FORMS 5

Effective Form Design

June 2011 – ReadSoft AB

Questions or comments about this document may be emailed to **documentation@readsoft.com**.

30062011.0510

ReadSoft AB (Head office) | Södra Kyrkogatan 4 | SE-252 23 Helsingborg | Sweden | Phone: +46 42 490 21 00 | Fax: +46 42 490 21 20
info@readsoft.com | www.readsoft.com

Contents

Introduction

FORMS is the most advanced forms processing and automatic data capture software available today. It is flexible enough to provide good interpretation results from forms that are not specifically designed for scanning or forms that could not be satisfactorily scanned with our competitors' software.

Many customers wanting to design completely new forms have asked us how to optimize their design for FORMS, while allowing easy set-up and scanning and yielding excellent interpretation results.

In response to this, we have created this guide, combining the knowledge of many customers with years of FORMS experience and ReadSoft's own expertise. This guide contains some suggestions and ideas on how to:

- Design a form in a way best suited to the paper handling and image properties of a standard entry-level scanner.
- Design a form so that FORMS can easily identify it and distinguish it from other forms.
- Design individual fields that respondents are likely to fill in clearly and correctly, to achieve excellent interpretation and few validation errors.
- Make decisions about color and printing that affect data capture favorably.

This guide is not definitive and should not be seen as dictating "the only way" to design a form. Many of our customers are happily scanning forms that were designed by an in-house team who had no knowledge that the forms would be scanned.

For information about topics not covered here, please refer to *FORMS Help*.

Scanners

The most important piece of hardware in any FORMS system is the scanner. There are many different manufacturers and models of scanners compatible with FORMS, ranging from small, inexpensive desktop scanners to very large high-volume scanners. All supply FORMS with an image it can interpret. However, needless to say, a cheaper scanner is unlikely to produce the same quality of image or possess the features of a higher-end model.

Image quality and enhancement features

Many of the middle- to high-end scanners provide special features to improve the quality of a scanned image. Some of the interface cards used to connect the scanner to your computer also provide such features. You can access most of these features from FORMS, and they aid interpretation. Despeckle, deskew, destreak, border removal, and many other special image enhancements can also help with poor quality forms, allowing you to get acceptable interpretation results where cheaper scanners may substantially reduce performance. For more information, look up “Scanner settings” in the *FORMS Help* index.

Paper feeding

The feed quality of a scanner is another very important feature when you consider the design of a “scanner-friendly” form. Cheaper scanners are suitable for standard paper weights and sizes and good quality paper. More expensive scanners are generally more forgiving and are able to handle a much wider range of paper types, including lower quality paper.

Dropout colors

One specific feature of a scanner is the ability to treat some colors as white. Such colors are known as *dropout colors*. Dropout colors are usually solid pastel colors like pale greens, reds and yellows that are invisible to the scanner.

This feature goes back to the early days of Optical Mark Recognition (OMR). As an example, many gaming coupons (such as lottery coupons) use dropout colors on the form background. These were ignored by the scanner, leaving only the mark to be interpreted.

Remember: FORMS, unlike other software products, does *not* need to use this feature. In fact, for mark fields, FORMS prefers boxes with solid borders. On the other hand, the interpretation rates of ICR character fields can be significantly improved by using these colors as field separators.

Each scanner can “drop out” a different range of colors—check with your manufacturer for complete lists of tested Pantone™® colors for reliable dropout.

As a general rule, cheaper scanners handle black-on-white or 100% dropout colors. High-end scanners are less critical when it comes to these factors and can selectively ignore lighter colors that are not specifically designated by the scanner manufacturer as dropout colors.

The form, the respondent, and you

With any new form, the emphasis should be placed on design and testing. The design of a form greatly influences:

- The way respondents write (for example their handwriting). When respondents write clear and well-separated characters, interpretation improves.
- The correctness of what they write (for example whether the information has the format you expect). When the information they provide is correct and formatted the way you want it, there are few validation errors and therefore very little processing in Verify.

Therefore, form design can have a profound impact on the efficiency of the data capture process.

Below are some basic principles that describe how to *guide* and *constrain* respondents so that their answers result in fast and accurate processing.

