EDIFACT

Index

Definition of UN/EDIFACT

<u>UN/EDIFACT</u>: United Nations rules for Electronic Data Interchange for Administration, Commerce and Transport. They comprise a set of internationally agreed standards, directories and guidelines for the electronic interchange of structured data, and in particular that related to trade in goods and services, between independent computerised information systems.

Recommended within the framework of the United Nations, the rules are approved and published by the UN/ECE/WP.4 (United Nations / Economic Commission for Europe / Working Party 4) in the United Nations Trade Data Interchange Directory (UNTDID) and are maintained under agreed procedures. UNTDID includes:

- EDIFACT syntax rules (ISO 9735);
- Message design guidelines;
- Syntax implementation guidelines;
- EDIFACT Data Elements Directory, <u>EDED</u> (a subset of UNTDED);
- EDIFACT Code list, <u>EDCL</u>;
- EDIFACT Composite data elements Directory, <u>EDCD</u>;
- EDIFACT standard Segments Directory, <u>EDSD</u>;
- EDIFACT United Nations Standard Messages Directory, <u>EDMD</u>;
- Uniform Rules of Conduct for the Interchange of Trade Data by Teletransmission (UNCID);
- Explanatory material, as appropriate.

EDIFACT syntax overview

This section is a summary of the ISO 9735 document: "EDIFACT Application level syntax rules", first released on 1988-07-15, amended and reprinted on 1990-11-01, and Amendum 1 of 1992.

The EDIFACT syntax rules set the standards for structuring data into segments, segments into messages, and messages into an interchange.

Structure of an interchange

An interchange may consist of the following segments:

UNA		
UNB	Interchange Header	Mandatory
UNG	Functional Group Header	Conditional
UNH	Message Header	Mandatory
	USER DATA SEGMENTS	
UNT	Message Trailer	Mandatory
UNE	Functional Group Trailer	Conditional
UNZ	Interchange trailer	Mandatory

Segments starting with "UN" are called service segments. They constitute the envelope or the "packing" of the EDIFACT messages.

User data segments contain the information itself, in a format specific to each message type.

Structure of a message

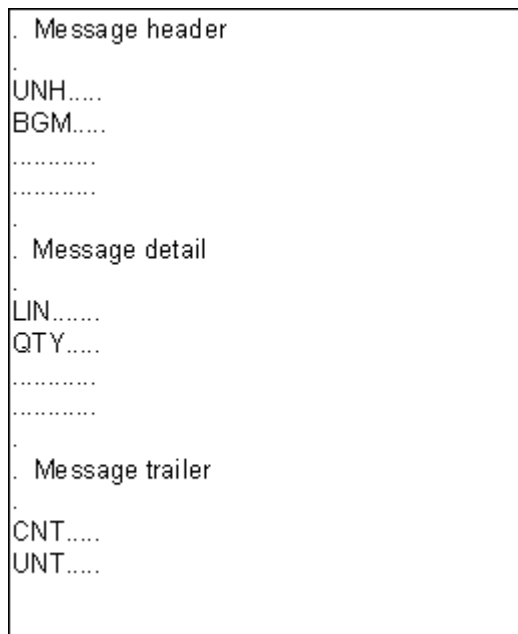Each data segment has a specific place within the sequence of segments in the message. They may occur in any of the following three sections of the message:

a. Heading section - A segment occurring in this section relates to the <u>entire message</u>.

b. Detail section - A segment occurring in this section relates to the detail information only.

c. Summary section - Only segments containing <u>totals or control</u> information may occur in the summary section, e.g. invoice total amount, number of lines in a purchase order, etc.

The sequence of the three message sections can be represented by the following simple example;

```
. Message header
.
UNH.....
BGM.....
...........
...........
.
. Message detail
.
LIN.......
QTY.....
...........
...........
.
. Message trailer
.
CNT.....
UNT.....

```

The same segment type may occur in more than one of the message sections, e.g. in the header and in the detail section, and/or more than once in the same section.

Some segments may be repeated a certain number of times at their specific location in the message. The status, Mandatory or Conditional, and the maximum number of repetitions of segment types are indicated in the message structure.

Within a message, specific groups of functionally related segments may be repeated; these groups are referred to as "segment groups". The maximum number of repetitions of a particular segment group at a specific location is included in the message definition.

A segment group may be nested within other segment groups, provided that the inner segment group terminates before any outer segment group terminates.

