TECHNICAL SPECIFICATION

# ISO/TS 12029

Second edition
2007-03-15

# Electronic imaging — Forms design optimization for electronic image management

*Imagerie électronique — Optimisation de conception de formulaires pour la gestion d'images électroniques*



Reference number
ISO/TS 12029:2007(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.org
Web www.iso.org

Published in Switzerland

 

# Contents

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 standardization.

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

The main task of technical committees is to prepare International Standards. 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.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.



An ISO/PAS or ISO/TS is reviewed after three years in order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn. If the ISO/PAS or ISO/TS is confirmed, it is reviewed again after a further three years, at which time it must either be transformed into an International Standard or be withdrawn.

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

ISO/TS 12029 was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Application issues*.

This second edition cancels and replaces the first edition (ISO/TS 12029:2002), which has been technically revised.

# Introduction

This Technical Specification discusses issues and provides guidance for the design of forms that are used for electronic capture of information. Features include

— colour dropouts,

— type fonts,

— printing screen tints,

— line width,

— data storage, and

— other interrelated issues.

It is necessary to balance conflicting requirements of user-friendliness and electronic capture. Making a form appealing by use of colour or graphics could assist users when they complete the form, but could also decrease the form's scannability or other automated related functions. This conflict might require compromise in design of a form.



While this Technical Specification focuses on electronic forms' design and structure, it should be noted that paper based forms' design and structure can have different characteristics which ensure usability and readability. It is advisable that the user use these specifications when developing electronic forms while keeping in mind that paper based forms can be easily replicated in an electronic format (with the same content as in the paper based form), but with differing fonts and spacing.

# Electronic imaging — Forms design optimization for electronic image management

## 1 Scope

This Technical Specification provides guidelines for the design of forms to be completed by users and scanned for processing by electronic image management (EIM) systems. These guidelines are limited to forms using roman characters.

## 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 1073-1, *Alphanumeric character sets for optical recognition — Part 1: Character set OCR-A — Shapes and dimensions of the printed image*

ISO 1073-2, *Alphanumeric character sets for optical recognition — Part 2: Character set OCR-B — Shapes and dimensions of the printed image*

ISO 12651, *Electronic imaging — Vocabulary*



## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 12651 and the following apply.

**3.1**
**alphanumeric**
pertaining to a character set that contains letters, numbers and other characters, such as punctuation marks and symbols

NOTE See ISO/IEC 2382-4.

**3.2**
**dropout ink**
ink of a colour that cannot be detected by a scanner

**3.3**
**font**
complete set of characters of a given size, weight and style of type, including capitals, small capitals and lower-case characters, together with figures, punctuation marks, ligatures, etc.

NOTE See ISO/IEC 2382-23.

**3.4**
**magnetic ink character recognition**
**MICR**
machine recognition of digits printed with magnetizable ink

NOTE See ISO 2033.

**3.5**
**character pitch**
number of characters per unit length of a line of print

**3.6**
**recognition zone**
area around a recognition data field that is free of other data

**3.7**
**optical mark recognition**
**OMR**
machine recognition of a mark, such as a tick, cross or spot based on minimum area rather than shape of the mark

## 4 Layout and design

### 4.1 General



The design of a form that is easiest for a person to complete can be in conflict with the most machine-readable form. For example, in a user-friendly layout, the following items, all interspersed with printed instructions next to specific areas, might be desirable:



— large print,

— colour-coded areas, and

— areas to be completed with both alphabetic and numeric information.

However, in a form designed for EIM, strict segregation of spaces for numeric and alphabetic information and instructional text within dropout colour areas can all be essential features. Optimum design can require a compromise between ideal user and scanner requirements. All logically connected information should be placed on the same page.

### 4.2 Data storage requirement

The designer should be aware of the impact on data storage requirements of line borders, screened tints and logos or other design elements with large areas of reversed print. Reversed print will make heavy demands on data storage. For all designs, particularly those having large areas of reversed print, the amount of data storage required should be determined and compared to the amount of data storage available within the system.

### 4.3 Page format

On each page of a form, the margin all around should not be less than 10 mm. If the document is bound, padded or has punched holes or die cuts, the margin at that edge should be not less than 25 mm, and holes and die cuts should be restricted to that margin.

Text and entry fields should not be within 6,5 mm of any crease or perforation.

## 4.4 Type

### 4.4.1 Typeface

There are two styles of typeface commonly used on forms, serif and sanserif, as illustrated in Figure 1.


**Figure 1 — Comparison of serif and sanserif typefaces**

Serif style is designed for ease of legibility, has variable line width within a character and a cross-line finishing a stroke of a letter. Sanserif has uniform line width within a character and no cross-lines. It is the style used in International Standards.

Serif type will inherently take more data storage capacity in a compressed image than sanserif type because more information has to be recorded for each character. With the most commonly used compression techniques, approximately 10 % more storage is required for a page printed in serif, as compared with one printed in sanserif type.

