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order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn.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.


Serif Sanserif

**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

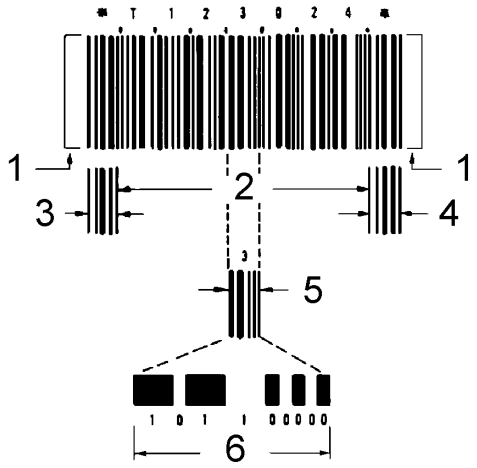
MICR fonts (see example in Figure 3) are limited, highly stylized character sets that are printed using magnetic ink. Among several fonts available, E-13B and CMC-7 are the fonts frequently used on financial transaction documents, such as cheques. E-13B is highly machine readable because of its solid 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.




**Key**

1 quiet zone (10X)
2 data
3 start
4 stop
5 typical data character
6 9 bits

**Figure 4 — Sample bar code**

#### 4.5.4.2 Using bar codes on forms

Bar codes are commonly used as a faster, cheaper and more accurate alternative for the capture of textual information about an individual, a place or a thing. There are many different types of bar code symbology available, with the choice of symbology being applications driven.

Bar codes can be printed using most of the printing processes. They can also be applied as labels at different stages of forms processing. When printing bar codes, the specifications governing the applicable symbology should be strictly followed to ensure that the codes will be successfully read.

Bar codes may be incorporated to automate forms-handling by representing the form number, certain fields of the user-entered information, and/or a control or tracking number to facilitate inventory requirements.

#### 4.5.4.3 When to add bar code

If the bar code represents information that is contained on the blank form, e.g. a form number, the bar code should be placed on the original prior to making distribution copies from it. A quiet zone, containing no dark marks, lines, print, etc., should surround the bar code area.

If the bar code represents information that is filled in by the user, e.g. an invoice number, the bar code should be created by the user, and the only form-design requirements would be to allow a sufficiently large quiet zone and instructions for the placement of a bar code label.

#### 4.5.4.4 Location of bar code

Ideally, the form should be designed so that a bar code representing specific information will always appear in the same area on each form. Designing the form in this manner facilitates automatic and single sheet scanning.

An eye-readable version of what the bar code represents should be printed along with the bar code, so that users of the form know what the symbol represents. This should be considered and space allowed when designing the form.

### 4.6 Hand-printed information

Machine recognition of hand-printed characters is used in subsystems that recognize either constrained or unconstrained hand-printed alphanumeric characters. This capability should be taken into account when a user is required to fill in blocks with letters or numbers from a list.

Hand-printed numerals can be interpreted with a high degree of accuracy because each has a distinct appearance. However, several of the twenty-six alphabetic characters can look very similar to each other, leading to difficulty in differentiation. It is not uncommon for users to connect letters even when they are printing them, and this decreases accuracy of hand-print recognition. It has been observed that the more hand-printing required to complete a form, the more the quality of printing deteriorates.

### 4.7 Location or registration marks

When designing a form intended for automated information capture, marks or codes should be used to ensure correct orientation and positioning of the form during scanning. There are several different marks suitable for location or registration purposes. These include Kermit code, cross hairs, rectangular blocks and triangular blocks. Using pre-printed text or graphic symbols is also possible depending on the particular software selected.

Samples of Kermit code are shown in Figure 5.

Location or registration marks should be placed on the document, spread as far apart as possible, avoiding edges of the page and corners where staples or clips might be placed.