Segment structure

A segment consists of:

- A segment tag: identifies the segment type
- Data element separators
- Simple, composite, or component data elements
- A segment terminator

Data elements can be defined as having a fixed or variable length.

A composite data element contains two or more component data elements.

A component data element is a simple data element used in a composite data element.

A data element can be qualified by another data element, the value of which is expressed as a code that gives specific meaning to the data. The data value of a qualifier is a code taken from an agreed set of code values.

Separators

In EANCOM four service characters (extracted from UNOA) have a special meaning and act as the default separators for EANCOM;

Apostrophe	'	=	segment terminator
Plus sign	+	=	segment tag and data element separator
Colon	:	=	component data element separator
Question Mark	?	=	release character; immediately preceding one of the service characters, it restores their normal meaning. E.g. 10?+10=20 means 10+10=20. Question mark is represented by ??

Should trading partners agree to use any of the character sets from B to F (inclusive) and the default separators from UNOA, then the UNA segment must be provided to explicitly state the default separator values.

When using any of the character sets A through F then the syntax version number (DE 0002) in the UNB segment should be set to 3.

Example of an EDIFACT segment:

DTM+137:19940101:102'

DTM	=	Tag of the "Date/Time Period" segment;
+	=	separator;
137	=	Qualifier to indicate the date is the Document/Message Date/Time;
:	=	separator of data elements within a composite (here, the date qualifier and the date);
19940101	=	the date;
:	=	separator of data elements within a composite (here, the date and the date format qualifier);
102	=	Qualifier to indicate the format of the date (CCYYMMDD);
'	=	Segment terminator.

Compression of data

In data elements for which the Trade Data Elements Directory specifies variable length and no other restrictions, non-significant character positions shall be suppressed. In the case of non-significant characters, leading zeroes and trailing spaces shall be suppressed.

TAG = segment tag; DE = data element; CE = component data element.

- **Exclusion of segments.** Conditional segments containing no data shall be omitted (including their segment tags).
- **Exclusion of data elements by omission.** Data elements are identified by their sequential position within the segments as stated in the Segment Directory. If a conditional data element is omitted and followed by another data element, its position shall be indicated by retention of its data element separator.

Eg: TAG+DE+DE+DE+CE:CE:CE' complete segment including all data elements

```
TAG+DE++DE+CE:CE:CE'
        ↑——————————— one DE has been omitted
```

- **Exclusion of data elements by truncation.** If one or more conditional data elements at the end of a segment are omitted, the segment may be truncated by the segment terminator.

Eg: TAG+DE+DE+DE+DE' Original including all data elements

```
TAG+DE+DE'
          ↑——————————— truncation
```

- **Exclusion of component data elements by omission.** If a conditional CE is omitted and followed by another CE, its given position must be represented by its CE separator.

Eg: TAG+DE++DE+CE:CE:CE' Original including all CE's

```
TAG+DE++DE+CE::CE'
         ↑_____ one CE has been omitted
```

- **Exclusion of component data elements by truncation.** One or more conditional CE at the end of a composite DE may be excluded by truncation by the DE separator or, if at the end of a segment, by the segment separator.

Eg: TAG+DE++DE+CE:CE:CE' Original including last CE

```
TAG+DE++DE+CE:CE'
         ↑_____ truncation
```

Representation of numeric values

- **Decimal sign.** The representation for decimal sign is the point on the line (.). The decimal sign shall not be counted as a character when computing the maximum field length of a data element. When a decimal sign is transmitted, there shall be at least one digit before and after the decimal sign.

To assist in-house file designers and data interchange partners, the following lengths may be used as a guideline:

Numeric Class	Format	Integer Digit	Decimal Digit
Amounts	n..18	15	3
Control Values	n..18	14	4
Cubes	n..9	5	4
Currency Rates	n..12	6	6
Other Range Value	n..18	15	3
Percentages	n..10	6	4
Percentage Range Value	n..18	14	4
Quantities	n..15	12	3
Rate per Unit	n..15	12	3
Tax Rates	n..17	13	4
Unit Prices	n..15	11	4
Unit Price Basis	n..9	6	3
Weights	n..18	15	3

- **Triad separator.** Triad separators shall not be used in interchange. (Allowed: 2500000; Not allowed: 2,500,000 or 2.500.000 or 2 500 000).

