L.2 STRUCTURE OF AN EDIFACT TRANSMISSION

This section is substantially based on the ISO 9735 document: "*EDIFACT application level syntax rules*", first released on 1988-07-15, amended and reprinted on 1990-11-01, and *Addendum 1* of 1992.

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

L.2.1 STRUCTURE OF AN INTERCHANGE

An interchange may consist of the following segments:

UNA	Service String Advice	Conditional*
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

* Mandatory if EDItEUR recommendations are followed.
** Not used in EDItEUR applications.

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

The UNA segment defines the separator characters used in the transmission, if they are not the default set for the character set defined in the UNB segment. See section L.2.2 below.

The UNB segment identifies the sender and receiver of the transmission, specifies the character set used, and carries other "housekeeping" data for the transmission. See section L.2.3 below.

The UNG and UNE segments are used only if the transmission carries several groups of messages of different types. *EDItEUR strongly recommends that a transmission should be limited to carrying only one group of messages of a single type. The UNG and UNE segments are not, therefore, used in EDItEUR practice.*

The UNZ segment ends the transmission. See section L.2.4 below.

The UNA, UNB and UNZ segments will normally be generated in outgoing transmissions, and processed in incoming transmissions, by a standard EDI software package. The user application need not be aware of their content. However, many implementers prefer to generate the whole of the EDIFACT transmission file as part of their own application software. Full specifications are given in this document.

L.2.2 CHARACTER SETS AND SERVICE STRING ADVICE: THE UNA SEGMENT

EDIFACT standards define a number of character sets, coded in the UNB segment as UNOA, UNOB, UNOC, UNOD etc. EDItEUR has adopted UNOC as the standard set for book and serials trading. This character set permits the representation of a full repertoire of special characters, including accents, for most European languages which use the Latin alphabet. It corresponds to the international standard character set ISO 8859.1.

EDItEUR recommends that the following five characters should be used as standard separators etc:

Apostrophe	'	=	segment terminator
Plus sign	+	=	segment tag and data element separator
Colon	:	=	component data element separator
Question mark	?	=	release character. Immediately preceding any of the special characters '+:? the release character returns it to its normal meaning. For example, 10?+10=20 means 10+10=20. Question mark is represented by ?? A release character is not counted when calculating the maximum length of the data element in which it occurs.
Period	.	=	decimal point. The period does not require a release character when it occurs as a normal text punctuation mark.

These are NOT the default separators for the UNOC character set. Consequently the UNA segment, as detailed below, must be sent at the start of each transmission to declare the separator values. Note that the segment has a spare character in position 5, "reserved for future use", which must be sent as a space.

UNA - C 1 - SERVICE STRING ADVICE				
Function : To define the characters selected for use as delimiters and indicators in the rest of the interchange that follows.				
	EDIFACT	EAN	*	Description
UNA1 Component data element separator	M an1	M		The separator between component data elements within a composite data element
UNA2 Data element separator	M an1	M		The separator between two simple or composite data elements
UNA3 Decimal notation	M an1	M		The character used for the decimal point
UNA4 Release character	M an1	M		The character used to restore separator and terminator signs to their normal meaning
UNA5 Reserved for future use	M an1	M		Sent as a space
UNA6 Segment terminator	M an1	M		The character used to indicate the end of a segment

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 is being used.

When the default set is used, the UNA segment need not be sent. When the UNA segment is sent, it must immediately precede the UNB segment and contain the six service string characters used in the interchange.

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

The UNA segment itself does not have any element separators.

UNA segment based on EDItEUR standard practice: :+.? '

Note that there is a space between the last two characters.

L.2.3 INTERCHANGE HEADER: THE 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	*	Controlling Agency (UNO = UN/ECE), followed by character set level. *Always UNOC in EDItEUR practice.*
0002	Syntax version number	M n1	M	*	3 = EDIFACT syntax version number 3
S002	**INTERCHANGE SENDER**	M	M		
0004	Sender identification	M an..35	M		EAN location number (n13) is preferred. *Alternatively a US book trade SAN may be used, or another mutually agreed ID.*
0007	Partner Identification code qualifier	C an..4	R		14 = EAN International *31B = US SAN Agency (EDItEUR code) 91 = assigned by supplier 92 = assigned by buyer*
0008	Address for reverse routing	C an..14	O		*Not normally used in EDItEUR practice.*
S003	**INTERCHANGE RECIPIENT**	M	M		
0010	Recipient identification	M an..35	M		EAN location number (n13) is preferred. *Alternatively a US book trade SAN may be used, or another mutually agreed ID.*
0007	Partner Identification code qualifier	C an..4	R		14 = EAN International *31B = US SAN Agency 91 = assigned by supplier 92 = assigned by buyer*
0014	Routing address	C an..14	O		*Not normally used in EDItEUR practice.*
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		Uniquely identifying the interchange. Created by the interchange sender.
S005	**RECIPIENT'S REFERENCE PASSWORD**	C	O		*Not normally used in EDItEUR practice.*
0022	Recipient's reference/password	M an..14	M		*Not normally used in EDItEUR practice.*
0025	Recipient's reference/password qualifier	C an2	O		*Not normally used in EDItEUR practice.*

