INTERNATIONAL STANDARD

ISO 10303-21

Second edition
2002-01-15

Industrial automation systems and integration — Product data representation and exchange —

Part 21: Implementation methods: Clear text encoding of the exchange structure

Systèmes d'automatisation industrielle et intégration — Représentation et échange de données de produits —

Partie 21: Méthodes de mise en application: Encodage en texte clair des fichiers d'échange

Reference number
ISO 10303-21:2002(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

Contents

page

Tables page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardizations.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this part of ISO 10303 may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

International Standard ISO 10303-21 was prepared by Technical Committee ISO/TC 184, *Industrial automation systems and integration*, Subcommittee SC 4, *Industrial data*.

This second edition cancels and replaces the first edition (ISO 10303-21:1994), of which it constitutes a technical revision. It incorporates the corrections published in ISO 10303-21:1994/Cor.1:1996.

This International Standard is organized as a series of parts, each published separately. The structure of this International Standard is described in ISO 10303-1.

Each part of this International Standard is a member of one of the following series: description methods, implementation methods, conformance testing methodology and framework, integrated generic resources, integrated application resources, application protocols, abstract test suites, application interpreted constructs, and application modules. This part is a member of the implementation methods series.

A complete list of parts of ISO 10303 is available from the Internet:

<http://www.nist.gov/sc4/editing/step/titles/>

Annexes A, B, C, D, E and F form a normative part of this part of ISO 10303. Annexes G and H are for information only.

Introduction

ISO 10303 is an International Standard for the computer-interpretable representation of product information and for the exchange of product data. The objective is to provide a neutral mechanism capable of describing products throughout their life cycle. This mechanism is suitable not only for neutral file exchange, but also as a basis for implementing and sharing product databases, and as a basis for archiving.

This part of ISO 10303 specifies a mechanism that allows product data described in the EXPRESS language, specified in ISO 10303-11, to be transferred from one computer system to another.

Major subdivisions in this part of ISO 10303 are:

— specification of the exchange structure syntax;

— mapping from an EXPRESS schema onto this syntax.

NOTE The examples of EXPRESS usage in this part of ISO 10303 do not conform to any particular style rules. Indeed, the examples sometimes use poor style to conserve space or to concentrate on the important points. The examples are not intended to reflect the content of the information models defined in other parts of this International Standard. They are crafted to show particular features of EXPRESS or of the exchange structure. Many examples are annotated in a way that is not consistent with the syntax rules of this part of ISO 10303. These annotations are introduced by symbolic arrows, either horizontal '-->', or vertical. These annotations should be ignored when considering the parse rules. Any similarity between the examples and the normative models specified in other parts of this International Standard should be ignored. Several mapping examples have been provided throughout this document. Additional *spaces* and new lines have been inserted into some of these examples to aid readability. These *spaces* and new lines need not appear in an exchange structure.

This edition incorporates the following technical modifications to ISO 10303-21:1994:

— the SCOPE structure (&SCOPE / ENDSCOPE) has been eliminated;

— the exchange structure may now contain multiple data sections;

— the exchange structure header section may now identify the default language for string attributes of entity instances encoded in a data section;

— the exchange structure header section may now identify information describing contexts within which the entity instances encoded in a data section are applicable;

— enumeration values may now be encoded using short names if such names are available.

All exchange structures that are encoded according to the previous edition of ISO 10303-21 and that do not use the SCOPE structure also conform to this edition.

Industrial automation systems and integration — Product data representation and exchange — Part 21: Implementation methods: Clear text encoding of the exchange structure

1 Scope

This part of ISO 10303 specifies an exchange structure format using a clear text encoding of product data for which the conceptual model is specified in the EXPRESS language (ISO 10303-11). The exchange format is suitable for the transfer of product data among computer systems.

The mapping from the EXPRESS language to the syntax of the exchange structure is specified. Any EXPRESS schema can be mapped onto the exchange structure syntax.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this part of ISO 10303. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this part of ISO 10303 are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 639-2:1998, *Codes for the representation of names of languages — Part 2: Alpha-3 code.*

ISO 3788:1990, *Information processing — 9-Track, 12,7 mm (0,5 in) wide magnetic tape for information interchange using phase encoding at 126 ftpmm (3 200 ftpi) — 63 cpmm (1 600 cpi).*

ISO 8601:2000, *Data elements and interchange formats — Information interchange — Representation of dates and times.*

ISO/IEC 8824-1:1998, *Information technology — Abstract Syntax Notation One (ASN.1): Specification of basic notation.*

ISO/IEC 8859-1:1998, *Information technology — 8 bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1.*

ISO/IEC 8859-2:1999, *Information technology — 8 bit single-byte coded graphic character sets — Part 2: Latin alphabet No. 2.*

ISO/IEC 8859-3:1999, *Information technology — 8 bit single-byte coded graphic character sets — Part 3: Latin alphabet No. 3.*

ISO/IEC 8859-4:1998, *Information technology — 8 bit single-byte coded graphic character sets — Part 4: Latin alphabet No. 4.*