Provide enough space

Think of how many forms you are asked to fill in where space is cramped. The effect is that the quality of your writing decreases. A good rule of thumb is that if you cannot read the response, FORMS will not be able to either. Leaving a 5 mm by 7 mm box per character may seem like a luxury and take up a lot of space, but the improvements in interpretation rates are substantial.

Read more about the optimal size of character fields in “Choose a suitable field size”, which begins on page 14.

Influence how respondents write

All ICR interpretation engines work best on clear, uppercase, well-separated characters. Bearing this in mind, try to design your form so that most of your respondents write in this way. You can do this by:

- Designing character fields that force respondents to complete the form in a certain way, as described in "Constrain the respondent", which begins on page 14.
- Providing instructions, as described on page 8.

Control the format of data

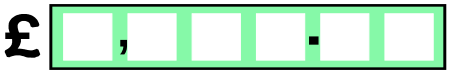
Many of the problems with capturing data automatically arise when the person filling in the form responds in an unpredictable way or tries to be "helpful" by providing more information than you need.

In other cases, a field on a form might request information but not specify its format. For example, if the question is your salary, but the field is a simple empty box, think of all the different responses someone earning £12,945 a year might provide:

£12,945.67p £13— £13K £12945—

…and so on! The respondent could tell you how much they earn per week or per month, whether they receive commission or bonuses, etc.—all in a field where you expected a single type of response.

There are many ways to get around this problem. One method is by making the question much clearer. However, the best way is to limit the possible responses using clever field design. Consider the following field design:



This makes it much less likely that any of the unexpected ways of writing a salary figure will be used. The respondent is constrained to supplying only the necessary information (numeric only, in this case) for good ICR data capture.

For a question like this, you might consider using a mark (tick box) field instead. If your data analysis is based around ranges of salaries, for example, a series of mark boxes with those valid ranges quoted would always give better results.

The basic rule is: Always structure the form in such a way that it becomes difficult for your respondent to use it in any way other than the way *you* want them to.

Another tip is to provide space for customers to write additional information. You may wish to ignore this extra information, but respondents will be less likely to write all over your data-capture field areas.

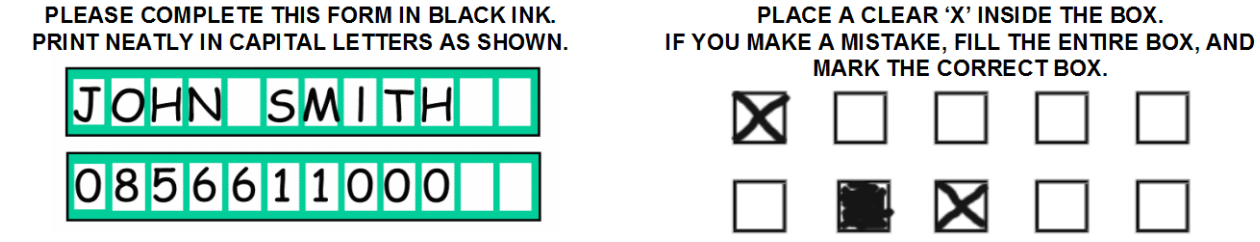
An example of this could be a “Work Phone Number” field. Often people try and squeeze in their extension number into the field, preceded by the word “Ext” or “Extension”. Since FORMS is probably expecting a purely numeric telephone number, this type of response results in significantly more verification due to the lower interpretation rate, which in turn is caused by unexpected alphabetic characters and/or more digits than expected.

It would be much better to provide a box for four or five digits, labeled with the word “Ext.”, so that the respondent can provide that information. Whether you capture this data or not is up to you.

Provide instructions

All forms should contain clear and concise instructions on how the form is to be filled in. Above handwritten fields, stress the importance of writing all information within the boxes.

Graphical instructions also help. If respondents see how to do it, they are more likely to follow the example. Notice that these instructions show the respondent how to cross out an accidentally ticked mark field:



You can also use dropout color to provide instructions. In the field below, DDMMYY is printed in a dropout color (described on page 5) that is not detected by the scanner.