UNB - M 1 - INTERCHANGE HEADER				
	EDIFACT	EAN	*	Description
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 *Not normally used in EDItEUR practice.*
0031 **Acknowledgement request**	C n1	O		*Not normally used in EDItEUR practice.*
0032 **Communications agreement ID**	C an..35	O	*	EANCOM...... *Not normally used in EDItEUR practice.*
0035 **Test indicator**	C n1	O		1 = Interchange is a test

The UNB 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 character set for EDItEUR applications is character set C (UNOC).

DE 0004 and 0010: The EAN location number is recommended for the identification of the interchange sender and recipient. Where an EAN number cannot reasonably be used, trading partners may use an SAN or any other mutually agreed number, with an appropriate value in DE 0007.

DE 0008: May specify an address within the sender's system to which any response should be routed. It is recommended that an EAN location number be used for this purpose. *Not normally used in EDItEUR applications.*

DE 0014: May specify an address within the recipient's system to which this transmission should be routed. It is recommended that an EAN location number be used for this purpose. *Not normally used in EDItEUR applications*.

DE S004: The date and time at which the sender prepared the transmission. This date and time may not necessarily be the same as the date and time of contained messages. The date in Composite S004, DE 0017, is not Year 2000 compliant, since it carries only a two digit year. (This will be changed in Version 4 of EDIFACT syntax, which has not yet become available for general adoption.) EDIFACT has determined that it will not take any specific action in respect of this data element in Version 3 syntax, since it is considered that in practice no ambiguity will arise. However, users should verify that any software which uses the UNB segment, including third-party EDI software, will interpret the two-digit year correctly.

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

DE S005: Any use of passwords must first be agreed bilaterally by the parties to the exchange. *Not normally used in EDItEUR applications*.

DE 0026: The application reference is used to identify the application on the recipient's system to which the interchange is directed. This data element may be used if, and only if, the interchange contains only one type of message, eg invoices. The reference used in this data element is assigned by the interchange sender. *In EDItEUR applications it is recommended that (a) each transmission should carry only one type of message, and (b) DE 0026 should carry the six-character EDIFACT name for the message, eg INVOIC or ORDRSP.*

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. *Not normally used in EDItEUR applications.*

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 bilateral agreements. *Not normally used in EDItEUR applications.*

Example of a UNB segment based on EDItEUR recommended practice:

`UNB+UNOC:3+5012345678901:14+4598765432198:14+000316:1402+INV73529++INVOIC'`

L.2.4 INTERCHANGE TRAILER: THE 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 notes				
This segment is used to end 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.				

Example of a UNZ segment:

`UNZ+3+INV73529'`   Three messages in this interchange.

L.2.5 STRUCTURE OF A MESSAGE

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

(a)   **Header section** - A segment occurring in this section relates to the *entire message.*

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

(c)　**Summary section** - Only segments containing *totals* or *control information* may occur in the summary section, eg 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:

```
Header section

UNH.....
BGM.....
............

Detail section

LIN.......
QTY.....
...........

Summary section

CNT.....
UNT.....
```
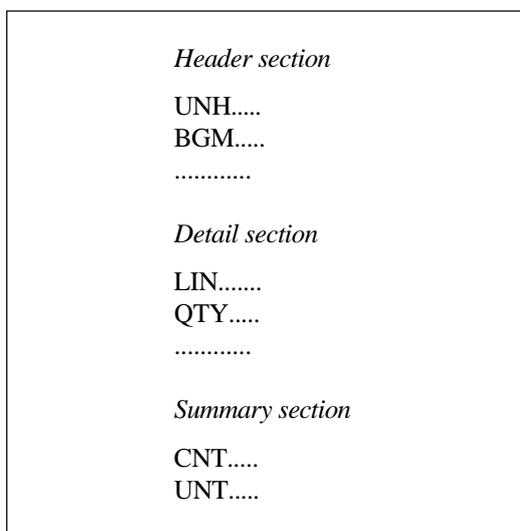
The same segment type may occur in more than one of the message sections, eg in the header and in the detail section, and/or more than once in the same section.

Some segments may be repeated a certain number of times at their location in the message. The status, *mandatory* or *conditional*, and the maximum number of repeats are indicated in the message specification.