ISO/IEC 8859-5:1999, *Information technology — 8 bit single-byte coded graphic character sets — Part 5: Latin/Cyrillic alphabet.*

ISO/IEC 8859-6:1999, *Information technology — 8 bit single-byte coded graphic character sets — Part 6: Latin/Arabic alphabet.*

ISO 8859-7:1987, *Information processing — 8 bit single-byte coded graphic character sets — Part 7: Latin/Greek alphabet.*

ISO/IEC 8859-8:1999, *Information technology — 8 bit single-byte coded graphic character sets — Part 8: Latin/Hebrew alphabet.*

ISO/IEC 8859-9:1999, *Information technology — 8 bit single-byte coded graphic character sets — Part 9: Latin alphabet No. 5.*

ISO 10303-1:1994, *Industrial automation systems and integration — Product data representation and exchange — Part 1: Overview and fundamental principles.*

ISO 10303-11:1994, *Industrial automation systems and integration — Product data representation and exchange — Part 11: Description methods: The EXPRESS language reference manual.*

ISO/IEC 10646-1:2000, *Information Processing — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane.*

3 Terms, definitions, and abbreviations

3.1 Terms defined in ISO 8859-1

This part of ISO 10303 makes use of the following terms defined in ISO 8859-1.

— byte;

— character;

— graphic character.

3.2 Terms defined in ISO 10646

This part of ISO 10303 makes use of the following term defined in ISO 10646.

— basic multilingual plane.

3.3 Terms defined in ISO 10303-1

This part of ISO 10303 makes use of the following terms defined in ISO 10303-1.

— application protocol;

— exchange structure.

3.4 Terms defined in ISO 10303-11

This part of ISO 10303 makes use of the following terms defined in ISO 10303-11.

— complex entity instance;

— data type;

— entity;

— partial complex entity instance;

— simple entity instance;

— token.

3.5 Other definitions

For the purposes of this part of ISO 10303, the following definitions apply.

3.5.1
basic alphabet
the set of characters G(02/00) through G(07/14) of ISO 8859-1.

3.5.2
clear text encoding
the encoding of information, using a sequence of codes for characters in the basic alphabet.

3.5.3
control directive
a sequence of characters in the basic alphabet.

3.5.4
keyword
a special sequence of characters identifying an entity or a defined type in the exchange structure.

3.5.5
section
a collection of data of the same functional category of information.

3.5.6
sequential file
a file that can only be accessed in a sequential manner.

3.5.7
token separator
a sequence of one or more 8-bit bytes that separate any two tokens.

3.6 Abbreviations

For the purposes this part of ISO 10303, the following abbreviations apply:

BMP	Basic multilingual plane
WSN	Wirth Syntax Notation

4 Exchange structure fundamental concepts and assumptions

4.1 Introduction

The exchange structure is described by an unambiguous, context-free grammar to facilitate parsing by software. The grammar is expressed in Wirth Syntax Notation that is described in annex B. The form of product data in the exchange structure is specified using a mapping from the EXPRESS language to the exchange structure syntax.

4.2 Notational and typographical conventions

Any *quotation marks* used in this part of ISO 10303 are not part of the text that appears in the exchange structure but serve to delimit that text. This statement applies to all places in the text where *quotation marks* are used. Table 2, Table 3, and Table 4 form an exception to this rule as the *quotation marks* used in those tables form part of the WSN rules.

In ISO 8859, each character is assigned an identifying name. When that name is used in this part of ISO 10303, it is typeset in *italics* to distinguish if from ordinary text. Thus *comma* is used to refer to ",", *low line* refers to "_", and *capital letter A* refers to "A".

Within examples in this part of ISO 10303, an annotation is introduced by the sequence `----->` where clarification is required.

4.3 Conformance

Two levels of conformance are specified:

— syntactical conformance of the exchange structure: an exchange structure conforms to ISO 10303-21 if the requirements of this part of ISO 10303 are satisfied;

— schema conformance of the exchange structure: the instances represented in the exchange structure conform to the schemas listed in the header section of the exchange structure if every requirement or constraint of these schemas is satisfied with respect to each instance or grouping of instances to which it shall apply whatsoever and the mapping requirements defined in clauses 9 and 10 of this part of ISO 10303 are satisfied.

NOTE Annex F presents methods for evaluating schema conformance when an exchange structure contains multiple data sections based on different EXPRESS schemas.

Syntactical conformance is a prerequisite for schema conformance.

Two classes of syntactical conformance are defined by this part of ISO 10303, depending on the method chosen for the encoding of complex entity instances (see 10.2.5). An implementation that claims syntactical conformance to this part of ISO 10303 shall read or write files or both that exhibit syntactical conformance in (at least) one of these two conformance classes.

An implementation that claims schema conformance to this part of ISO 10303 shall read or write files or both that exhibit schema as well as syntactical conformance.

5 Formal definitions

5.1 Formal notation

Wirth Syntax Notation (WSN) is used in this part of ISO 10303 to specify the syntax of the exchange structure in a formal notation. WSN is described in annex B.