Design for identification

Before any data on a form can be captured, FORMS must identify the form. Unique adjustment fields and recognition fields on forms maximize their potential for identification.

Optimal design of adjustment fields

In most cases, the layout of different forms is unique. FORMS uses *adjustment fields* to identify and deskew forms and match them to the form definition. All the other fields on a form have a fixed position on the form relative to the adjustment fields.

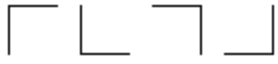
Traditional adjustment fields contain characters, symbols, or intersecting lines that are unique to the specific form (compared to other forms to be processed at the same time). Duplex (two-sided) forms require five adjustment fields on each side.

Flexible line maps are another kind of adjustment field, but these are not covered here. You can find more information about flexible line maps in *FORMS Help*.

When forms have no natural ideal adjustment fields (for example when a there is very little text outside of the field areas, or when strong background colors interfere with potential adjustment fields), it can be advantageous to have special ones printed on the forms.

Ideally, traditional adjustment objects are angular characters. We recommend using 90-degree right angles of about 5mm x 5mm to 7mm x 7mm.

Example (not to scale):



The objects should be black on a white background, surrounded by white space (no other characters, lines or dots). One adjustment object is placed in each corner of the form. Place adjustment objects at least 10 mm (3/8”) from the edges of the form. Ensure that your adjustment fields are placed differently (asymmetrically), or place different characters in each corner of the form, otherwise a form that is placed in the scanner upside down may be identified incorrectly.

Optimal design of recognition fields

Why use recognition fields?

In cases where one form is so similar to another that *unique* adjustment fields cannot be found, FORMS uses *recognition fields* instead to identify the form. A recognition field contains a unique identification number, a form number. (Note that adjustment fields are still required for deskewing.)

FORMS uses recognition fields to identify forms when the forms are too similar to be identified by their adjustment fields alone, or when a standard set of non-unique adjustment fields has been preprinted on every form.

However, recognition fields are good to have on forms even when adjustment fields are optimal. Here's why:

If you add a recognition field to every form, and change the number for every minor revision or reprint of the form, you will have insurance against a printing problem or publishing error that causes your new form to be incorrectly identified. In normal situations the recognition field may not be included in the form definition. But if such a problem occurs, you simply create a new form definition, adding the recognition field for the new form. Without the recognition field, the printing error may have meant that a whole print run was useless.

Therefore, add a recognition field to every new form you design.

Requirements for recognition fields

As recognition fields are particularly sensitive, they must:

- Contain characters larger than 2 mm and smaller than 10 mm (9-point to 42-point) in height, with good contrast with the background. (For best results, use a 10-point to 14-point typeface.)
- Not be written on when the form is filled in.
- Not be so near the edge of the form that FORMS might miss them.

The ideal recognition field

Ideally, the identification number has the following attributes:

- It consists of only a few digits, as digits are interpreted with more accuracy than other characters.
- It is about 3 mm in height (12-point), uses a plain typeface without serifs, and is not italicized or bolded.
- It is preprinted with high quality on the form in a white field with a black rectangular box around it.
- It is not placed too near the edge of the form or any other characters or lines.

Plan to use the same scanner settings on all forms

FORMS frequently handles jobs where several forms are placed together in *sets*. (A set is two or more form definitions that are scanned, interpreted, and verified together, such as market research questionnaires consisting of several pages, or forms with enclosures.)

Even when sets are not used, different form definitions can be included in the same job, making pre-sorting unnecessary.

In these cases, FORMS requires consistent settings for paper size, resolution (DPI), lightness, and contrast. You cannot mix these parameters in one job. Therefore, try to keep these factors the same on all forms you create.

Design for interpretation

After a form is identified, FORMS captures and validates data from individual fields. This section tells you the best way to design the most common field types interpreted by FORMS.

Optimal design of mark fields

Mark fields (or tick boxes) are usually boxes that a respondent fills in with a check mark or ”x”.

If you are designing a new form, use mark fields whenever possible instead of fields containing characters to be interpreted. For example, rather than having a respondent indicate his or her sex by writing an “M” or “F” in a character field, offer one mark field for male and one for female.

