
**Electronic imaging — Forms design
optimization for electronic image
management**

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

STANDARDSISO.COM : Click to view the full PDF of ISO/TS 12029:2002



PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

STANDARDSISO.COM : Click to view the full PDF of ISO/TS 12029:2002

© ISO 2002

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

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
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Layout and design	2
5 User guidance	7
6 Post-printing features	9
7 Materials	11
8 Form testing	12
9 New technology	12

STANDARDSISO.COM : Click to view the full PDF of ISOTS 12029:2002

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 3.

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 with a view to deciding whether it should be confirmed for a further three years, revised to become an International Standard, or withdrawn. In the case of a confirmed ISO/PAS or ISO/TS, it is reviewed again after six years at which time it has to be either transposed into an International Standard or withdrawn.

Attention is drawn to the possibility that some of the elements of this Technical Specification 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 imaging applications*, Subcommittee SC 2, *Application issues*.

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 the design of a form.

STANDARDSISO.COM : Click to view the full PDF of ISO/TS 12029:2002

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 normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Specification. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this Technical Specification 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 1073-1:1976, *Alphanumeric character sets for optical recognition — Part 1: Character set OCR-A — Shapes and dimensions of the printed image*

ISO 1073-2:1976, *Alphanumeric character sets for optical recognition — Part 2: Character set OCR-B — Shapes and dimensions of the printed image*

ISO 2471:1998, *Paper and board — Determination of opacity (paper backing) — Diffuse reflectance method*

ISO 11108, *Information and documentation — Archival paper — Requirements for permanence and durability*

ISO 12651:1999, *Electronic imaging — Vocabulary*

ISO 12653-1, *Electronic imaging — Test target for the black-and-white scanning of office documents — Part 1: Characteristics*

ISO 12653-2:2000, *Electronic imaging — Test target for the black-and-white scanning of office documents — Part 2: Method of use*

3 Terms and definitions

For the purposes of this Technical Specification, 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

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.

3.4

magnetic ink character recognition

MICR

machine recognition of digits printed with magnetizable ink

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 which is the 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. Large reverses 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 their use 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 that of “i” compared with “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 the 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 it 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, and 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 so that the processing system can recalculate the subtotal to 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 according to the relevant International Standards, as given in 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. For an example of an OCR font, see Figure 2.

A B C D E F G H

Figure 2 — Sample OCR code

4.5.3 MICR fonts

MICR fonts, shown 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.

255074988 1598300 096

Figure 3 — Sample MICR code

4.5.4 Bar codes

4.5.4.1 General

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

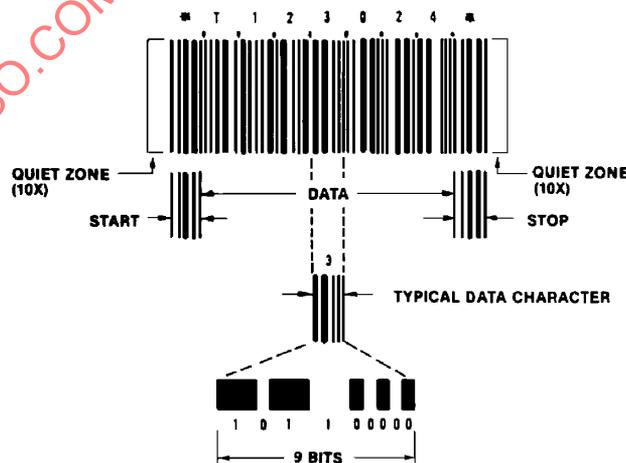


Figure 4 — Sample bar code

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 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, such as 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 always appears 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 numeral has a distinct appearance. However, several of the 26 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 spread out on the document as far apart as possible, avoiding edges of the page and corners where staples or clips might be placed.