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

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

L.2.6 SEGMENT STRUCTURE

A segment consists of:

- A segment tag, which 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 element, which carries a code value giving specific meaning to the data element which it qualifies. The qualifier is taken from an agreed set of code values.

Example of an EDIFACT segment:

```
DTM+137:19940101:102'
```

DTM	=	segment tag, identifying the "Date/time/period" segment
+	=	segment tag and data element separator
137	=	date qualifier, indicating that the date is the document/message date/time
:	=	separator for component data elements within a composite
19940101	=	date, in the format specified by the date format qualifier
:	=	separator for data elements within a composite
102	=	date format qualifier, indicating the format of the date (CCYYMMDD)
'	=	segment terminator.

L.2.7 COMPRESSION OF DATA

In data elements for which the EDIFACT Directory specifies variable length and no other restrictions, non-significant character positions (eg leading zeroes and trailing spaces) are suppressed.

In the following examples, TAG = segment tag; DE = data element; CE = component data element.

Exclusion of segments. Conditional segments containing no data are omitted.

Exclusion of data elements by omission. Data elements are identified by their sequential position within the segment as stated in the EDIFACT Directory. If a conditional data element is omitted and followed by another data element, its position must be indicated by retention of its data element separator.

Example:

`TAG+DE+DE+DE+CE:CE:CE'`	Complete segment including all data elements
`TAG+DE++DE+CE:CE:CE'`	The second data element 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 is truncated by placing the segment terminator after the last data element for which data is present.

Example:

`TAG+DE+DE+DE+DE'`	Segment including all data elements
`TAG+DE+DE'`	Segment truncated after the first two data elements

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.

Example:

`TAG+DE+DE+CE:CE:CE'`	Segment including all CE's
`TAG+DE+DE+CE::CE'`	Segment with the penultimate CE omitted

Exclusion of component data elements by truncation. One or more conditional CE's 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.

Example:

`TAG+DE+DE+CE:CE:CE'`	Segment including last CE
`TAG+DE+DE+CE:CE'`	Segment truncated by the segment separator

L.2.8 REPRESENTATION OF NUMERIC VALUES

Decimal sign. The decimal sign is represented by a full point on the line (.). The decimal sign is not counted as a character when computing the maximum field length of a data element. When a decimal sign is transmitted, there must 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 digits	Decimal digits
Amount (monetary)	n..18	15	3
Control value	n..18	14	4
Cube	n..9	5	4
Currency rate	n..12	6	6
Other range value	n..18	15	3
Percentage	n..10	6	4
Percentage range value	n..18	14	4
Quantity	n..15	12	3
Rate per unit	n..15	12	3
Tax rate	n..17	13	4
Unit price	n..15	11	4
Unit price basis	n..9	6	3
Weight	n..18	15	3

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

Sign. Numeric data element values are to be sent as positive. Although conceptually a deduction is negative, it is represented by a positive value: eg in a credit note all values are sent as positive amounts, and the application software will take note of the message name code (DE 1001) and process the values accordingly. In addition, some data element and code combinations will lead to implied negative values, eg DE 5463 with code value "A" (allowance) in an ALC segment in an invoice. Again, the values are sent as positive amounts.

If, however, a value has to be explicitly represented as negative, it must be sent immediately preceded by a minus sign, eg -112. The minus sign is not counted as a character when computing the maximum field length of the data element.

L.2.9 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

Layout of message subset specifications

In sections L.3 onwards, the message subsets required for this application are defined and illustrated.

To keep the application guidelines as clear and compact as possible, while making it easy to cross-refer to the full specifications if so required, the following conventions have been used:

(a)   Only those segments which are used in this application are included in the message definition.

(b)   Each segment is identified by its EANCOM segment number, and by its EDIFACT segment code (eg **UNH**) and name (eg **MESSAGE HEADER**).

(c)   Each data element or composite data element is identified by its EDIFACT code number (eg 1001) and name (eg Document/message code).

(d)   The status of each data element *for the purposes of this application* is shown by one of the following codes:

   M   Mandatory. This includes data elements which are *mandatory* in EDIFACT; data elements which, although not mandatory in EDIFACT, are *required* in EANCOM; and data elements which EDItEUR considers to be *required for the application*, although both EDIFACT and EANCOM treat them as optional or conditional.

   D   Dependent. Data elements which are used only under specified conditions, when they may in fact become mandatory.

   O   Optional. Data elements whose presence is optional, ie to be agreed between trading partners.

   N   Not used. This category includes data elements which are *not used* in EANCOM, and data elements which are *optional* or *dependent* in EANCOM, but which EDItEUR considers to be unnecessary for the application.

(e)   Where a data element is not used in this application, details of its format are omitted.

(f)   Where a composite data element is not used in this application, its constituent data elements are not listed.