The preparation of documents for electronic distribution[1]

By Frédéric PELLETIER and Daniel POULIN

[1]   Distribution of documents in printed form is characterized by its simplicity. The recipient will only see what the printed version of the document shows and will have access to no more information than the words appearing on the pages. This situation is far different from that of electronic distribution of documents.

[2]   The visual presentation of electronic documents is more flexible. The appearance of the document may change, but, more significant for our purpose, is the fact that electronic documents may contain various types of more or less hidden information. In fact, documents prepared with office software such as word processors, spreadsheet and presentation applications insert into files information commonly referred to as "metadata". Since sharing such documents may involuntarily reveal information to the recipient, it is of the utmost importance that necessary precautions are taken when preparing them, especially if they are to be distributed to the public.

[3]   After an overview of the problem of the presence of metadata in electronic documents (I), we will suggest procedures to follow while preparing such documents for electronic distribution, in order to reduce the risk of distributing documents containing confidential information (II). We will then describe software tools that automatically minimize metadata in documents (III). Finally, we will show different possible approaches to the insertion of objects in documents (IV).

[4]   To provide the reader with sufficient detailed explanations, the authors have only considered the problems related to the use of the most popular word processors: Corel WordPerfect ("WordPerfect") and Microsoft Word ("Word"). Those who use other office software will have to adjust the instructions accordingly.

I. Exchange of electronic documents and the risk of involuntary disclosure of information

[5]   An electronic document contains much more information than its on-screen or printed form may reveal. Users of most popular applications suites such as WordPerfect Office and Microsoft Office have to take some precautions to prevent a situation where the presence of metadata in distributed files threatens confidentiality.

A. "Metadata"

[6]   Generally speaking, the term "metadata" means: information regarding the nature of other information. In the publishing world, it relates more precisely to information regarding a document. For simplicity, we will use the term in its widest possible meaning, which includes all hidden text added by software to the visible text of a document.

[7]   Documents prepared with word processing software often contain metadata, even if only for determining the visual presentation of text. Over the years, increasing sophistication of office software demanded an increased use of metadata. This information is no longer limited to the visual presentation of documents, but is also now used to facilitate their classification, retrieval, visualization, editing and conversion.

[8]   We can categorize metadata under three types: text format attributes, document properties and editing data. **Text format** attributes include, for instance, textual characters and paragraph attributes; and the document format or the version of the application used to create it. Generally, this type of data is not confidential. **Document properties** consist of statistics, a summary of the document and personal information about the software user. **Editing data** includes text fragments and annotations resulting from the editing operations used in some advanced software features.

[9]   Here is a list, for each type previously described, of the most important metadata that WordPerfect and Word may insert in a document:

1. Text format attributes:
    - Font attributes (size, family…)
    - Paragraph attributes (spacing, style…)
    - Version of the software in use

2. Document properties:
    - User's (or "author's") name and initials
    - Reviewers (or past "authors") names and initials
    - Name of the user's workplace
    - Name of the user's computer
    - Document summary and description
    - Names of templates and styles used to create the document

- Document's creation, modification and consultation dates
- Routing slip
- File size and path
- Document statistics (page, word, character count…)
- Properties of embedded objects

3. Editing data:
   - Former versions of the text
   - Removed or added text fragments
   - Revision marks and annotations
   - Undo/redo history
   - Comments
   - Hidden text
   - Field codes from hyperlinks and linked objects
   - Unique numerical identifier

[10]   It is easy to see how the presence of such information in distributed documents can lead to the disclosure of useless or undesirable information. Let's take the example of a lawyer retained by a client to write a contract. A colleague who recently completed a similar retainer for another client proposes to take his version as a precedent. He then transmits an electronic version of the precedent in order to accelerate the preparation of the document. Our lawyer opens the document and, by modifying several clauses, adapts it to fit the particular needs of the client. She therefore writes a new contract, which she then transmits for approval to the client by e-mail. The client reviews the document, makes some modifications, and adds comments of strategic nature in order to guide the lawyer in the preparation of the final version. He also adds a table by copying it from a worksheet opened with a spreadsheet application. After a few exchanges of the document between the lawyer and client, the final version is sent by e-mail to the other contracting party's lawyer. In some circumstances, if no precautions are taken by our lawyer before distributing this document, it could contain the following information:

- Names of the lawyers that have worked on the document since its creation and names of the legal offices in which they practice
- Names of the clients that have worked on the document since its creation and names of their workplaces
- Document text fragments that have been added then deleted since its creation
- Former document versions
- Content of workbook tabs in which we find the worksheet that was used to add the table to the document
- Names and addresses of the senders and recipients of the document transmitted by e-mail

[11]   It is important to remember that presentation software (e.g. Corel Presentations and Microsoft PowerPoint) and spreadsheets (e.g. Corel Quattro Pro and Microsoft Excel) also insert this type of metadata in documents. Therefore, although the rest of this text is

centered only on word processing software, the same caution is needed regarding the distribution of documents produced with other office applications.

B. Access to potentially confidential metadata

[12]   Confidential metadata associated with a document can generally be consulted using the same word processing software used to create the document. It is however impossible to check for the presence of all the information using only this software. The use of a generic text editor may be necessary.