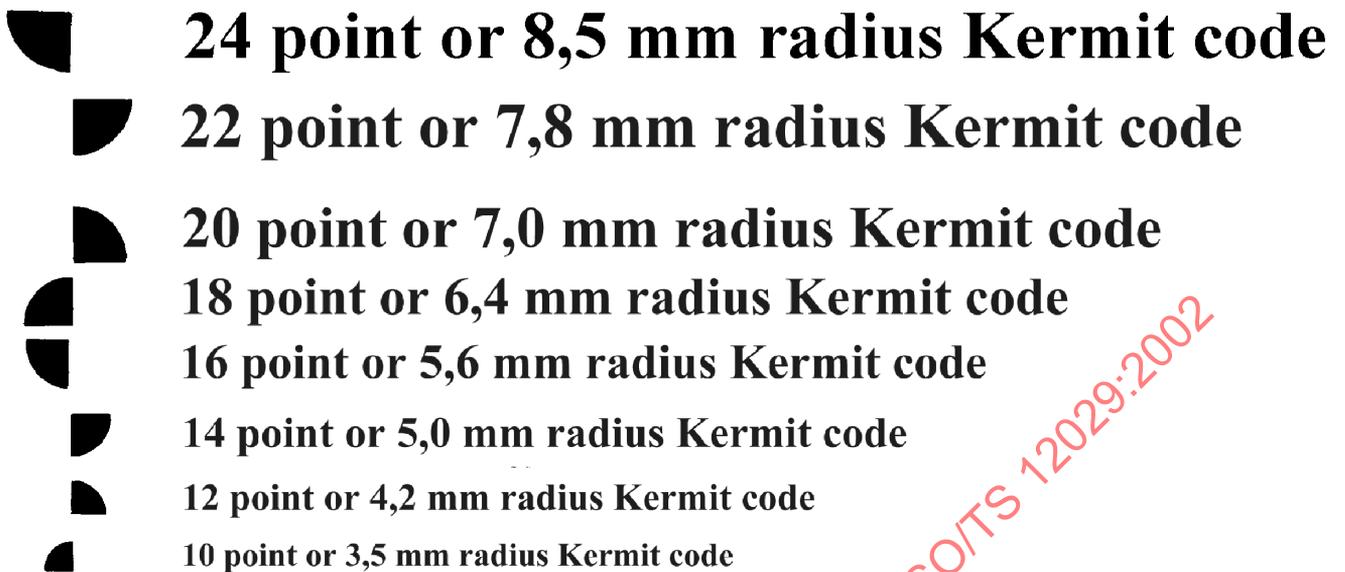


Figure 5 — Samples of Kermit code

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 Handwritten entries

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.

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

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

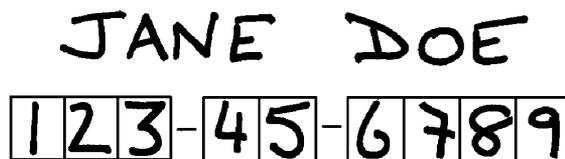


Figure 6 — Illustration of desired hand-printing

5.2 Printed instructions

Printed instructions should be clear, concise and adjacent to the corresponding entry field. Extensive, detailed instructions and 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 should be avoided.

5.3 Form features

Examples of form features that assist completion include the following:

- design of the entry fields or boxes in the target field itself;
- logical sequence of requested information;
- highlighting with colour (often a dropout ink);
- ruled borders;
- textual help, e.g. periods, commas and currency sign for amount fields;
- check boxes or circles used for multiple choices;
- optical mark recognition (OMR) zones.

Data-entry spaces are not only a guide to the user, they are critical to using hand-printed characters as recognition objects.

Constraint boxes are used in an attempt to avoid the following types of hand-printing problems that can affect recognition:

- touching characters;
- overlapped characters;
- nested characters;
- joined characters;
- broken characters.

To facilitate capture from hand-printing, field boxes have been devised to constrain and clearly separate characters. For examples of constraint boxes see Figures 7 and 8. These may be printed using a dropout ink.



Figure 7 — Sample constraint boxes

Constraint boxes should be separated by small spaces. This is because boxes without spaces between them can still result in the problems described previously. Boxes intended for hand-printed entry should not be too small. Some people find it difficult to print clearly, and this problem becomes worse when printing is confined to small entry spaces. It is recommended that constraint boxes be not less than 6,4 mm high and 5,1 mm wide.

Highlighting parts of a document with the use of colour and ruled border lines are examples of form features guiding a user's eyes to another part of the document.

Some guidelines are contained in textual instructions. These consist of printed symbols that further indicate what is needed. Examples are:

- currency sign, commas and decimal marker in an amount field;
- colon in time field;
- oblique in date field.

5.4 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.

Internal length of box sides and internal diameter of circles should be $7,0 \text{ mm} \pm 2 \text{ mm}$ and minimum line thickness 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 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 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.