- **Sign.** Numeric data element values shall be regarded as positive. Although conceptually a deduction is negative, it shall be represented by a positive value, e.g. in a credit note all values will be positive, the application software will using the message name coded (DE 1001) convert all values into negative. In addition some data element and code combinations will lead to implied negative values, e.g. data element 5463 with code value 'A, Allowance' in an ALC segment in an invoice.

If a value is to be represented as negative, it shall in transmission be immediately preceded by a minus sign e.g. -112. The minus sign shall not be counted as a character when computing the maximum field length of a data element.

Character set

For the characters in the set below, the 7-bit codes in the basic ISO 646 standard shall be used, unless the corresponding 8-bit codes in ISO 6937, ISO 8859.1, ISO 8859.2, ISO 8859.5 and ISO 8859.7, or other bit codes are specifically agreed between the interchanging partners. If another bit code set is being used this must be specifically stated in the interchange agreement between the parties.

The following character set, known as character set level A, is __recommended__ to be used within EANCOM for international exchanges:

Letters, upper case	A to Z
Numerals	0 to 9
Space character	
Full stop	.
Comma	,
Hyphen/minus sign	-
Opening parentheses	(
Closing parentheses	)
Oblique stroke (slash)	/
Equal sign	=
Exclamation mark	!
Quotation mark	"
Percentage sign	%
Ampersand	&
Asterisk	*
Semi-colon	;
Less-than sign	<
Greater-than sign	>

When using the Level A service characters (UNOA), the characters Apostrophe ('), Plus sign (+), Colon (:) and Question mark (?) are reserved for special use (see section 2.2.4.).

In addition to the Level A characters detailed above, a number of additional characters may be required by EANCOM users. These characters are contained in the International Reference Version (IRV) of ISO 646.

EAN International recommends that should any EANCOM user wish to use any of these additional characters, that agreement must first be obtained from their trading partners in order to ensure correct processing by the receiving application.

Following you will find the set of characters contained in the IRV:

Number sign	#
Commercial at	@
Left square bracket	[.
Reverse solidus	\
Right square bracket	] .
Circumflex accent	^
Grave accent	`
Left curly bracket	{
Vertical line	|
Right curly bracket	}

The identification of the character set being used must be made in data element 0001 of the UNB segment.

When using any of the character sets A through F then the syntax version number (DE 0002) in the UNB segment should be set to 3.

The following table explains which languages are catered for in which version of ISO-8859;

ISO 8859 version	Languages
1	Danish, Dutch, English, Faroese, Finnish, French, German, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish, Swedish
2	Albanian, Czech, English, Hungarian, Polish, Romanian, Serbo-Croatian, Slovak, Slovene
5	Bulgarian, Byelorussian, English, Macedonian, Russian, Serbo-Croatian, Ukrainian
7	Greek

Directory status, version and release

All EANCOM 1997 messages are based on the draft directory set D.96A released by the United Nations in March 1996.

This directory set includes techniques for the identification of the status, version and release of the directory. The data element 0052 is used to indicate the status of the directory which is being used, 'D'raft. The EDIFACT Directory Set release is identified by a three character alpha/numeric in data element 0054, of which the first two characters identifies the Directory Set Issue number (which is equivalent to the last two digits of the year in which the Directory Set was issued), followed by a single alpha character indicating the release.

EANCOM message version

Each EANCOM message carries its own version number which allows the unambiguous identification of different versions of the EANCOM messages. The EANCOM version number is indicated in Data Element 0057 in the UNG and UNH segments. It is structured as follows:

> EANnnn

| where: | EAN | = Indicates EAN as the agency controlling the subset. |
| | nnn | = Three-digit version number of the EANCOM subset. |

Version numbers for formally released EANCOM messages start at the number '001' and are incremented by one for each subsequent release of the message.

In addition to the EANCOM message version numbers '001' and greater, EANCOM also caters for messages at version '000', or pre-release. Version '000' EANCOM messages are messages which have completed their development phase and which are available to potential users for trial purposes. Because updates to EANCOM are issued on average once per year it is not possible to make available immediately every EANCOM message developed to every EANCOM user. The existence of version '000' messages is communicated by EAN International to its numbering organisations and copies are made available to the numbering organisations only following specific requests. All version '000' messages are automatically included in the next update of EANCOM when they formally become part of the standard, i.e. move to version '001'.