[13]   First, let us distinguish two types of software: word processing and generic text editor. The word processing software makes it possible to write text in a file and then control its appearance. Thus two types of information are written in the file: text characters typed with the keyboard, and data regarding text format or previously mentioned metadata. Those are often embedded in the form of proprietary codes unique to this software. A generic text editor only allows the entry of textual characters in any file. Only the typed text is written to the file, and it is impossible to specify a particular visual presentation for this text.

1. Using a generic text editor

[14]   The generic text editor gives access the all the information contained in a file. It is not a user-friendly way of consulting a document prepared with word processing software since the sequence of characters and lines does not necessarily reflect the structure and contents of the document that was created with its original application. Moreover, some data is not presented in an intelligible form. It is, however, possible to read very substantial sequences of characters. An experienced user can decipher and retrieve a great deal of information that is invisible to most users.

2. Using word processor features

[15]   When a document is opened with the word processor that was used for its creation, the majority of the metadata it contains may be consulted using the same features that were used to insert them, by the application menus. However, there remains metadata that cannot be accessed this way. For example, Word inserts a list of the last 10 "authors" of the document. This list is not accessible from the Word menus, but it is possible to see it by opening the document with a generic text editor.

[16]   A user's personal information can be consulted using the "file/properties" feature. Editing data that is potentially confidential can generally be viewed using the menus with which the corresponding features were inserted. The following paragraphs briefly describe these features.

a) Properties (**Word** and **WordPerfect**)

[17] The **Properties** feature makes it possible to associate to the file some document data, including statistics and a user's personal information. For example, while saving the document using Word or WordPerfect, the name of the user is by default inserted into a field called **Author**.

[18] With Word, properties are accessible by clicking on **File ► Properties**, and then by looking at the **General**, **Summary**, **Statistics**, **Contents** and **Custom** tabs. Personal information is inserted by default into the fields **Author** and **Company** of the **Summary** tab and in the **Last saved by** field of the **Statistics** tab.

[19] Using WordPerfect, the properties are visible by clicking on **File ► Properties…**. Personal information is found in the **Summary** tab. The **Author** and **Typist** fields are filled by default during the document creation. The **Information** tab contains document statistics.

b) Fast saves (**Word**)

[20] With Word's **Fast saves** feature, saving a file is faster since instead of replacing the old version of the document by the new one, the application only appends changes to the end of the file. The effect is that text fragments that had been removed and do not appear on screen are still present in the document. It is impossible to access these text fragments using the Word menus.

c) Undo/redo history (**WordPerfect**)

[21] This WordPerfect feature makes it possible for the user to undo or redo up to 300 editing operations made on a document. The user can also choose to save this history with the document. The consultation of the undo/redo history is done by clicking on **Edit ► Undo/redo history…**. It is also possible to see one by one these operations by clicking on **Edit ► Undo** or **Edit ► Redo**.

d) Versioning (**Word** and **WordPerfect**)

[22] Both Word and WordPerfect make it possible to save several versions of a document within the same file. Saving a former document version adds the complete text of this version in the file.

[23] With Word, former versions may be consulted by clicking on **File ► Versions…**. If previous versions were saved with the document, a list of the versions appears in the dialog box.

[24] In WordPerfect, the **Version Control** feature is not available through the "standard" installation; it had to have been specifically chosen using the "personalized" installation. Without this feature, the user does not have access to possible former versions recorded by another user in a document, nor can the user check for the presence

of old versions in a document. After installing the **Version control** feature, it is possible to access the saved versions by clicking on **File ► Version control ► Retrieve Current…** If the document contains former versions, the **Retrieve version** dialog box displays a list; if not, a message indicates that the document does not contain any revisions.

e) Document Review, Track Changes, Comments and Hidden text (**Word** and **WordPerfect**)

[25] These Word or WordPerfect features generate text in the form of revision annotations, comments or hidden text. Often useful to the author or group of authors in preparing a document, this information is normally not intended for distribution with its final version.

[26] In certain circumstances, this information is not shown during document preparation. For example, using Word, the display of revision marks, comments and the formatting marks on screen can depend on the software version. Let us suppose that a user revises a document while inserting some comments, and then sends it back to its author without informing him that this document contains comments. These comments may not be displayed on the author's screen, and the document's final version will be distributed with these comments.

[27] Fortunately, these features are not activated by default with Word or WordPerfect installation. If only one user has worked on a new document, he or she doesn't have to fear the inclusion of revision marks, comments or hidden text in the document without his knowledge. On the other hand, if the document was created or modified by another user, it is advisable to make sure that the document is clear before its distribution.

[28] Using Word 2002, it is possible to see this type of metadata by clicking on:

- **View ► Markup** to see revision marks and comments; and on
- **Tools ► Options…**, and then in the **View** tab, checking the *Hidden Text* box to display hidden text with dotted underline.

[29] Using Word 97 and 2000, it is possible to see this type of metadata by clicking on:

- **View ► Comments** to see comments, unless it is impossible to click on this button, which means that the document does not contain any comment; on
- **Tools ► Track Changes ► Highlight Changes**, and by checking the *Highlight changes on screen* box to see revision marks; and on
- **Tools ► Options…**, and then in the **View** tab, checking the *Hidden Text* box to display hidden text with dotted underline.