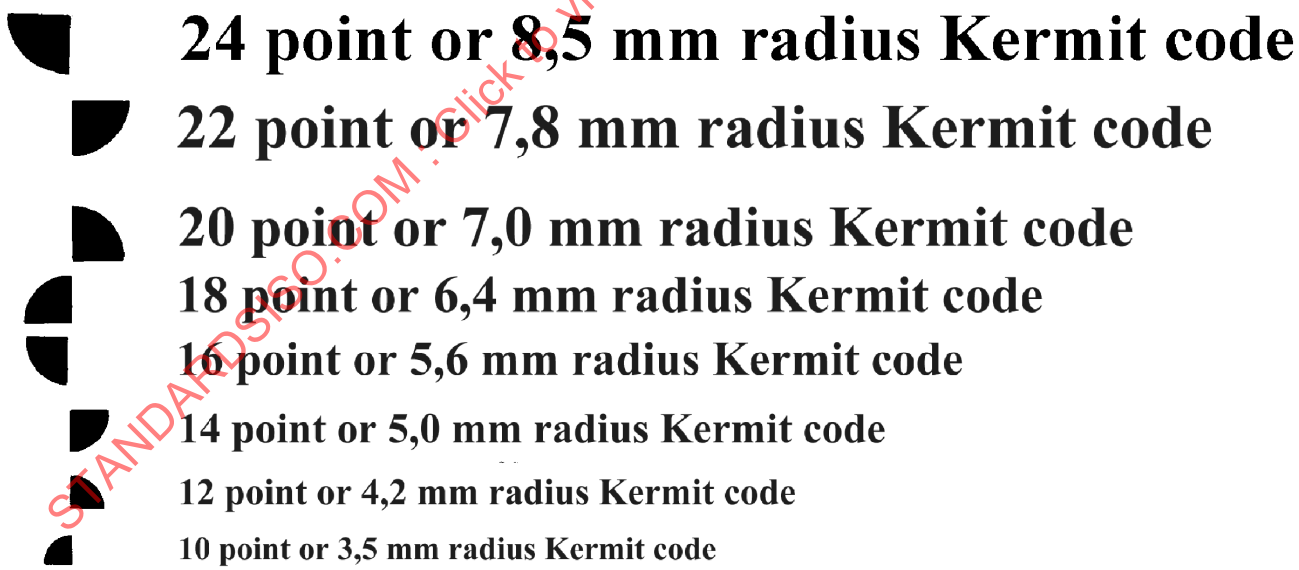



**Figure 5 — Samples of Kermit code**

These marks should be offset to give the software the best opportunity to de-skew the document before the scanning process.

There should be at least four, preferably six, location or registration marks, in case some are defaced.

### 4.8 Highlighting

Methods for highlighting an area of the form, using screened solid or borders of recordable colour, can seriously increase the data storage requirement. When dotted or lined screened areas are used, the data storage requirement increases with the width of the dot or line. A solid of high density could reduce contrast with black print to an undesirably low level, making capture difficult. A line border is more economical of data storage than a solid of the same area. Using a dropout ink that will consistently be read as white by the scanner is extremely efficient for image storage.

### 4.9 Printed rules

Line thickness should not be less than 0,7 mm. All lines should be parallel to an edge of the form.

## 5 User guidance

### 5.1 General

Instructions and form attributes guide a user to enter appropriate information in each of the entry fields. Instructions are either printed text or design features of a form that lead the user to complete the form in the way intended. These are not considered information to be captured and should be printed in a dropout colour.

### 5.2 Hand-written entries

Instructions for hand-printing alphabetic or numeric characters on forms may include statements such as the following:

— “PLEASE PRINT USING CAPITAL LETTERS ONLY”;

— “PLEASE PRINT LIKE THIS”;

— “PRINT YOUR CHARACTERS LIKE THIS; ONE CHARACTER PER BOX”;

— “PRINT CHARACTERS AS SHOWN”;

— “PRINT ALL NUMBERS AS SHOWN”;

— “PRINT ALL LETTERS AS SHOWN”;

— “PRINT CAREFULLY ALMOST FILLING BOXES AS BELOW”;

— “DO NOT USE FELT TIP PENS”.

An instruction of the kind given above, in conjunction with a sample of hand-printing style such as the example in Figure 6, provides useful aids to form completion. The user is instructed as to what to do, and given an example of how to do it using hand-printing of the required style.

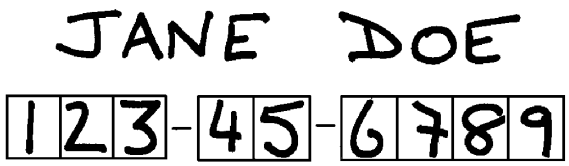



**Figure 6 — Illustration of desired hand-printing**

## 5.3 Printed instructions

Printed instructions should be clear, concise and adjacent to the corresponding entry field. Extensive, detailed instructions should be avoided, as should those that refer to a separate space on a form, the reverse side of a form, a completely separate sheet or a completely separate set of instructions.

