure 3, “The structure of an MD-array value illustrated on a sample 3x3 array”, shows the structure of a sample MD-array value.

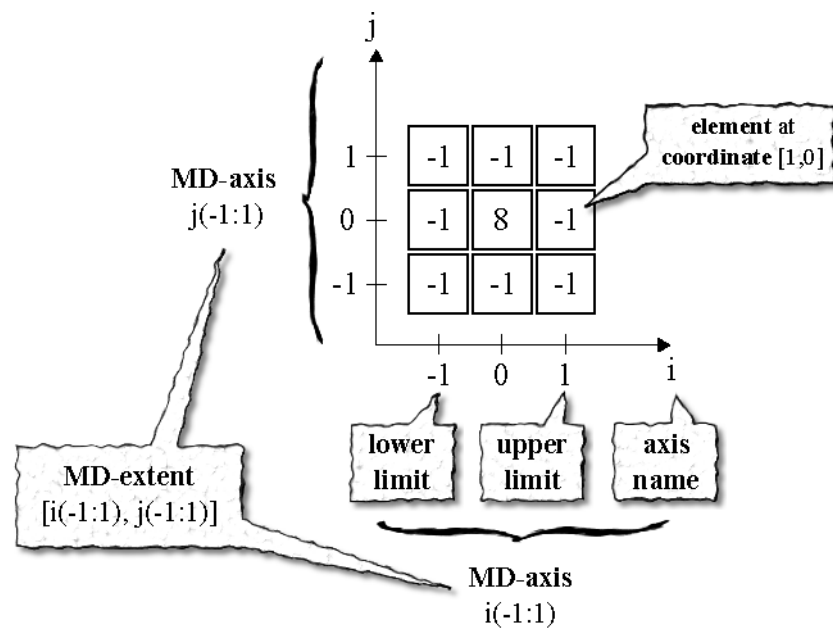


Figure 3 — The structure of an MD-array value illustrated on a sample 3x3 array

5.3 MD-array type definition

5.3.1 Type definition concepts

The definition of an MD-array (see [Clause 3, “Terms and definitions”](#)) is a good starting point in order to understand what components are needed for the type of an MD-array:

- 1) “An MD-array is an ordered collection of elements of the same type ...” So, one thing needed to specify the type of an MD-array is the type of its elements, more specifically known as the element type. This is no different from the existing ARRAY and MULTISSET.
- 2) “... where each element is 1:1 associated with some coordinate within its MD-extent.” Hence, the other part needed is an MD-extent that delimits the coordinates of the elements in an MD-array.

5.3.2 Element type

MD-arrays stand out from the spectrum of collection types in that the storage location of an element can be derived directly from its coordinates, which makes storage and access particularly efficient. This requires that all elements are of the same length. Therefore, variable-size collection elements like sets and multisets do not qualify as element types. MD-arrays as element type is disallowed as well for the following reasons:

- 1) Nesting an MD-array of MD-dimension d_1 into an MD-array of MD-dimension d_2 can equivalently be modeled as a single MD-array of MD-dimension d_1+d_2 .
- 2) It keeps the data model simpler and more consistent in that all collection types are disallowed, and no handling specifically of MD-arrays is needed.

All in all, any SQL data type is allowed to be an element type of an MD-array, except for collection-containing types. A data type *TY* is *collection-containing* if exactly one of the following conditions is true:

- *TY* is a collection type.
- *TY* is a row type, and the declared type of some field of *TY* is a collection-containing type.
- *TY* is distinct type, and the source type of *TY* is a collection-containing type.
- *TY* is a structured type and the declared type of some attribute of *TY* is a collection-containing type.

5.3.3 MD-dimension

The MD-dimension is an essential property of an MD-array that indicates how many MD-axes it has. Two MD-arrays of different MD-dimensions are fundamentally different. Therefore, an MD-array type that specifies a certain MD-dimension admits only MD-array values of that MD-dimension.

An MD-array has an MD-extent that is a list of MD-axes. Each MD-axis has a name, a lower limit, and an upper limit.

5.3.4 MD-axis names

The name of an MD-axis uniquely identifies that MD-axis, which becomes relevant in operations that refer to the MD-axes of an MD-array. In operations on two or more MD-arrays, the names of corresponding MD-axes are required to be the same; a regular 2D x/y image is completely different from a transposed y/x image, after all. It might happen that some MD-arrays correspond semantically, while the corresponding MD-axis names are different (for example, “t” in one MD-array and “time” in another); SQL/MDA provides a CAST variant for such cases that allows explicitly renaming the MD-axis names.