Documentation conventions

Format and picture of data elements

The following conventions apply in the present documentation:

a	alphabetic characters
n	numeric characters
an	alpha-numeric characters
a3	3 alphabetic characters, fixed length
n3	3 numeric characters, fixed length
an3	3 alpha-numeric characters, fixed length

a..3	up to 3 alphabetic characters
n..3	up to 3 numeric characters
an..3	up to 3 alpha-numeric characters

Status indicators

(M)andatory data elements or composites in EDIFACT segments retain their status in EANCOM.

Additionally, there are five types of status with a (C)onditional EDIFACT status, whether for simple, component or composite data elements. They are listed below and can be identified when relevant by the abbreviations.

-	REQUIRED	**R**	Indicates that the entity is required and must be sent.
-	ADVISED	**A**	Indicates that the entity is advised or recommended.
-	DEPENDENT	**D**	Indicates that the entity must be sent in certain conditions, as defined by the relevant explanatory note.
-	OPTIONAL	**O**	Indicates that the entity is optional and may be sent at the discretion of the user.
-	NOT USED	**N**	Indicates that the entity is not used and should be omitted.

Code values have two status's in EANCOM:

-	**Restricted (*)**	A data element marked with an asterix (*) in the fourth column of the segment details of a message indicates that the listed codes in column five are the only codes available for use with the data element at the same level as the asterix, in the segment, in the message.
-	**Open**.	All data elements where coded representation of data is possible and a restricted set of code values is not indicated are open. The available codes are listed in the Data Elements and Code Sets Directory (Part III of this manual). Code values may be given as examples or there may be a note on the format or type of code to be used.

Temporary Codes.

Several codes within the EANCOM 1997 codes list are identified as being temporary code values through the use of '**(EAN Code)**', or other codes (such as EAN Code Book or SWIFT code), immediately after the code value. These codes are values which do not exist in the UN/EDIFACT codes list. In many instances the data element containing the temporary code value will be followed in a composite by the data element 3055. This data element allows the explicit identification of temporary codes.

When EAN codes are used in coded EDIFACT data elements, which exist in composite data elements containing data elements 1131 and 3055, the code value '9 = EAN' EAN' (for finance messages code value '17 = SWIFT') <u>must</u> be used in data element 3055 to identify

unambiguously the fact a temporary code is being used. See C273 in IMD, C524 in HAN, for examples of this.

All temporary codes will be forwarded to UN/EDIFACT for official EDIFACT values. It must be noted however that the code values allocated by EDIFACT will not normally be the same as the temporary code allocated and that some alignment may be needed when formal EDIFACT codes are issued.

