INTERNATIONAL STANDARD

ISO/IEC 16022

Second edition
2006-09-15

Information technology — Automatic identification and data capture techniques — Data Matrix bar code symbology specification

Technologies de l'information — Techniques d'identification automatique et de capture des données — Spécification de symbologie de code à barres Data Matrix

Reference number
ISO/IEC 16022:2006(E)

Contents

Page

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

ISO/IEC 16022 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 31, *Automatic identification and data capture techniques*.

This second edition cancels and replaces the first edition (ISO/IEC 16022:2000), which has been technically revised. It also incorporates the Technical Corrigendum ISO/IEC 16022:2000/Cor.1:2004.

Introduction

Data Matrix is a two-dimensional matrix symbology which is made up of nominally square modules arranged within a perimeter finder pattern. Though primarily shown and described in this International Standard as a dark symbol on light background, Data Matrix symbols can also be printed to appear as light on dark.

Manufacturers of bar code equipment and users of the technology require publicly available standard symbology specifications to which they can refer when developing equipment and application standards. The publication of standardised symbology specifications is designed to achieve this.

Information technology — Automatic identification and data capture techniques — Data Matrix bar code symbology specification

1 Scope

This International Standard defines the requirements for the symbology known as Data Matrix. It specifies the Data Matrix symbology characteristics, data character encodation, symbol formats, dimensions and print quality requirements, error correction rules, decoding algorithm, and user-selectable application parameters.

It applies to all Data Matrix symbols produced by any printing or marking technology.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 15424, *Information technology — Automatic identification and data capture techniques — Data Carrier Identifiers (including Symbology Identifiers)*

ISO/IEC 19762-1, *Information technology — Automatic identification and data capture (AIDC) techniques — Harmonized vocabulary — Part 1: General terms relating to AIDC*

ISO/IEC 19762-2, *Information technology — Automatic identification and data capture (AIDC) techniques — Harmonized vocabulary — Part 2: Optically readable media (ORM)*

ISO/IEC 15415, *Information technology — Automatic identification and data capture techniques — Bar code print quality test specification — Two-dimensional symbols*

ISO/IEC 15416, *Information technology — Automatic identification and data capture techniques — Bar code print quality test specification — Linear symbols*

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*

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

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

AIM Inc. ITS/04-001 International Technical Standard: *Extended Channel Interpretations — Part 1: Identification Schemes and Protocol*

3 Terms, definitions, symbols and mathematical/logical notations

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19762-1, ISO/IEC 19762-2 and the following apply.

3.1.1
codeword
symbol character value, an intermediate level of coding between source data and the graphical encodation in the symbol

3.1.2
module
single cell in a matrix symbology used to encode one bit of data, nominally a square shape in Data Matrix

3.1.3
convolutional coding
error checking and correcting (ECC) algorithm that processes a set of input bits into a set of output bits that can recover from damage by breaking the input bits into blocks, then convolving each input block with the contents of a multi-stage shift register to produce protected output blocks

NOTE These encoders can be constructed in hardware using input and output switches, shift registers, and exclusive-or (XOR) gates.

3.1.4
pattern randomising
procedure to convert an original bit pattern to another bit pattern, intended to reduce the probability of repeating patterns occurring in the symbol, by inverting selected bits

3.2 Symbols

For the purposes of this document, the following mathematical symbols apply unless defined locally.

d number of error correction codewords

e number of erasures

k (for ECC 000 - 140) the number of bits in a complete segment input to the state machine to generate the convolutional code (for ECC 200) total number of error correction codewords

m the memory order of the convolutional code

n (for ECC 000 - 140) the number of bits in a complete segment generated by the state machine producing the convolutional code (for ECC 200) total number of data codewords

N the numerical base in an encodation scheme

p number of codewords reserved for error detection

S symbol character

t number of errors

u the input bit segment to the state machine, taken k bits at a time

v the output bit segment from the state machine, generated n bits at a time

X horizontal and vertical width of a module

ε error correction codeword

3.3 Mathematical/logical notations

For the purposes of this document, the following notations and mathematical operations apply.

div integer division operator

mod integer remainder after division

XOR exclusive-or logic function whose output is one only when its two inputs are not equivalent.