5.2 Basic alphabet definition

The alphabet of the exchange structure is defined as the characters from G(02/00) to G(07/14) of ISO 8859-1. This alphabet is represented in the exchange structure by the set of 8-bit bytes with decimal values 32 to 126. Table 1 divides the basic alphabet into subsets. G(x/y) is a notation for the character in position (16 times x) + y in the code table in ISO 8859-1.

NOTE Table D.1 gives the correspondence between the 8-bit bytes and their graphic representation in ISO 8859-1.

Table 1 - WSN defining subsets of the basic alphabet

```
SPACE     = " " .

DIGIT     = "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7"
          | "8" | "9" .

LOWER     = "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h"
          | "i" | "j" | "k" | "l" | "m" | "n" | "o" | "p"
          | "q" | "r" | "s" | "t" | "u" | "v" | "w" | "x"
          | "y" | "z" .

UPPER     = "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H"
          | "I" | "J" | "K" | "L" | "M" | "N" | "O" | "P"
          | "Q" | "R" | "S" | "T" | "U" | "V" | "W" | "X"
          | "Y" | "Z" | "_" .

SPECIAL   = "!" | """" | "*" | "$" | "%" | "&" | "." | "#"
          | "+" | "," | "-" | "(" | ")" | "?" | "/" | ":"
          | ";" | "<" | "=" | ">" | "@" | "[" | "]" | "{"
          | "|" | "}" | "^" | "`" | "~" .

REVERSE_SOLIDUS  = "\" .

APOSTROPHE = "'" .

CHARACTER = SPACE | DIGIT | LOWER | UPPER | SPECIAL
          | REVERSE_SOLIDUS | APOSTROPHE
```


5.3 Exchange structure

The exchange structure shall be a sequential file using a clear text encoding. The exchange structure shall consist of at least two sections: the header section and one or more data sections. The header section provides data relating to the exchange structure itself. The structure of the header section is specified in clause 8. The data section provides the data to be transferred. The structure of the data section is specified in clause 9. The exchange structure is defined by the WSN in Table 3.

The exchange structure is a stream of 8-bit bytes that are encodings of the graphic characters of the basic alphabet. The graphic characters are collected into recognizable sequences called tokens. Tokens may be separated by token separators. The exchange structure can be considered as a sequence of tokens and token separators.

5.4 Definition of tokens

The tokens used in the exchange structure are defined by the WSN in Table 2.

5.5 WSN of the exchange structure

The syntax of the exchange structure is specified in Table 3. Table 3 references the tokens defined in Table 2. The relationship between the syntax and the EXPRESS schema is specified in clause 10.

Table 2 - WSN of token definitions

```
KEYWORD             = USER_DEFINED_KEYWORD | STANDARD_KEYWORD .

USER_DEFINED_KEYWORD = "!" UPPER { UPPER | DIGIT } .

STANDARD_KEYWORD    = UPPER { UPPER | DIGIT } .

SIGN                = "+" | "-" .

INTEGER             = [ SIGN ] DIGIT { DIGIT } .

REAL                = [ SIGN ] DIGIT { DIGIT } "." { DIGIT }
                      [ "E" [ SIGN ] DIGIT { DIGIT } ] .

NON_Q_CHAR          = SPECIAL | DIGIT | SPACE | LOWER | UPPER .

STRING              = "'" { NON_Q_CHAR |
                      APOSTROPHE APOSTROPHE |
                      REVERSE_SOLIDUS REVERSE_SOLIDUS |
                      CONTROL_DIRECTIVE } "'" .

ENTITY_INSTANCE_NAME = "#" DIGIT { DIGIT } .

ENUMERATION         = "." UPPER { UPPER | DIGIT } "." .

HEX                 = "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" |
                      "8" | "9" | "A" | "B" | "C" | "D" | "E" | "F" .

BINARY              = """" ( "0" | "1" | "2" | "3" ) { HEX } """" .
```


5.6 Token separators

A token separator is a element that separates two tokens. Token separators are *space*, the explicit print control directives, and comments. A token separator may appear between the terminals or non-terminals of the productions of Table 3. Any number of token separators may appear wherever one token separator may appear. A token separator shall not appear within tokens except that explicit print control directives may also appear within binaries and within strings. Print control directives are defined in clause 11.

NOTE *Space* is the only whitespace character within the basic alphabet described in 5.2. Line-delimiters such as *line feed* or *carriage return* are permitted in the exchange structure by annex A, but do not belong to the basic alphabet and are required by annex A to be ignored when processing the exchange structure. Consequently, line breaks may appear anywhere within the structure, including within tokens.

A comment shall be encoded as a *solidus asterisk* "/*" followed by any number of characters from the basic alphabet, and terminated by an *asterisk solidus* "*/". Any occurrence of *solidus asterisk* following the first occurrence shall not be significant, i.e. comments cannot be nested. All graphic characters appearing inside a comment shall not be significant to the exchange structure and are only intended to be read by humans.