Message branching diagrams and structure charts

Within every EANCOM message two diagrams are presented which explain the structure and sequence of the message. These diagrams are known as the Message Structure Chart and the Message Branching Diagram.
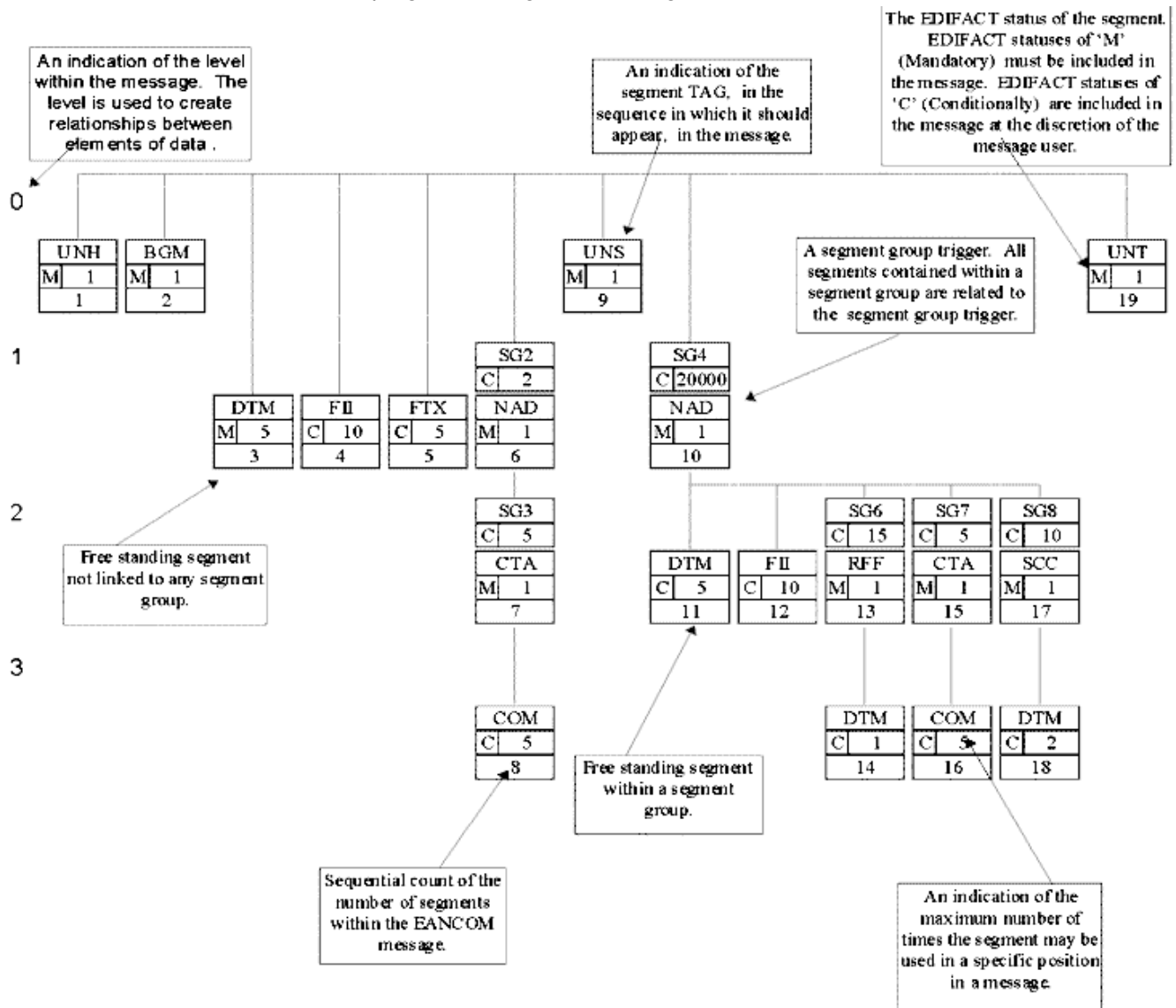
The message structure chart is a sequential chart which presents the message in the sequence in which it must be formatted for transmission. Every message is structured under three headings, header, detail, and summary. An example of a message structure chart follows;
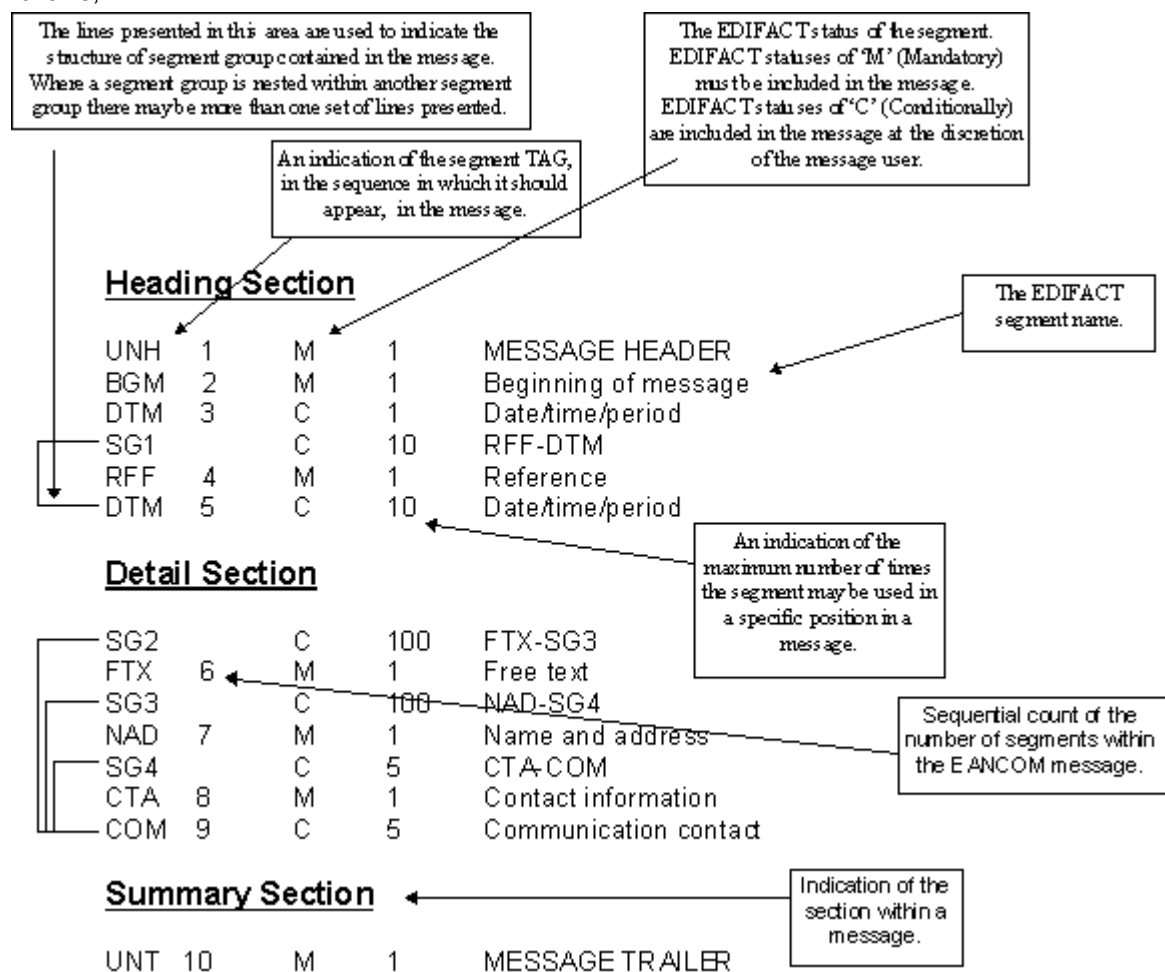
The structure chart should always be read from top down and left to right (please note that the message detailed is simply an example message and does not bear any relevance to real EANCOM messages).