[30] Using WordPerfect, this information is visible by clicking on:

- The balloon shaped icons or grey shaded boxes in the document margin to see the comments; on
- **File ► Document ► Review…**, and on the **Author…** button to see revision annotations; and on
- **View ► Hidden Text** to display hidden text with dotted underline.

f) Insertion of objects (**Word** and **WordPerfect**)

[31] WordPerfect and Word allow the insertion of objects in a document. An object is any set of information (e.g. a picture, a worksheet, a diagram or table) whose "source" is another document. The inserted object is either "embedded", "linked" or simply "pasted" in the document. Only the simply pasted object is free of metadata.

[32] With inserted objects, it is not always easy to know which type it is. Moreover, metadata associated with the object may be difficult, if not impossible, to access. This is why it is recommended to simply paste objects as pictures or text without linking or embedding them; see section IV.

C. How to avoid the distribution of files containing metadata

[33] Since any file created or modified with word processing software always contains some metadata, it is impossible to completely eliminate it. In a legal context, the objective is to minimize the presence of metadata in distributed files in order to avoid the involuntary disclosure of confidential information.

[34] The nature of the precautions to be taken to minimize metadata in document preparation depends on each legal institution. Indeed, the confidentiality of information may vary according to the work environment. For example, the file path of a document can contain totally neutral terms, but in other cases this type of information can generate breaches of confidentiality or embarrassing misunderstandings for the author. Moreover, the specific preparation methods of each organization may attract different consequences respectiong the risks of involuntary insertion of metadata in documents. In general, the more people who work on a document, the greater the risk.

[35] The better solution consists of minimizing the upstream insertion of metadata regarding all newly created documents on a workstation. This objective can be carried out with the workstation configuration of each software user or by the systematic use of a document template especially designed to minimize metadata insertion.

[36] In organizations where metadata is used for document workflow management purposes and, to the contrary, in decentralized organizations where the standardization or preparation procedures is difficult, it will be necessary to work out a particular process to "clean" each file, downstream, before it is distributed. This can be done manually, following the procedures described in Section II. It is also possible and often desirable to do it automatically using the software tools mentioned in Section III.

II. Metadata minimization procedures in Word and WordPerfect documents

[37] This section offers a description of the procedures to be followed in order to minimize the presence of metadata likely to generate involuntary disclosure of information in the documents prepared using the most popular word processing software, namely Microsoft Word and Corel WordPerfect. It describes a series of distinct procedures for Word 2002 (A), Word 97, Word 2000 (B), and WordPerfect (C). Certain procedures are aimed at the workstation configuration in order to avoid upstream insertion of metadata; others are aimed at the downstream cleaning of documents which already contain metadata.

[38] Who is the person best placed to carry out metadata minimization? Although suppression of most metadata doesn't have any impact on the text content and can be done by office staff, the suppression of editing data requires decision-making regarding the final content of the document. This is why we establish a distinction between the person who is responsible for the content of the document and the one who is responsible for its distribution.

[39] The person "responsible for the contents" is the text author or its final reviser. It is, for example, the judge who works on the judgment, either by creating the electronic document using word processing software, or by revising a previously handwritten or transcribed electronic document.

[40] The person "responsible for the distribution" is the one in charge of the final material preparation of the electronic document. This person may be the same as the one responsible for the contents, but in most legal institutions it is a member of the court's office staff.

[41] Preliminary remarks:

1) The following procedures are aimed at metadata most likely to escape the author's control. For further details, we suggest consulting online documentation available through the distributors of Word and WordPerfect:

- *OFF: How to Minimize Metadata in Microsoft Office Documents (Q223396)*, URL: http://support.microsoft.com/default.aspx?scid=kb;en-us;Q223396;
- *Minimizing metadata in WordPerfect 10 documents*, URL: http://cache.corel.com/Storage/CorDocument/Minimizing_metadata_in_WordPerfect_10_documents,0.pdf.

2) The problem arising from the confidentiality of metadata is becoming better and better known in the office suites industry. Consequently, software distributors are gradually facilitating the management of metadata in most recent versions of their office applications. This is particularly the case for Word's 2002 version (included in Office

XP) and also for WordPerfect's versions 9 and 10. That is why it is recommended to use the most recent version of the software.

A. For Word 2002 users

[42] After having their workstation configured (1), software users may create and modify existing documents and avoid the insertion of several metadata. It is however possible that some remains. The person responsible for the document contents will have to ensure that the editing data is gone (2), and then other metadata minimization can be carried out by the person responsible for the document's distribution (3).

1. Workstation configuration

[43] Word 2002 has "Confidentiality options" which facilitate metadata management. It is now possible to remove personal information when saving a document. Moreover, the user can choose to be warned of the presence of editing data before distributing a document. In addition, although the fast saves feature is inactive by default, it is safer to make sure that it is turned off. Lastly, it is preferable to make sure that hidden text is always displayed on screen.

[44] These workstation configuration procedures are addressed to both the person responsible for the document contents and the one responsible for its distribution.

a) Removing personal information when saving

[45] In Word 2002, this option is associated to the current document. For systematic suppression of personal information in all documents based on a given template, it is necessary to modify this template. The Word default document template is a file named "Normal.dot" and is found in a folder named "Templates". To find its exact location, which can vary from one environment to another, click on **Tools ► Options…**, then under the **File Locations** tab, follow the access path associated with *User Templates*.