## 5.4 Form features

### 5.4.1 Form features for layout extraction

In order to achieve best results, both in the human interface phase (in which the forms are being filled in) and in the recognition phase, it is recommended to use predefined printing techniques.

The design of the forms should be planned in such a way that the lines which create the layout of the constraint boxes will be later removed (dropout), leaving only the filled-in information to be recognized.

There are different methods to ensure the dropout of the constraint boxes. One of the most common methods is the use of dropout ink colours, which are not captured or which are minimally captured by the scanner during the scanning phase. All non-colour scanners have either a red, green or blue dropout colour, and most scanners can provide the ability to configure which dropout colour is to be used through a control panel, switch or “bulb” or “light” change. While most scanners have provided the ability to change the “bulb” or “light” to change the dropout colour being used, most scanners on today's market utilize firmware/software configuration capabilities to achieve this change, and not hardware changes, as in the past.

Another common method is to use only 20 % of the regular ink through a net. This technique produces a light dotted line. The result is the same.

The ways to implement the above-mentioned methods vary from one project to the other, e.g.

— the areas *around* the constraint boxes could be coloured in light blue dropout ink, leaving a white empty area inside the constraint boxes;

— the areas *inside* the constraint boxes could be coloured in light blue dropout ink, while the rest of the page is white;

— the light dotted line technique could be used to delimit the written areas in constraint boxes.

### 5.4.2 Form feature for document preparation

It is recommended that the printing house cut off one of the corners of each paper form. This will help the preparation phase personnel, as well as the scanner operators, to verify easily that all the forms are in the same position before entering the batch into the scanner. This will also save valuable time on scanning the forms in the wrong orientation.

### 5.4.3 Registration marks

Taking into account that most scanners result in some skewing and minor distortion when the document is being fed through the transport mechanism, each form should include “registration marks”. “Registration marks” are small “dots” in the corners of the document which allow for the scanner to “align” the page, ensuring accurate data capture and recognition. “Registration marks” should be a minimum of 10 mm in diameter and of a solid dark colour, or black. There should be a minimum of two registration marks on each form, and preferably four. If only two “registration marks” are to be used, they should be placed on opposite corners of the document, outside all information to be captured and processed. If four marks are to be used, each mark should be placed in a corner of the document, ensuring proper image alignment prior to processing.

### 5.4.4 Form identification

One of the most important issues during the data capture phase is to recognize and identify correctly the identification field(s) which uniquely identify the specific forms. There are several ways to identify a specific page or image.

The following methods have been used in different projects around the world:

— printing bar codes on each side of a form;

— printing bar code stickers on each side of the form;

— printing a unique serial number on each side of the form;

— including a hand-written, unique serial number on each side of the form.

It is recommended to include a check digit in the serial number to allow logical check of the recognized information.

All forms within a flow should be compared, in order to verify that no two forms are too similar, thus making form identification more difficult. In cases in which two forms are highly similar, it is recommended that both an identification mark and a unique serial number be printed on each form.