A message branching diagram is a pictorial representation (in flow chart style) which presents the logical sequence and relationships contained within a message.

Branching diagrams should be read, starting at the UNH segment, from left to right and top to bottom. The lines contained within a branching diagram should be considered as guides which must be followed in order to progress through the message.

The EDIFACT status of the segment. EDIFACT statuses of 'M' (Mandatory) must be included in the message. EDIFACT statuses of 'C' (Conditionally) are included in the message at the discretion of the message user.

An indication of the level within the message. The level is used to create relationships between elements of data.

An indication of the segment TAG, in the sequence in which it should appear, in the message.

A segment group trigger. All segments contained within a segment group are related to the segment group trigger.

Free standing segment not linked to any segment group.

Free standing segment within a segment group.

Sequential count of the number of segments within the EANCOM message.

An indication of the maximum number of times the segment may be used in a specific position in a message.

Interchange structure and service segments

The interchange structure in an EDIFACT transmission is organised in several grouping levels. The service segments are the envelope of the groups.

The first service segment possible in an interchange is the 'UNA' segment which is used to define the separators being used in the interchange.

The second service segment, "UNB", indicates the beginning of the interchange.

The next one, "UNG", indicates the beginning of a group of messages of the same type, for example invoices.

The last service segment, "UNH", indicates the beginning of a given message.

To each beginning service segment corresponds an ending service segment (note, UNA is not a beginning segment).

Service string advice:	UNA
Interchange envelope:	UNB .... UNZ
Group envelope:	UNG .... UNE
Message envelope:	UNH .... UNT

The interchange can thus be represented like this:
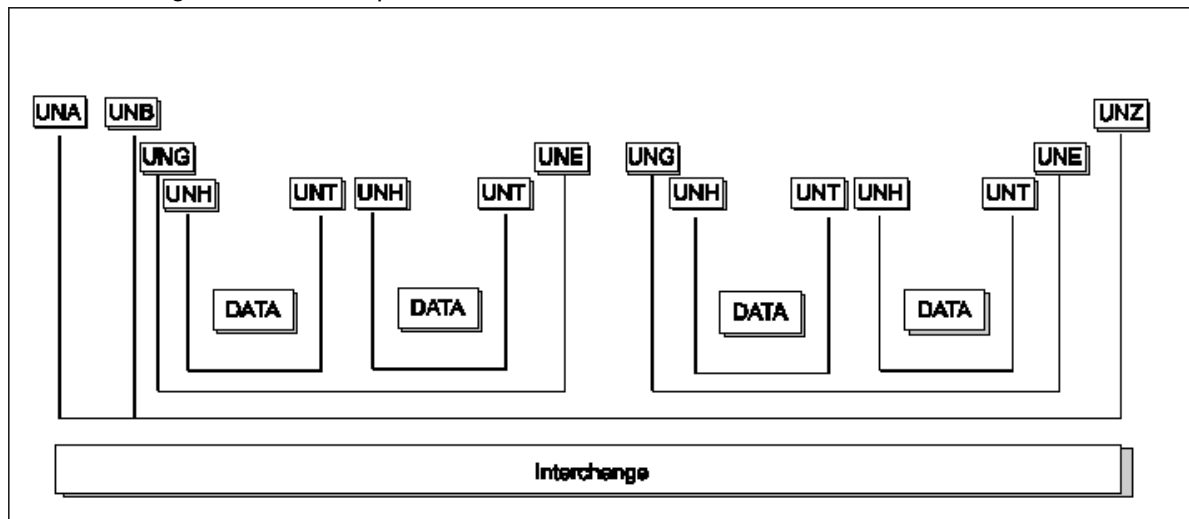
Segment UNA is dependent on the character set being used. If the EANCOM default character set A is being used then the UNA segment is not required.