[46] To systematically remove personal information of any document based on a given document template:

1. Open the document template;
2. Click on **Tools ► Options…**;
3. In the **Security** tab, check the ***Remove personal information from this file on save*** box;
4. Click on **OK**;
5. Save the document template.

b) Activating the warning prior to distribution option

[47]   When a user chooses to be warned when revision marks or comments are present in any document before distribution, he or she does so for all the documents opened on the same workstation by the same user. To activate this option:

1. Open the document;
2. Click on **Tools ► Options…**;
3. In the **Security** tab, check the ***Warn before printing, saving or sending a file that contains changes or comments*** box;
4. Click on **OK**.

c) Deactivating fast saves

[48]   Deactivating fast save affects all the documents opened or created on the same workstation by the same user. Any save will thereafter be a *full save*. To turn off fast save:

1. Open a document;
2. Click on **Tools ► Options…**;
3. In the **Save tab**, make sure that the ***Allow fast save*** check box is clear;
4. Click on **OK**;

d) Displaying hidden text

[49]   The systematic on-screen display of hidden text ensures this information is not hidden to the user. The displaying of hidden text applies to all documents opened or created by the same user on the same workstation. To turn on the display of hidden text:

1. Open a document;
2. Click on **Tools ► Options…**;
3. In the **View** tab, check the ***Hidden text*** box;
4. Click on **OK**;

2. Procedures to be executed under the supervision of the person responsible for the contents

[50]   The use of advanced editing features by an author or a group of authors and revisers implies potential insertion of several editing data in the prepared documents. When the person responsible for the document contents is not the only one to have worked on it, he or she must look out for the presence of text fragments and if necessary remove them from the document. This person must also make sure that objects were suitably inserted. The following procedures must be carried out under his or her supervision, since they can modify the contents of the document.

a) Removing revision marks and comments, and deactivating the Track changes feature

[51] If necessary, the revision marks and comments must be removed. With Word 2002, this is managed using the **Reviewing** toolbar, which contains the following buttons:

 .

[52] To display this toolbar, just open a document that contains revision marks or comments. If it is not displayed, click on **View ► Toolbars ► Reviewing**. The **Reviewing** toolbar make it possible to browse from one annotation or comment to another, to accept or reject changes and to remove comments. It also makes it possible to deactivate the Track changes feature.

[53] To accept or reject changes, remove the comments and deactivate Track changes:

1. Open the document;
2. Place the cursor at the beginning of the document;
3. On the **Reviewing** toolbar, click on the suitable buttons so that you can accept or reject changes or remove comments, until exhaustion of all revision marks and comments;
4. On the **Reviewing** toolbar, if the **Track Changes** button is framed as follows: , click on it to turn off this feature;
5. Save the document.

b) Removing hidden text

[54] Hidden text in the document must be removed or inserted in the document. It is displayed on the screen with a dotted underlining. In order to find all hidden text in a document, using Word's **Find** function can be very helpful. To seek, then remove or reveal hidden text in a document:

1. Open the document;
2. Click on **Edition ► Find…**;
3. In the **Find and Replace** dialog box, click on the **More** button in the **Find** tab;
4. Click in the **Find** text box;
5. In the **Find and Replace** dialog box, click on **Format ► Font…**;
6. In the **Find Font** dialog box, locate the **Effects** list of the **Font, style and Effects** tab, and check the *Hidden* box;
7. Click on **OK**;
8. In the **Find and replace** dialog box, click on the **Next** button of the **Find** tab;
9. Either erase the hidden text or reveal it by turning its the *Hidden* format;
10. Repeat operations 8 and 9 until depletion of the hidden text;
11. Save document.

c) Converting embedded or linked objects

[55] If an object was embedded in or linked to the document, it must be converted into a simple pasted object as text or picture, not linked, as explained in section IV. If the embedded or linked object is a graphic such as a diagram, it is better to convert it into a picture. If it is text (such as a worksheet zone), it is preferable to convert it into formatted text, thus preserving the font and table attributes. To convert an embedded or linked object:

1. Open the document;
2. Click on the object and cut it by clicking on **Edit► Cut**;
3. Click on **Edit ► Paste special…**;
4. In the **Paste special** dialog box, choose the appropriate object format:
   a. *Paste / As / Picture (Enhanced metafile)* for graphic objects
   b. *Paste /As / Formatted Text (RTF)* for textual objects;
5. Click on **OK**;
6. Save the document.