LSB least significant bit

MSB most significant bit

4 Symbol description

4.1 Basic characteristics

Data Matrix is a two-dimensional matrix symbology.

There are two types:

ECC 200 which uses Reed-Solomon error correction. ECC 200 is recommended for new applications.

ECC 000 - 140 with several available levels of convolutional error correction, referred to as ECC 000, ECC 050, ECC 080, ECC 100 and ECC 140 respectively. ECC 000 - 140 should only be used in closed applications where a single party controls both the production and reading of the symbols and is responsible for overall system performance.

The characteristics of Data Matrix are:

a) Encodable character set:

   1) values 0 – 127 in accordance with the US national version of ISO/IEC 646

NOTE 1 This version consists of the G0 set of ISO/IEC 646 and the C0 set of ISO/IEC 6429 with values 28 – 31 modified to FS, GS, RS and US respectively.

   2) values 128 - 255 in accordance with ISO 8859-1. These are referred to as extended ASCII.

b) Representation of data: A dark module is a binary one and a light module is a zero.

NOTE 2 This International Standard specifies Data Matrix symbols in terms of dark modules marked on a light background. However, subclause 4.2 provides that symbols may also be produced with light and dark modules reversed in colour (see 4.2), and in such symbols references in this International Standard to dark modules should be taken as references to light modules, and vice versa.

c) Symbol size in modules (not including quiet zone):

ECC 200	10 x 10 to 144 x 144 even values only
ECC 000 – 140	9 x 9 to 49 x 49, odd values only

d) Data characters per symbol (for maximum symbol size in ECC200):

   1) Alphanumeric data: up to 2 335 characters

   2) 8-bit byte data: 1 555 characters

   3) Numeric data: 3 116 digits.

e) Selectable error correction:

   ECC 200: Reed-Solomon error correction.

   ECC 000 - 140: Four levels of convolutional error correction, plus the option to apply only error detection

f) Code type: Matrix

g) Orientation independence: Yes

4.2 Summary of additional features

The following summarises additional features which are inherent or optional in Data Matrix:

a) Reflectance reversal: (Inherent): Symbols are intended to be read when marked so that the image is either dark on light or light on dark (see Figure 1). The specifications in this International Standard are based on dark images on a light background, therefore references to dark or light modules should be taken as references to light or dark modules respectively in the case of symbols produced with reflectance reversal.

b) Extended Channel Interpretations: (ECC 200 only, optional): This mechanism enables characters from other character sets (e.g. Arabic, Cyrillic, Greek, Hebrew) and other data interpretations or industry-specific requirements to be represented.

c) Rectangular symbols: (ECC 200 only, optional): Six symbol formats are specified in a rectangular form.

d) Structured append: (ECC 200 only, optional): This allows files of data to be represented in up to 16 Data Matrix symbols. The original data can be correctly reconstructed regardless of the order in which the symbols are scanned.

4.3 Symbol structure

Each Data Matrix symbol consists of data regions which contain nominally square modules set out in a regular array. In larger ECC 200 symbols, data regions are separated by alignment patterns. The data region, or set of data regions and alignment patterns, is surrounded by a finder pattern, and this shall in turn be surrounded on all four sides by a quiet zone border. Figure 1 illustrates an ECC 140 and two representations of an ECC 200 symbol.

(a) ECC200, dark on light (b) ECC200, light on dark (c) ECC140, dark on light

Figure 1 — ECC 200 (a & b) and ECC 140 (c) encoding "A1B2C3D4E5F6G7H8I9J0K1L2"

4.3.1 Finder pattern

The finder pattern is a perimeter to the data region and is one module wide. Two adjacent sides, the left and lower sides, forming the L boundary, are solid dark lines; these are used primarily to determine physical size, orientation and symbol distortion. The two opposite sides are made up of alternating dark and light modules. These are used primarily to define the cell structure of the symbol, but also can assist in determining physical size and distortion. The extent of the quiet zone is indicated by the corner marks in Figure 1.