Sanserif generally requires less horizontal line space and more vertical height than the same point size serif type. Because of its uniform line width, it is preferred for photocopying, microfilming and scanning.



For forms which might be used in optical character recognition (OCR) applications, sanserif typefaces should be used.



For information on a form that is not required to be captured by scanning, the style of typeface used is not important.

### 4.4.2 Symbols

An OCR program can use a particular symbol to prompt an action. The forms designer should be aware of any such symbols and avoid the use of them other than as a prompt.

### 4.4.3 Spacing

In typesetting, character spacing can be either fixed or proportional. In fixed typesetting, each character takes up equal horizontal space. Proportional typesetting allows for characters of different width, such as the width of "*i*" compared with that of "*w*", and automatically adjusts space between the individual characters to give a more natural appearance.

There should be a clear gap between characters. The recommended minimum gap is not less than the width of the vertical stroke of characters of the font.

The designer should also be concerned with vertical spacing requirements of an OCR system. Although 4,2 mm vertical spacing is usually sufficient for typewritten entries, at least twice that amount of space is necessary for hand-printed entries and for separating entries for OCR.

### 4.4.4 Character pitch

Form design should allow no more than 0,4 characters per millimetre for character pitch.

#### 4.4.5 Character size

In the printing industry, type size is usually specified in millimetres. In a computer or typewriter, type size is usually indicated in points. Fortunately, the printing industry is familiar with both systems and can easily translate requirements. The point (0,35 mm) is a unit derived from the height of metal slugs, once commonly, but now rarely, used to set type. The size of character is not directly related to point size. For a given point size, the actual heights of the same upper-case character can be different for various typefaces. There is also variation in the ratio of heights of lower-case “e” to upper-case characters. This means that for a given point size, even if upper-case characters of two different typefaces have the same height, there is a possibility that this is not so for lower-case characters. Because it is the size of the lower-case characters that will limit scannability, minimum acceptable point size should be determined by the height of the lower-case “e”. The recommended minimum height of the lower-case “e” is 1,4 mm.

If the EIM system is used as a transfer medium as part of overall processing of the information extracted from a form, the minimum type size used shall allow for any degradation of image quality resulting from subsequent parts of the process.

#### 4.4.6 Weight

The weight of a type font is its relative line thickness, ranging from light to extra bold. Font weight directly affects the number of dots or pixels used to display a character of an electronic image. Different weights can also be used to emphasize or reduce significance of text blocks or captions for the user.

#### 4.4.7 Type family



Design variations on a basic typeface can include italic, condensed, expanded and others. Form designers should try to keep the number of type families used within a form to a minimum to project an uncomplicated appearance that is pleasing to the eye. EIM systems, particularly OCR software, can also benefit from limited use of type families.



### 4.5 Machine printed stylized information

#### 4.5.1 General

Information may be presented on a form as a bar code, or in OCR or MICR characters. These bar codes and stylized character sets are especially designed for automated processing and are machine readable with high accuracy.

Machine reading is not always wholly accurate. The degree of accuracy achievable can be improved if forms have error-checking features built into their design. Whenever possible, forms should be designed to use a second source of information for cross checking. When calculation is involved, both subtotal and entry figures should appear on forms in order that the processing system can recalculate the subtotal and compare it with the amount read. Other examples of information for cross checking are account number/customer name and version number/issue date.

#### 4.5.2 OCR fonts

OCR fonts, such as Farrington 7B, OCR-A and OCR-B, are available with numeric only and alphanumeric character sets for automated recognition. Data encoded using OCR fonts shall be printed in accordance with the relevant International Standards (see Table 1).

**Table 1 — OCR font standards**

| Code types supported | International Standard |
|---|---|
| OCR-A Numeric | ISO 1073-1 |
| OCR-A Alphanumeric | ISO 1073-1 |
| OCR-B Numeric | ISO 1073-2 |
| OCR-B Alphanumeric | ISO 1073-2 |

OCR characters should be printed by a laser printer at the highest resolution whenever possible. The use of a dot matrix printer will generally give poorer print quality and reduce the accuracy of the OCR reading. Black characters printed on a white or light-coloured background are preferred.

OCR characters may be placed anywhere on the form, however, they should preferably be in a clearly defined recognition zone and printed parallel to the other text. Figure 2 provides an example of an OCR font.

**Figure 2 — Sample OCR code**



### 4.5.3 MICR fonts

MICR fonts (see example in Figure 3) are limited, highly stylized character sets that are printed using magnetic ink. Among several fonts available, E-13B and CMC-7 are the fonts frequently used on financial transaction documents, such as cheques. E-13B is highly machine readable because of its solid clear character format. CMC-7, on the other hand, is more difficult for optical recognition because the characters are made of discrete, thin vertical lines.



**Figure 3 — Sample MICR code**

### 4.5.4 Bar codes

#### 4.5.4.1 General

Bar codes (see example in Figure 4) may be placed anywhere on a form, provided appropriate steps are taken to protect their integrity.