Segments UNB..UNZ and UNH..UNT are mandatory.

Segments UNG..UNE are conditional. Within EANCOM the use of the UNG..UNE segments is not recommended as the grouping of same message types is not considered to add significant value to simply including multiple same message types within one interchange, i.e. between UNB..UNZ.

If the UNG..UNE segments are used then it should be noted that it is not possible in the EANCOM CONTRL message to syntactically report on a functional group.

The message itself is structured with a Header, a Detail and a Summary section. In messages where there may be ambiguity between the sections the UNS segment may be used as a separator.

The layout of the service segments UNA, UNB..UNZ, and UNG..UNE is presented in this section.

The segments UNH, UNS, and UNT being different for each message, are detailed in the message description part of this manual.

<u>Example of an interchange:</u>

An interchange contains two sets of messages: three despatch advices and two invoices. The interchange is sent on 2 January 1997 by a company identified by the EAN location number 5412345678908 to a company identified by the EAN location number 8798765432106.

UNA:+.? '

UNB+UNOA:3+5412345678908:14+8798765432106:14+970102:1000+12345555+++++EAN COMREF52'

....

UNH+66025+DESADV:D:96A:UN:EAN005'

.....

.....

UNT+35+66025'

UNH+66420+DESADV:D:96A:UN:EAN005'

.....

.....

UNT+26+66420'

UNH+1588+INVOIC:D:96A:UN:EAN008'

....

....

UNT+46+1588'

UNH+2063+INVOIC:D:96A:UN:EAN008'

....

....

UNT+87+2063'

UNH+67020+DESADV:D:96A:UN:EAN005'

.....

.....

UNT+102+67020'

....

UNZ+5+12345555'

EANCOM PART I

Segment Layout   - UNA segment

		EDIFACT	EAN	*	Description
UNA - C 1 - SERVICE STRING ADVICE					
Function: To head, identify and specify a message.					
UNA1	Component data element separator	M an1	M		Is used as a separator between component data elements contained within a composite data element (default value :)
UNA2	Data element separator	M			Is used to separate two simple or composite data elements (default value : +)
UNA3	Decimal notation	M an1	M		Is used to indicate the character used for decimal notation (default value : .)
UNA4	Release character	M an1	M		Used to restore the separator and the terminator signs to their original specification (default value: ?)
UNA5	Reserved for future use	M an1	M		(default value : space)
UNA6	Segment terminator	M an1	M		Used to indicate the end of segment data (default value : ')

Segment Notes:

This segment is used to inform the receiver of the interchange that a set of service string characters which are different to the default characters are being used.

When using the default set of service string characters the UNA segment need not be sent. When it is sent it must immediately precede the UNB segment and contain the six service string characters selected by the interchange sender.

Regardless of whether or not all of the service string characters are being changed every data element within this segment must be filled, i.e. if some default values are being used with user defined ones then both the default and user defined values must be specified.

When expressing the service string characters in the UNA segment it is not necessary to

include any element separators.