### 5.4.5 Check boxes

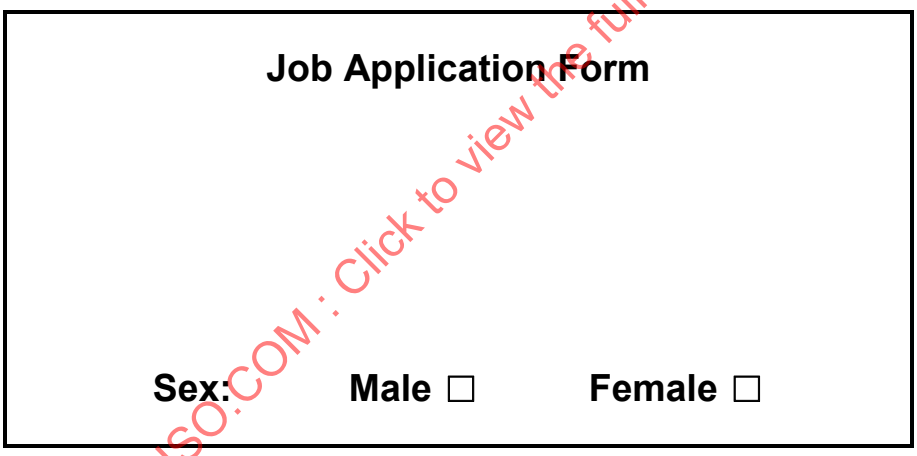



**Figure 7 — Optical mark recognition zones**

In general, in order to improve recognition results, most of the questions on the form should be in “check box” format (see Figure 7), rather than in alpha or numeric format. The check boxes should be one beside the other, rather than one above the other, e.g. the design of the form should follow the example in Figure 8 b) rather than that in Figure 8 a) below.

What age group do you belong to:

0-15

16-28

29-49

50 and above

**a)**

What age group do you belong to:

☐ 0-15 ☐ 16-28 ☐ 29-49 ☐ 50 and above

**b)**

**Figure 8 — Use of check boxes on forms**

### 5.4.6 Importance of spacing

One of the most important goals of the form is to ensure the easy interpretation of the filled-in information during the processing phase. To achieve this, it is recommended to implement a method that will segment each mark and/or character from the other. One of the most efficient ways for segmentation is by using a separated constraint box for each filled-in symbol.

The common standard for optimal size of the constraint boxes should be as follows:

for OMR (tick box): 3 mm × 3 mm;

for written numerals and alpha: 6 mm height × 5 mm width.

There should be a space of 1 mm between neighbouring constraint boxes.

It is recommended to have a minimum margin of not less than 1,5 cm between the information written or printed on the form and the edge of the page.

Figures 9 and 10 provide illustrations of constraint boxes.