Mark fields are processed very reliably if have the following characteristics:

- A mark field should be a white square surrounded by a solid black outline (box). A 1-point line is recommended, not hairline. The white background should not have any shading or background pattern. Though FORMS can handle smaller fields, an ideal mark field is about 3 x 3 mm (1/8” square) in size.
- The spacing between fields is important. A vertical gap of at least ½ the height of the field and a horizontal gap of more than twice the width of the field are recommended.
- A cross or “x” is preferable to a tick or check mark for filling in mark fields, since the former tend to run across into other boxes or neighboring text fields, causing inaccurate results. An untidy respondent is more likely to be constrained by having to make a cross or an “x” in the appropriate box. We recommend that you include on your form an illustration of how to fill in the box. See the example in “Provide instructions.”
- Using the field definition options **Detect strikeouts** and **Detect noise**, you can direct FORMS to ignore fields that are either very heavily filled in or too lightly filled in. These can reduce the amount of verification you do on mark fields. To increase the usability of these options, print clear instructions above each field, outside the actual field area. Sometimes respondents mark the wrong box, so explain and show an example of how to shade out a box they have marked by accident. Again, see the example in “Provide instructions.”

Optimal design of character fields

Character fields contain handwritten or machine-printed characters to be interpreted.

Decide what is to be written in the field, and how

First, weigh the importance of each field. Do not ask for information that you do not need.

Consider whether the same information can be obtained using mark fields rather than characters to be interpreted. For example, if the question is about salary, and your data analysis is based around ranges of salaries, go with a series of mark fields showing those ranges. It is easier to interpret a mark in a box than a salary written by hand.

When you do need a character field, try to find the single most appropriate *format* for the information in the field. FORMS' validation routines are most efficient if the format of the information in the field is identical on every form. For example, if a number is expected, for maximum efficiency it should always have the same number of digits, not three digits in one case and five in another. Ways of controlling the format of the data are discussed in “Control the format of data”, which begins on page 7.

Machine-printed fields that are to be interpreted

In principle, FORMS can interpret all printed typefaces. The level of interpretation is affected by:

- Which typeface you choose. For best results, select a clear, sharp font. Typefaces without serifs are often more easily interpreted than fonts with large serifs. Avoid bold and italicized fonts.
- The font size. Select a height of 2.5-3 mm (10-12 point) but not larger than 12 point.
- Its accessibility. Place text at least 10 mm (3/8”) from the edge of the paper.
- Scanning resolution.

Keep in mind that this advice is only valid for fields from which you want FORMS to interpret data. The position, typeface, and size of text that will not be interpreted does not matter.

Consider the respondent

Consider the person who fills in the form. Forms that are easy to complete often result in good interpretation rates.

Print clear instructions above each field, outside the actual field area, as described in “Provide instructions”, which begins on page 8. Keep in mind that ICR interpretation engines work best on clear, uppercase, well-separated characters.

Choose a suitable field size

The size and format of a character field greatly influences how respondents fill it in and, therefore, the interpretation rate.

Experience shows that the best results are obtained when a character field is about 7 mm high, with a width of at least 5 mm for each character that is expected to be written in the field. For example, make a field that is to contain a 7-digit product number 7 mm high x 35 mm wide.

While the above measurements are ideal, small adjustments in these sizes can be made without interfering too much with interpretation.

Field delineation and background

Put a solid black box around character fields. A line thickness of 1 or 1.5-point works best.

Give the field a white or dropout background, free from shading or patterns. Avoid background text such as unit measurements inside the field, or place such text so that the field contents do not overwrite it.

Constrain the respondent

Ideally, all hand-written alphabetic and numeric fields should be *constrained* to discourage respondents from connecting characters together. That is, they should have one clear area per character to be filled in.

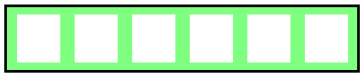
Three ways of constraining the field are shown below. These are the best configurations for handwritten character fields, in order of preference.

Note: Even higher interpretation rates can be achieved with semi-constrained and fully constrained and fields if you select the **Fixed character width** option in the field definitions for these fields.