5.3.5 MD-axis lower and upper limits

The lower and upper limits of the MD-axes are not fundamental to the nature of an MD-array. MD-arrays with different lower and upper limits might still be related to each other, as the following example illustrates.

Suppose there exist greyscale satellite images of each country in the world in the same resolution. In SQL/MDA they would be 2-dimensional MD-arrays of different sizes (the “width” of the first MD-axis and “height” of the second MD-axis), as there are smaller and larger countries. In a “map” of the whole world in the same resolution¹, the MD-array for each country would be placed at a different position on the overall map (Figure 4, “Placement of satellite images of each country on a world map (from Geographic Bounding Boxes)”, i.e., the lower and upper limits of its MD-axes would be different from those of other MD-arrays. Nevertheless, they are related to each other, and it would be beneficial to possibly to put them in a single MDARRAY column, connecting them to further columns holding metadata like the country name, geographic boundaries, population, etc.

¹ “resolution” refers to the real size of a single pixel, e.g., 30 meters

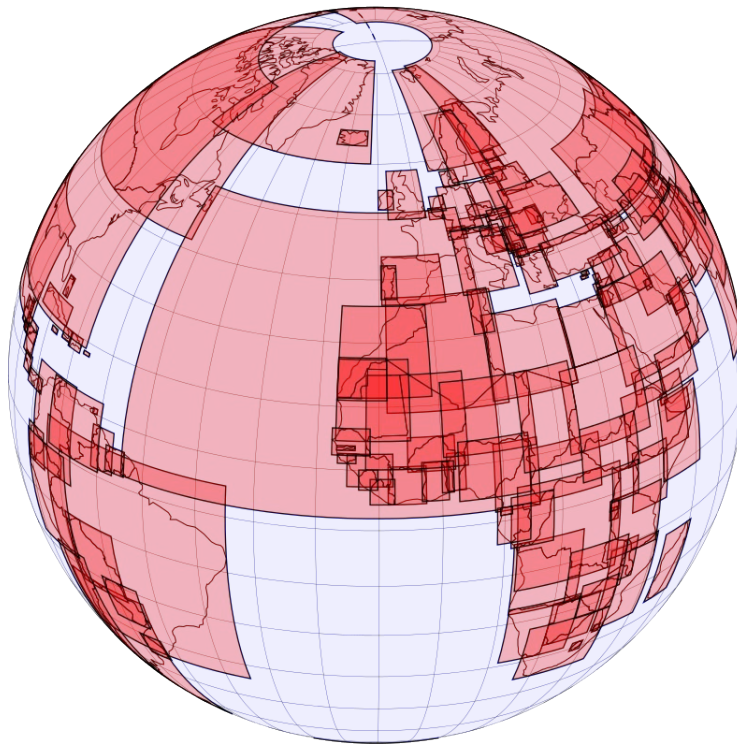


Figure 4 — Placement of satellite images of each country on a world map (from [Geographic Bounding Boxes](#))

Therefore, MD-array values can have varying lower and upper limits. The MD-array type can optionally be declared with minimum lower and maximum upper axis limits; if no limit is defined for an axis, it can extend to the implementation-defined axis limit.

5.3.6 Putting it all together

Stepping through the MD-array type definition rules of [Subclause 8.1](#), “<data type>”, in ISO/IEC 9075-15:

```
<md-array type> ::=  
  <data type> MDARRAY <maximum md-extent>
```

So specifying a column of MD-array type requires specifying first the element type, followed by the keyword MDARRAY, and a maximum MD-extent at the end. Continuing with the specification details of <maximum md-extent>:

```
<maximum md-extent> ::=  
  <left bracket or trigraph> <maximum md-axis list> <right bracket or trigraph>  
  | <left bracket or trigraph> <maximum md-axis list anonymous> <right bracket or trigraph>
```

```
<maximum md-axis list> ::=  
  <maximum md-axis> [ { <comma> <maximum md-axis> }... ]
```

```
<maximum md-axis list anonymous> ::=  
  <maximum md-axis anonymous> [ { <comma> <maximum md-axis anonymous> }... ]
```

A maximum MD-extent is either a list of “regular” maximum MD-axes, or a list of “anonymous” maximum MD-axes. The difference becomes clear in the grammar rules below. It’s worth mentioning here that the list of