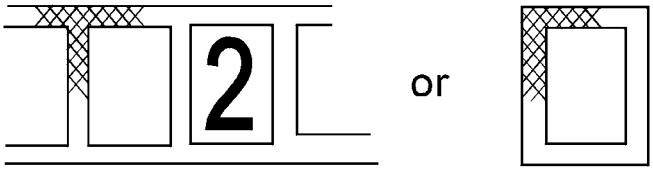



**Figure 9 — Sample constraint boxes showing minimum margin**

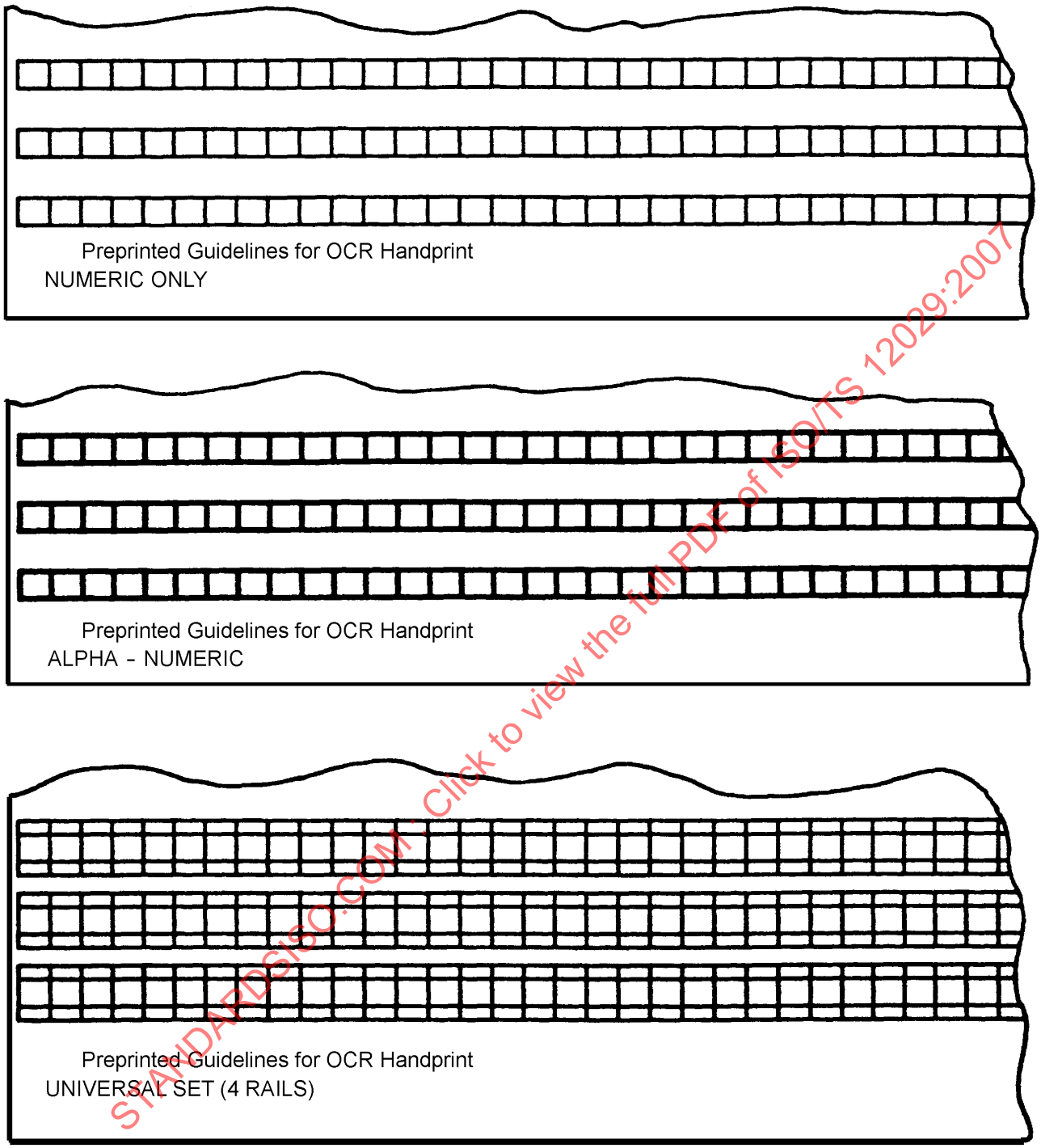





**Figure 10 — Sample constraint boxes**

## 5.5 Optical mark recognition (OMR)

Check boxes and circles shall be large enough to show a clear and distinct mark. They should be free from background shading and edge shadow. For accurate information capture, square boxes are preferred.

The internal length of box sides and the internal diameter of circles should be 7,0 mm ± 2 mm and the minimum line thickness should be 0,7 mm. The minimum distance between adjacent zones should be 7,0 mm.

# 6 Post-printing features

## 6.1 Perforations

Perforations facilitate detachment of part of a form to meet other requirements. Slit perforations are preferred to hole punched perforations because slits are less likely to interfere with the document feeding than holes, which can cause jamming by snagging sheet corners. Perforations should be avoided in high-volume processing applications.

## 6.2 Holes

Holes may be punched to allow forms to be placed in ring-binders. Holes should be placed with centres not less than 12,5 mm from any area to be scanned, in order to avoid eliminating information requiring capture. Drilling or punching should be made cleanly, leaving no loosely attached debris. Loose paper confetti discs can interfere severely with the operation of scanners. Avoid placing holes where they can cause jamming, or inadvertently signal the feeding mechanism to stop by allowing a light beam to pass through the hole.

## 6.3 Die-cuts

Die-cuts, used for filing systems and windows, can cause jamming in the document feeder of a scanner. They should be avoided.

## 6.4 Folds

Not more than two folds should be allowed in a form. Any fold should preferably be made parallel to the machine (grain) direction of the paper, and positioned not less than 6,4 mm from information required to be captured.

## 6.5 Multi-part forms

Multi-part forms that are to be scanned should be collated in logical order for the system. The form design should provide identification of form and part number on each sheet.

## 6.6 Padding

A low-tack, flexible, high solid content adhesive should be used to make form pads. When detaching a form from a pad, some of the adhesive often remains attached to the form. If it is transferred to the feed mechanism of a scanner, the adhesive can cause jamming. Low solids content water-based adhesives can cause waviness of the edges of the paper to which it is applied. After detachment, the wavy edge can cause jamming in the feeder.

## 6.7 Carbonizing

Forms having patches of a wax-based carbon compound, e.g. airline tickets, are sometimes processed through automatic feed equipment. The carbon compound accumulates on the rollers and belts with which it comes into contact, necessitating frequent shutdowns for cleaning. Whenever possible, black chemically or physically mated carbonless copy should be used for this purpose.