Dropout color as constraint

Use a dropout color (described on page 5) to guide respondents' placement of handwritten characters. These fields should be 4 mm wide by 7 mm high (white space) per character position, with shaded spacing around the fields of 2 mm and gaps between the fields of 2 mm, all printed in a dropout color. Put a 1-point solid black line around the total field area.

Example:



The principle of a dropout color box is that while filling in the field, respondents are constrained to the white areas of the box, which keeps the characters separate and clear.

Characters that extend into the colored areas do not cause problems, because the scanner ignores the dropout color (that is, it sees the color as white), leaving FORMS with only the filled-in box.

What the filled-in field looks like:

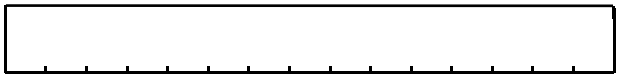
What FORMS sees:

For information about which dropout colors are applicable to your particular scanner, consult the scanner manufacturer. Testing of forms with dropout colors is recommended.

Comb marks (hash marks) as constraint

Lines can be placed on the bottom of the box or both the top and bottom. Although FORMS automatically excludes comb marks from interpretation, ideally the marks should be no more than 1.5 mm in height to ensure that they are never mistakenly interpreted as one ("1") or the letter "i".

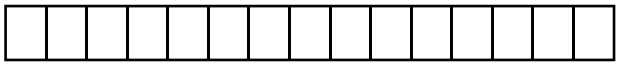
Example:



Vertical lines as constraint

It is possible to use black and white boxes.

Example:



In this example, you could double-click each small box in the field, but this would have two disadvantages:

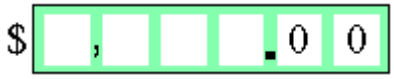
- With comma separation in the output, there would be commas between each character.
- Validation of the field as a whole would not be possible.

Special fields

Monetary amounts

Fields containing information such as monetary amounts should have preprinted currency symbols (£, $, etc.), thousand separators, and decimal points to discourage respondents from writing these symbols themselves. Decimal zeros should be preprinted if these do not need to be completed by the respondent.

Example:



Date fields

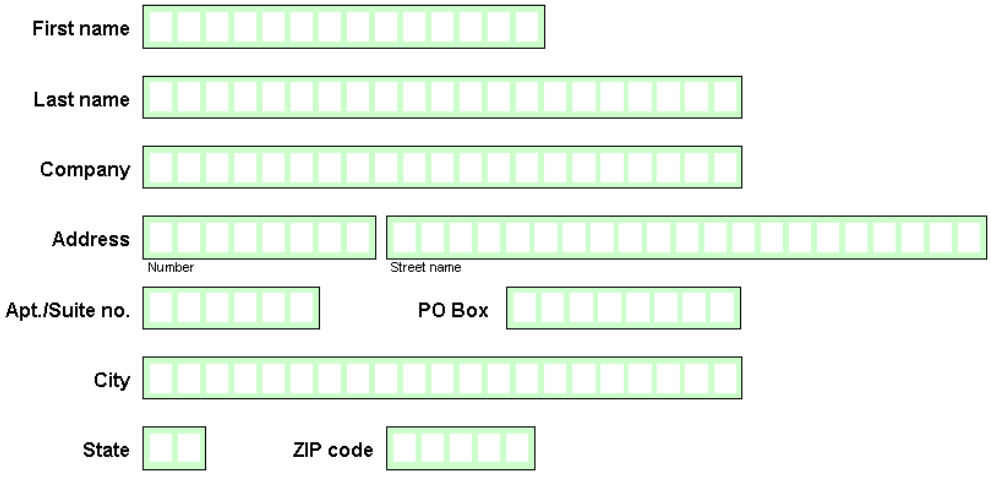
Date fields should have printed separators and a guide to show where to complete each element of the field.

Example:

Address fields

Address fields should have each component in a separate field. Important considerations are the house/building number and postal code. These fields can contain either comb marks or dropout colors.

Example:



Other field types

This document describes the optimal design of adjustment fields, recognition fields, mark fields, and character fields. You may want to capture other types of information. *FORMS Help* describes the optimal design of other field types such as barcode fields, image fields, object fields, and selection fields.