3. Procedures to be executed by the person responsible for the distribution

[56] The following procedures may be executed by office staff since they do not have any impact on the displayed content of the document.

a) Executing a full save

[57] Using Word, it is not possible to check if an existing document was saved with the Fast save option turned on. It is necessary to make sure that the document undergoes at least one full save. To execute a full save on a document before its distribution:

1. Open the document;
2. Make sure that Fast saves are turned off;
3. Make some minor changes in the document, e.g. type then remove characters at the end of the file;
4. Save the document.

b) Deleting or changing document properties

[58] It is advisable to make sure that document properties do not contain information not to be distributed. To remove or change document properties:

1. Open the document;
2. Click on **File ► Properties**;
3. In the **Summary**, **Contents** and **Custom** tabs, delete or change contents of the text boxes, as needed;
4. Click on **OK**;

5. Save the document.

c) Deactivating automatic versioning and deleting former versions

[59] A distributed document must contain only its current version, and the automatic version saving option must be turned off. To deactivate automatic versioning and delete former versions:

1. Open the document;
2. Click on **File ► Versions…**;
3. Make sure that the *Automatically save a version on close* check box is cleared;
4. Select **Existing Versions**, if applicable, and then click on **delete**;
5. Click on **Close**;
6. Save the document.

B. For Word 97 and 2000 users

[60] After having their workstation configured (1), users may create and modify existing documents and avoid the insertion of several types of metadata. It is however possible that some remains. The person responsible for the document contents will have to ensure that the editing data is gone (2), and then other metadata minimization can be carried out by the person responsible for the document's distribution (3).

1. Workstation configuration

[61] In order to avoid the systematic insertion of his or her name and initials as the document's reviser or author, the user of Word 97 or 2000 must remove or change some information in the profile. It is also safer to make sure that the Fast save option is inactive and to make sure that hidden text is always displayed on screen.

[62] These workstation configuration procedures are addressed to both the person responsible for the document contents and the one responsible for its distribution.

a) Removing user name and initials

[63] To remove or modify the user name and initials systematically inserted in the Properties all new documents:

1. Open the document;
2. Click on **Tools ► Options…**;
3. In the **User** tab, make the needed changes in the **Name**, **Initials** and **Address** text boxes, by typing at least one space or character;
4. Click on **OK**

b) Deactivating Fast saves

[64]   Deactivating Fast saves affects all the documents opened or created on the same workstation by the same user. Any save will thereafter be a *full save*. The **Allow fast saves** option is activated by default when installing the first distribution of Word 97. The users of this version should be particularly vigilant. In the 2000 and 2002 versions, it is turned off by default. To turn off Fast saves:

1. Open a document;
2. Click on **Tools ► Options…**;
3. In the **Save** tab, make sure the ***Allow fast saves*** check box is cleared;
4. Click on **OK**;

c) Displaying hidden text

[65]   The systematic on-screen display of hidden text prevents this information to be unseen by the user. The displaying of hidden text applies to all documents opened or created by the same user on the same workstation. To turn on the display of hidden text:

1. Open a document;
2. Click on **Tools ► Options…**;
3. In the **View** tab, check the ***Hidden text*** box;
4. Click on **OK**;

2. Procedures to be executed under the supervision of the person responsible for the contents

[66]   The use of advanced editing features by an author or a group of authors and revisers implies potential insertion of some editing data in the prepared documents. When the person responsible for the document contents is not the only one to have worked on it, he or she must look out for the presence of text fragments and if necessary remove them from the document. This person must also make sure that objects were properly inserted. The following procedures must be carried out under his or her supervision, since they can modify the contents of the document.

a) Removing revision marks and deactivating the Track changes feature

[67]   Where necessary, revision marks must be removed, and the Track changes feature deactivated. To remove revision marks and turn off Track changes:

1. Open the document;
2. Click on **Tools ► Track changes ► Highlight Changes**;
3. In the **Highlight Changes** dialog box, make sure the ***Highlight changes on screen*** box is checked, and that the ***Track changes while editing*** check box is cleared;
4. Click on **OK**;

5. Click on **Tools ► Track changes ► Accept or Rejects…**;
6. In the **Accept or Rejects Changes** dialog box, click on the appropriate buttons to accept or reject the modifications, unless it is impossible to click on these buttons, which means that the document does not contain any revision marks;
7. Click on **Close**;
8. Save the document.

b) Deleting comments

[68] Where necessary, comments can be removed. To remove comments:

1. Open the document;
2. Click on **View ► Comments**, unless it is impossible to click on this button, which means that the document does not contain any comments;
3. Click in the document window, at the beginning of the first page;
4. Click on the **Next Comment** button in the **Revision** toolbar;
5. Click on the **Delete Comment** button in the **Revision** toolbar;
6. Repeat steps 4 and 5 until all comments are deleted;
7. Save the document.

c) Removing hidden text

[69] Hidden text in the document must be removed or inserted in the document. It is displayed on the screen with a dotted underlining. In order to make sure that there remains no trace of hidden text in a document, using Word's **Find** feature can be very helpful. To seek, then remove or reveal hidden text in a document:

1. Open the document;
2. Click on **Edition ► Find…**;
3. In the **Find and Replace** dialog box, click on the **More** button in the **Find** tab;
4. Click in the **Find** text box;
5. In the **Find and Replace** dialog box, click on **Format ► Font…**;
6. In the **Find Font** dialog box, locate the **Effects** list of the **Font, style and Effects** tab, and check the *Hidden* box;
7. Click on **OK**;
8. In the **Find and replace** dialog box, click on the **Next** button of the **Find** tab;
9. Either erase the hidden text or reveal it by turning off its *Hidden* format;
10. Repeat operations 8 and 9 until no hidden text remains;
11. Save document.

d) Converting embedded or linked objects

[70] If an object was embedded or linked to the document, it must be converted into a simple pasted object as text or picture, not linked, as explained in section IV. If the embedded or linked object is graphic such as a diagram, it is better to convert it into a picture. If it is text (such as a worksheet zone), it is preferable to convert it into a

formatted text, thus preserving the font and table attributes. To convert an embedded or linked object:

1. Open the document;
2. Click on the object and cut it by clicking on **Edit ► Cut**;
3. Click on **Edit ► Paste special…**;
4. In the **Paste special** dialog box, choose the appropriate object format:
    a. *Paste / As / Picture (Enhanced metafile)* for graphic objects
    b. *Paste /As / Formatted Text (RTF)* for textual objects;
5. Click on **OK**;
6. Save the document.