4.3.2 Symbol sizes and capacities

ECC 200 symbols have an even number of rows and an even number of columns. Some symbols are square with sizes from 10 x 10 to 144 x 144 not including quiet zones. Some symbols are rectangular with sizes from 8 x 18 to 16 x 48 not including quiet zones. All ECC 200 symbols can be recognised by the upper right corner module being light. The complete attributes of ECC 200 symbols are given in Table 7 in Section 5.5.

ECC 000 - 140 symbols have an odd number of rows and an odd number of columns. Symbols are square with sizes from 9 x 9 to 49 x 49 (modules) not including quiet zones. These symbols can be recognised by the upper right corner module being dark. The complete attributes of ECC 000 - 140 symbols are given in Annex G.

5 ECC 200 requirements

5.1 Encode procedure overview

This section provides an overview of the encoding procedure. Following sections will provide more details. An encoding example for ECC 200 is given in Annex O. The following steps convert user data to an ECC 200 symbol:

Step 1: Data encodation

> Analyse the data stream to identify the variety of different characters to be encoded. ECC 200 includes various encodation schemes which allow a defined set of characters to be converted into codewords more efficiently than the default scheme. Insert additional codewords to switch between the encodation schemes and to perform other functions. Add pad characters as needed to fill the required number of codewords. If the user does not specify the matrix size, then choose the smallest size that accommodates the data. A complete list of matrix sizes is shown in Section 5.5, Table 7.

Table 1 — Encodation schemes for ECC 200

Encodation scheme	Characters	Bits per data character
ASCII	double digit numerics	4
	ASCII values 0 - 127	8
	Extended ASCII values 128 - 255	16
C40	Upper-case alphanumeric	5,33
	Lower case and special characters	10,66[a]
Text	Lower-case alphanumeric	5,33
	Upper case and special characters	10,66[b]
X12	ANSI X12 EDI data set	5,33
EDIFACT	ASCII values 32 - 94	6
Base 256	All byte values 0 - 255	8

[a] encoded as two C40 values as result of use of a shift character

[b] encoded as two Text values as result of use of a shift character

Step 2: Error checking and correcting codeword generation

For symbols with more than 255 codewords, sub-divide the codeword stream into interleaved blocks to enable the error correction algorithms to be processed as shown in Annex A. Generate the error correction codewords for each block. The result of this process expands the codeword stream by the number of error correction codewords. Place the error correction codewords after the data codewords.

Step 3: Module placement in matrix

Place the codeword modules in the matrix. Insert the alignment pattern modules, if any, in the matrix. Add the finder pattern modules around the matrix.

5.2 Data encodation

5.2.1 Overview

The data may be encoded using any combination of six encodation schemes (see Table 1). ASCII encodation is the basic scheme. All other encodation schemes are invoked from ASCII encodation and return to this scheme. The compaction efficiencies given in Table 1 need to be interpreted carefully. The best scheme for a given set of data may not be the one with the fewest bits per data character. If the highest degree of compaction is required, account has to be taken of switching between encodation schemes and between code sets within an encodation scheme (see Annex P). It should also be noted that even if the number of codewords is minimised, the codeword stream might need to be expanded to fill a symbol. This fill process is done using pad characters.

5.2.2 Default character interpretation

The default character interpretation for character values 0 to 127 shall conform to ISO/IEC 646. The default character interpretation for character values 128 to 255 shall conform to ISO 8859-1: Latin Alphabet No. 1. The graphical representation of data characters shown throughout this document complies with the default interpretation. This interpretation can be changed using Extended Channel Interpretation (ECI) escape sequences, see 5.4. The default interpretation corresponds to ECI 000003.