Segment Layout   - UNB segment

UNB - M 1 - INTERCHANGE HEADER				
Function: To start, identify and specify an interchange.				
		EDIFACT	EAN	*	Description
S001	SYNTAX IDENTIFIER	M	M		
0001	Syntax identifier	M a4	M	*	UNOA . UNOB . Controlling Agency (UNO = UNOC . UN/ECE) followed by character UNOD . set level UNOE . UNOF .
0002	Syntax version number	M n1	M	*	3 = Syntax version number 3
S002	INTERCHANGE SENDER	M	M		
0004	Sender identification	M an..35	M		EAN location number (n13)
0007	Partner identification code qualifier	C an..4	R	*	14 = EAN International
0008	Address for reverse routing	C an..14	O		
S003	INTERCHANGE RECIPIENT	M	M		
0010	Recipient identification	M an..35	M		EAN location number (n13)

0007	Partner identification code qualifier	C an..4	R	*	14 = EAN International
0014	Routing address	C an..14	O		
S004	DATE/TIME OF PREPARATION	M	M		
0017	Date	M n6	M		YYMMDD
0019	Time	M n4	M		HHMM
0020	Interchange control reference	M an..14	M		Unique reference identifying the interchange. Created by the interchange sender.
S005	RECIPIENT'S REFERENCE PASSWORD	M	O		
0022	Recipient's reference/password	M an..14	M		
0025	Recipient's reference/password qualifier	C n4	O		
0026	Application reference	C an..14	O		Message identification if the interchange contains only one type of message.
0029	Processing priority code	C a1	O		A = Highest priority
0031	Acknowledgement request	C n1	O		Message identification if the interchange contains only one type of message.
0032	Communications agreement identification	C an..35	O	*	EANCOM......
0035	Test indicator	C n1	O		1 = Interchange is a test

Segment Notes:

This segment is used to envelope the interchange and also to identify the party for whom the interchange is intended and the party who has sent the interchange. The principle of the UNB segment is the same as a physical envelope which covers one or more letters or documents and which details the address where delivery is to take place and the address from where the envelope has come.

DE 0001: The recommended (default) character set for use in EANCOM for international exchanges is character set A (UNOA). Should users wish to use character sets other than A then agreement on which set to use should be reached on a bi-lateral basis before communications begin.

DE 0004 and 0010: Within EANCOM the use of the EAN location number is recommended for the identification of the interchange sender and recipient.

DE 0008: The address for reverse routing is provided by the interchange sender to inform the interchange recipient of the address within the sender's system to which responding interchanges must be sent. It is recommended that EAN location numbers be used for this purpose.

DE 0014: The address for routing, which was originally provided by the interchange recipient, is used by the interchange sender to inform the recipient of the address of the recipients systems to which the interchange should be routed. The address provided here is normally informed to the interchange sender by the recipient in advance of the interchange transmission. It is recommended that EAN location numbers be used for this purpose.

DE S004: The date and time specified in this composite should be the date and time at which the interchange sender prepared the interchange. This date and time may not necessarily be the same as the date and time of contained messages.

DE 0020: The interchange control reference number is generated by the interchange sender and is used to uniquely identify each interchange. Should the interchange sender wish to re-use interchange control reference numbers it is recommended that each number be preserved for at least a period of three months before being re-used. In order to guarantee uniqueness the interchange control reference number should always be linked to the interchange sender's identification (DE 0004).

DE S005: The use of passwords must first be agreed bi-laterally by the parties exchanging the interchange.

DE 0026: This data element is used to identify the application on the interchange recipients system to which the interchange is directed. This data element may only be used if the interchange contains only one type of message, e.g. only invoices. The reference used in this data element is assigned by the interchange sender.

DE 0031: This data element is used to indicate whether an acknowledgement to the

interchange is required or not. The EANCOM CONTRL message should be used to provide acknowledgement of interchange receipt. In addition the EANCOM CONTRL message may be used to indicate when an interchange has been rejected due to syntactical errors.

DE 0032: This data element is used to identify any underlying agreements which control the exchange of data. Within EANCOM the identity of such agreements must start with the letters 'EANCOM' with the remaining characters within the data element filled according to bi-lateral agreements.

Segment Layout   - UNZ segment

UNZ - M 1 - INTERCHANGE TRAILER					
Function: To end and check the completeness of an interchange.					
		EDIFACT	EAN	*	Description
0036	Interchange control count	M n..6	M		Number of messages or functional groups within the interchange.
0020	Interchange control reference	M an..14	M		Identical to DE 0020 in UNB segment.
Segment Notes: This segment is used to provide the trailer of an interchange. DE 0036: If functional groups are used this is the number of functional groups within the interchange. If functional groups are not used this is the number of messages within the interchange.'					