3. Procedures to be executed by the person responsible for distribution

[71]   The following procedures, which may be executed by office staff, do not have any impact on the displayed content of the document.

a) Executing a full save

[72]   Using Word, it is not possible to check if an existing document was saved with the Fast saves option turned on. It is necessary to make sure that the document undergoes at least one full save. To execute a full save on a document before its distribution:

1. Open the document;
2. Make sure that Fast saves are turned off;
3. Make some minor changes in the document, e.g. type then remove characters at the end of the file;
4. Save the document.

b) Deleting or changing document properties

[73]   It is advisable to make sure that the document properties do not contain information not to be distributed. To remove or change document properties:

1. Open the document;
2. Click on **File ► Properties**;
3. In the **Summary**, **Contents** and **Custom** tabs, delete or change contents of the text boxes, as needed;
4. Click on **OK**;
5. Save the document.

c) Deactivating automatic versioning and delete former versions

[74]   A distributed document must contain only its current version, and the automatic version saving option must be turned off. To deactivate automatic versioning and delete former versions:

1. Open the document;
2. Click on **File ► Versions…**;
3. Make sure that the ***Automatically save a version on close*** check box is cleared;
4. Select **Existing Versions**, if applicable, and then click on **delete**;
5. Click on **Close**;
6. Save the document.

C. For WordPerfect 8, 9 and 10 users

[75]   After having their workstation configured (1), users may create and modify existing documents and avoid the insertion of several metadata. It is however possible that some remains. The person responsible for the document contents will have to ensure that the editing data is gone (2), and then other metadata minimization can be carried out by the person responsible for the document's distribution (3).

1. Workstation configuration

[76]   In order to avoid the systematic insertion of his or her name and initials as the document's reviser or author, the WordPerfect user must remove or change some information in the profile. It is also safer to make sure that the Version control feature is installed, the history undo/redo feature is inactive and that hidden text is always displayed on-screen.

[77]   These workstation configuration procedures are addressed to both the person responsible for the document contents and the one responsible for its distribution.

a) Installing Version control

[78]   In order to be able to check if a document contains old versions, the WordPerfect user must first have the Version control feature installed on his workstation. To install Version control:

1. Insert WordPerfect's installation software on the workstation;
2. Follow the installation wizard's instructions;
3. When choosing the installation components, consult the **Utilities** list, and check the ***Corel Versions*** box;
4. Resume installation following the on-screen instructions.

b) Removing user name and initials

[79]   To remove or modify the user name and initials systematically inserted into the Properties all new documents:

1. Open the document;
2. Click on **Tools ► Settings… ► Environment**;

3. In the **General** tab, make the needed changes in the **Name** and **Initials** text boxes, by typing at least one space or character;
4. Click on **OK**;
5. Click on **Close**.

c) Deactivating undo/redo history saving

[80]   Deactivating the history undo/redo saving option affects all document consulted or created by a user on the same workstation. When installing versions 6, 7 and 8 of WordPerfect, the undo/redo history saving option is turned on by default for the 10 last operations. In the 9 and 10 versions, it is turned off by default. To make sure that undo/redo history saving option is deactivated on a workstation:

1. Open the document;
2. Click on **Edit ► Undo/Redo History…**;
3. In the **Undo/Redo History** dialog box, click on **Options…**;
4. In the **Undo/Redo Options** dialog box, make sure the **Save Undo/Redo items with document** check box is cleared;
5. Click on **OK**.

d) Displaying hidden text

[81]   The systematic on-screen display of hidden text prevents this information from being hidden to the user. The display of hidden text applies to all documents opened or created by the same user on the same workstation. To turn on the display of hidden text:

1. Open the document;
2. Click on **Tools ► Settings… ► View**;
3. In the **Display Settings** dialog box, check the *Hidden text* box;
4. Click on **OK**;

2. Procedures to be executed under the supervision of the person responsible for the contents

[82]   The use of advanced editing features by an author or a group of authors and revisers implies potential insertion of some editing data in the prepared documents. When the person responsible for the document contents is not the only one to have worked on it, he or she must look out for the presence of text fragments and if necessary remove them from the document. This person must also make sure that objects were properly inserted. The following procedures must be carried out under his or her supervision, since they can modify the contents of the document.

a) Removing or inserting revision annotations

[83]   Where necessary, revision marks must be removed. With WordPerfect 9 and 10 versions, the **Review Document** dialog box appears systematically when opening a

document which contains revision annotations. If the box does not appear, the document has no revision annotations. Using version 8, the presence of revision annotations could be verified by clicking on **File ► Document ► Review…**. If choosing **Review…** proves to be impossible, then the document has no revision annotations.