Design for validation

To achieve the most accurate interpretation of your forms, you must define each field as specifically and accurately as possible on the form definition. Then take advantage of all of FORMS' many options for field- and form-level validation.

You can make this easier by building opportunities for validation into your form design. Read about FORMS' built-in field and form validation functionality in *FORMS Help*.

Printing the forms

Paper features and quality

Document feeders can be sensitive, so it is important to choose paper of the correct quality according to the recommendations of the manufacturer or dealer of your scanner.

Generally, the following are to be avoided:

- Slippery or high-gloss paper surfaces
- Low quality, dusty, and NCR or carbon (duplicate) type paper
- Ripped, crumpled, and torn papers (especially along the leading edge)
- Very thick or thin paper (70-80 g/m2 is an ideal weight; other weights should be tested)
- Gummed or heavily perforated edges (especially along the leading edge)
- Forms that commonly have attachments stapled or paper-clipped to them

Paper color and background

The goal for FORMS is always the same: to receive or process an image in which the information that needs to be preserved is always dark, clear, and (at the lowest level – after FORMS or the scanner applies any image filters) seen as solid black. This includes:

- The respondent's text
- Boxes surrounding fields
- Any other foreground of the form, such as preprinted currency symbols, decimals, or decimal points

FORMS should be able to ignore (treat as white) information that is not needed, especially where it either 1) creates a larger image file than necessary, or 2) causes noise in a field area, because this affects interpretation negatively.

It is easiest to process a form with a white background. However, light pastel colored paper may be acceptable. Testing is recommended to ensure scanner compatibility before using colored paper in a production environment.

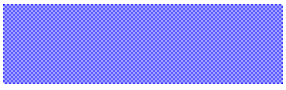
Dropout colors

If you print forms with dropout colors, choose your printing company carefully. Some companies do not understand all the ramifications of scanning such forms. Test all color dropout forms with full-color proofs before ordering large quantities of them.

Colored forms

Avoid *graphic washes*, where spaced dots of a darker color are used to create the impression of a lighter color, as shown below. This is never successful. Graphic washes cause lower interpretation if they cover or immediately surround fields where data is recorded. Also, if you save image files for any reason, these dotted backgrounds drastically increase storage requirements.

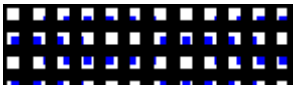
Original color.



A graphic wash of the same color. This is printed as a *screen*, meaning that it is actually finely spaced dots of a dark color, mixed with the paper color to create a lighter color.

Same as above, but magnified five times.

Problem: Your scanner's resolution is different from the printing company's screen. Visualize this as a mask or template with holes punched through.



Imagine this template placed over the image to be scanned. The scanner sees the graphic wash as white in some areas, and blue in others, because the patterns do not match.

The result is a *moiré pattern*, which is an uneven pattern of noise on the scanned image. This reduces FORMS' interpretation rates. When you encounter this problem on existing forms, use a despeckle filter to remove the noise.

Therefore, if you want a colorful form, use actual different colors rather than trying to save money by using a graphic wash. And when you order colored forms from your printing company, specify that use of a graphic wash to create the impression of different colors is not acceptable.

Print all forms in one batch

All forms to be processed using a single form definition should be printed at the same time and by the same printing company. Otherwise, minute differences on the forms might affect FORMS' ability to handle them with a single form definition. Adjustment objects are particularly sensitive to minute changes. Printing recognition fields on all forms is one way of protecting yourself from the consequences of this situation. See “Optimal design of recognition fields” on page 10 for more information.

Conclusion

Following your purchase of FORMS and the scanning hardware, you can process any existing form. At the same time, a good form design will help you obtain the maximum efficiency, effectiveness, and performance that FORMS has to offer. The relatively small amount of extra time required to produce forms tailored for FORMS repays itself in increased throughput, lower failure rates, and better overall workflow.

If you are faced with difficult choices when creating a form, please feel free to call or contact ReadSoft for more advice. We will be happy to help.

For additional information, please refer to *FORMS Help*. There you can find information about the optimal design of each field type, as well as other topics that are not covered in this guide. For example, look up *Optimize* in the help index and read topics that describe how to improve *form definitions* and improve *interpretation*.