[84] When an author revises a document, the revision toolbar shows the following buttons: .

[85] These buttons make it possible to browse from one annotation to another and to insert or remove them, all at once or one by one. To insert or remove revision annotations in a document:

1. Open the document;
2. In the **Review Document** dialog box, click on **Author…**;
3. On the revision toolbar, click on the appropriate buttons to insert or delete annotations either all at once or one by one;
4. Save the document.

b) Deleting comments

[86] Where necessary, comments can be removed. To remove comments:

1. Open the document;
2. Place the cursor at the beginning of the document;
3. Click on **View ► Reveal codes**;
4. Click on **Edit ► Find and replace…**;
5. In the **Find and replace…** dialog box, click on **Match ► Codes…**;
6. In the **Codes** dialog box, choose **Comment** in the **Find Codes** list;
7. Click on the **Insert and close** button;
8. In the **Find and replace…** dialog box, click on **Next**;
9. Read the comment by clicking on his icon;
10. To delete a comment, click on **Replace**;
11. Repeat operations 8 to 10 until all comments are deleted;
12. Click on **Close**;
13. Save the document.

c) Removing hidden text

[87] Hidden text in the document must be removed or inserted in the document. In order to make sure that there remains no trace of hidden text in a document, using WordPerfect's **Find and Replace** feature can be very helpful. To seek, then remove or reveal hidden text in a document:

1. Open the document;
2. Place the cursor at the beginning of the document;
3. Click on **View ► Reveal Codes**;

4. Click on **Edit ► Find and replace…**;
5. In the **Find and replace…** dialog box, click on **Match ► Codes…**;
6. In the **Codes** dialog box, choose **Hidden on** in the **Find Codes** list;
7. Click on the **Insert and Close** button;
8. In the **Find and replace…** dialog box, click on **Next**;
9. Read the text by clicking on the document window;
10. Either erase the hidden text or reveal it by turning off its *Hidden* format;
11. Repeat steps 9 and 10 until no hidden text remains;
12. Click on **Close**;
13. Save the document.

d) Converting embedded or linked objects

[88] If an object was embedded or linked to the document, it must be converted into a simple pasted object as text or picture, not linked, as explained in section IV. If the embedded or linked object is a graphic such as a diagram, it is better to convert it into a picture. If it is text (such as a worksheet zone), it is preferable to convert it into a formatted text, thus preserving the font and table attributes. To convert an embedded or linked object:

1. Open the document;
2. Click on the object and cut it by clicking on **Edit ► Cut**;
3. Click on **Edit► Paste special…**;
4. In the **Paste special** dialog box, choose the appropriate object format:
    a. *Paste / As / Picture* for graphic objects, or
    b. *Paste / As / Rich Text Format* for textual objects;
5. Click on **OK**;
6. Save the document.

3. Procedures to be executed by the person responsible for the distribution

[89] The following procedures, which may be executed by office staff, do not have any impact on the displayed content of the document.

a) Deleting the undo/redo history

[90] If the user can undo editing changes on a document as soon as it is opened in WordPerfect, this means that an undo/redo history was saved within it. To remove the undo/redo history of a document:

1. Open the document;
2. Make sure, if needed, that the undo/redo history saving option is turned off;
3. Save the document.

b) Deleting or changing document properties

[91] It is advisable to make sure that the document properties do not contain information not to be distributed. To remove or change document properties:

1. Open the document;
2. Click on **File ► Properties…**;
3. In the **Summary** tab, delete or change contents of the text boxes, as needed, especially the **Author** and **Typist** text boxes;
4. Click on **OK**;
5. Save the document.

c) Deleting document former versions

[92] To verify if the document contains former versions to be deleted, click on **File ► Version Control ► Retrieve Current…**. If necessary, remove the former versions and keep the current one. To delete the former versions:

1. Open the document;
2. Click on **File ► Save As…**;
3. In the **Save As** dialog box, enter a new file name or keep the same name but choose a different folder;
4. Click on **OK**;
5. Destroy the old document and rename the document to be distributed, if needed.

III. Software tools allowing the automation of the metadata minimization process

[93] Taking into account the risk incurred with the presence of metadata in distributed documents and the extent of work needed to clean documents, it is preferable for certain organizations to be provided with a software tool specifically designed to minimize the metadata in produced documents.

[94] In addition to the productivity gain generated by automating the process, software tools make it possible to remove information otherwise impossible to clear from documents. It is notably the case for the list of the last ten authors of a document, which is impossible to delete with Word.

[95] Word users benefit from complete software tools (A), while WordPerfect users must install macros on their workstations (B).

A. For Word users

[96] Word users have access to various commercial applications in order to minimize the metadata of their documents. Among those:

- Metadata Assistant, by Payne Consulting (see http://www.payneconsulting.com);
- Out-of-Sight, by SoftWise (see http://www.softwise.net); and
- ezClean, by KKLSoftware (see http://www.kklsoftware.com).

[97] These applications provide a complete metadata management utility as to detect, remove or convert them. Word and Excel documents can be cleaned with the click of a mouse when they are opened in these applications, or when they are sent as an attachment to an Outlook message.

[98] It should be noted that ezClean can process only the active document opened in Word and Excel, or sent by Outlook. Metadata Assistant and Out-of-Sight can process not only the active document, but also a batch of files in a given folder including, if needed, sub-folders. The possibility of processing batches of files may be a decisive factor for certain organizations which require the systematic cleaning of a great quantity of documents.

[99] Metadata Assistant distinguishes itself from the other two applications in the fact that it can be executed standalone, whereas Out-of-Sight and ezClean only work through Word, Excel or Outlook. Moreover, Metadata Assistant can manage documents protected by a password and has generally more options that facilitate the management of metadata minimization.

B. For WordPerfect users

[100] WordPerfect users do not have the benefit of particular software tools allowing metadata management. However, they can install specifically designed macro commands on their workstation to clean WordPerfect files.

[101] Some Internet resources freely distribute such macro commands, like the tips page from *Barry McDonnell's Toolbox* at the following address:

 <http://ourworld.compuserve.com/homepages/macdonnell/Tips/UndoRedo.html>

IV. Illustration of external object insertion methods in a document

[102] Word and WordPerfect allow "object" insertions in a document. Such insertions can be carried out in various ways in order to meet the needs of each type of software user. For the legal community, it is strongly recommended that techniques be used that limit the information added to the exchanged documents to what the author really intends to transmit.

[103] For the rest of the discussion and by way of illustration, we will take the example of a Word user who wishes to insert in his or her document a histogram prepared with Excel. We will describe three types of objects according to their method of insertion: "linked", "embedded" and "pasted". We will also discuss the advantages and disadvantages each of these types. The recommended approach – the insertion of a simply pasted object – does not introduce into the document any information other than what appears on the screen.

[104] Let us initially clarify the terminology. For the purpose of this document, "object" indicates any entity of information: fragment of text, histogram, or part of a spreadsheet. The document from which the object comes will be designated as the "source document" and the "target document" is the document that receives it. The application associated with the source document is the "server application", and the one associated with the target document is the "client application". In our example, the histogram is the object; the source document is a spreadsheet type document. Excel is the server application, whereas Word is the client application.

A. Insertion of the histogram as a "linked" object

[105] The object "linked" to the target document is in fact only a dynamic reference to the source document. The target document contains the access path and the format of the source file, but not the data itself. This linking information nevertheless makes possible the displaying in the target document of the information pointed out in the source document.

[106] This operation is executed by selecting in Excel a group of information, here a histogram, and by pasting it in Word using the **Special paste** feature of the **Edition** menu, while choosing "Paste with link".

[107] The principal advantage of this approach comes from the fact that the histogram will automatically be updated in the Word document whenever the data is modified in the Excel document.

[108] Documents that are to be distributed should not contain linked objects since the link between the target and the source document will break during distribution. Consequently, the readers of the document would not see the information the author wishes to present them.

B. Insertion of the histogram as an "embedded" object

[109] The "embedded" object in a target document inserts in this document not only the subset of information copied from the source document but also its relationship to the entirety of information contained in it. Thus, an embedded object introduces into the target document not only the copied and apparent object, but also all data used for its preparation, and even metadata associated with the source document.

[110] In the histogram example, its incorporation in a Word document involves the insertion - probably involuntary - not only of the whole content of the Excel tab to which the histogram is linked, but also of all the other tabs of the workbook and the metadata which could be attached to it.

[111] The advantage of this approach is to preserve the link between the embedded object and the software components which permit its manipulation and allow its modification in the target document by modifying the data that accompanied it. Since it can be harmful that a document final version can be modified and contain invisible data, it is highly recommended to avoid this method of insertion into distributed documents.

C. Insertion of the histogram as a "pasted" object, without linking or embedding it

[112] The simply "pasted" object is the copy of a source document which contains only its visible data. This data can be presented in the form of picture or formatted text. It is the recommended approach.

[113] This approach consists of using the methods of insertion that will produce only the desired effects, i.e. the addition in the target document of only the information selected in the source document.

[114] Here's how to proceed for our histogram and some other examples.

- The insertion of a histogram from an Excel document to a Word document:

  1. Select and copy the histogram in an Excel document;
  2. Go to Word and place the cursor at the insertion point;
  3. Click on **Edit ► Paste special…**;
  4. Choose *Paste / As / Picture (Enhanced Metafile)*;
  5. Click on **OK**.

- The insertion of a spreadsheet zone in a Word document:

  1. Select and copy the appropriate zone in the Excel document;
  2. Go to Word and place the cursor at the insertion point;
  3. Click on **Edit ► Paste special…**;
  4. Choose *Paste / As / Formatted Text (RTF)*;

5. Click on **OK**.

• The insertion of a PowerPoint image in a Word document :

1. Select and copy the appropriate zone in the PowerPoint document;
2. Go to Word and place the cursor at the insertion point;
3. Click on **Edit ► Paste special…**;
4. Choose *Paste / As / Picture (Enhanced Metafile)*;
5. Click on **OK**.

[115] These examples illustrate that it is possible to copy objects such as images or formatted text into a document without linking or embedding them.