ClinicalTrials.gov Protocol Redaction Report

Redaction Sub-Group Members:

Ian McGovern – Massachusetts General Hospital Cancer Center
Scott Patton – Stanford Center for Clinical & Translational Research & Education (SPECTRUM)
Alanda R Perry Jones - Virginia Commonwealth University
Jane Cho - Columbia University Medical Center
Melanie Chladny – University of Michigan
Michelle Morgan – MD Anderson Cancer Center

TABLE OF CONTENTS

NOTE: The information provided within this report is intended only to be used as general guidance. It is the responsibility of the 'Responsible Party' and/or the sponsor to ensure that all applicable laws and requirements are followed.

Document Redaction Overview

1. Required documents
    - A. Protocol, statistical analysis plan (SAP), and amendments
        - i. SAP can be part of the protocol (e.g., in Statistical Methods) or a separate document
        - ii. The final amended version that incorporates the original protocol and all amendments can be submitted
        - iii. See Section Required Documents for more information
2. Required format
    - A. Each document must include a cover page with official title, NCT number (if available), and version date of the document
    - B. Files must be in PDF/A format
    - C. See section Document Formatting for more information
3. Redaction tools
    - A. Adobe Acrobat Pro (most popular)
        - i. 7-day free trial available for those that need one-time use
    - B. Redact assist add-in (MS Word)
        - ii. Free trial available; recommend disabling add-in when not in use (affects system performance)
    - C. See section Redaction Software for more information about these and other tools
4. Redaction best practices
    - A. Use specific redaction tools that remove the text and all associated metadata.
    - B. Keep copies of the redacted and non-redacted versions
        - i. If redactions as submitted are not acceptable to ClinicalTrials.gov, a new redacted version will be required.
5. Methods that may not successfully redact
    - A. Printing a document, blacking out the information with a marker, then scanning the document
        - i. May not fully conceal the text underneath
    - B. Electronically putting a filled in shape, or using black highlight over text to be redacted leaves the text underneath intact.
        - i. Text can still be recovered even if the document is converted to a PDF.
    - C. Deleting text or changing the color of the text to a different color (like white)
        - i. Deleted text can be retained in the metadata and white text can be read by highlighting the text or copying and pasting it.
6. Information that can be redacted
    - A. Names, addresses, other Personally Identifiable Information (PII)

i. PII should always be redacted unless already disclosed (e.g. PI's name) or appropriate consent is obtained
ii. Studies of rare diseases may be more likely to contain PHI
iii. See section Redaction Practices: Personally Identifiable Information (PII)/ Protected Health Information (PHI) for more information

B. Trade Secrets and/or confidential commercial information
i. As defined in freedom of information Act (5 U.S.C. 552) and the Trade Secrets Act (18 U.S.C. 1905)
ii. Exploratory endpoints may be redactable since they are not required on ClinicalTrials.gov and may contain confidential commercial information
iii. Input from industry partner may be needed
a. Get input from contracts department
b. Understand privacy requirements
iv. See section Redaction Practices: Trade Secrets/ Confidential Commercial Information Guidance* for more information

C. Manufacturer details

7. What should NOT be redacted
A. Information that must otherwise be submitted as part of the study record, according to FDAAA 801/ Final Rule Regulations
B. Information that has already been disclosed publicly

8. Removing Metadata
A. Metadata gives information about other data in the document.
B. Even if the text in a document is deleted, it might still be retained in the document's metadata.
C. Metadata can reveal confidential information including: who reviewed the document and when; added or deleted text; imbedded images; comments; whether other versions of the document exist; etc.
D. Remove metadata prior to saving the document as a PDF/A
i. Please refer to Appendix 1 and Appendix 2 for steps to remove metadata using MS Word and Adobe Acrobat Pro. Other methods of removing metadata have not been evaluated in this report.

9. Institutional preparation
A. Identify resources to provide the needed software and support
B. Identify/establish infrastructure for document retention
C. Define roles and responsibilities
D. Create standard operating procedures (SOP) and/or guidance documents
E. See section Administrative Considerations

Required Documents

Requirements are defined in 42 CFR 11.48(a)(5): https://www.federalregister.gov/d/2016-22129/p-1542.

Comments, responses, and details on the Final Rules are found in 42 CFR 11.10, pages 64999-65002.

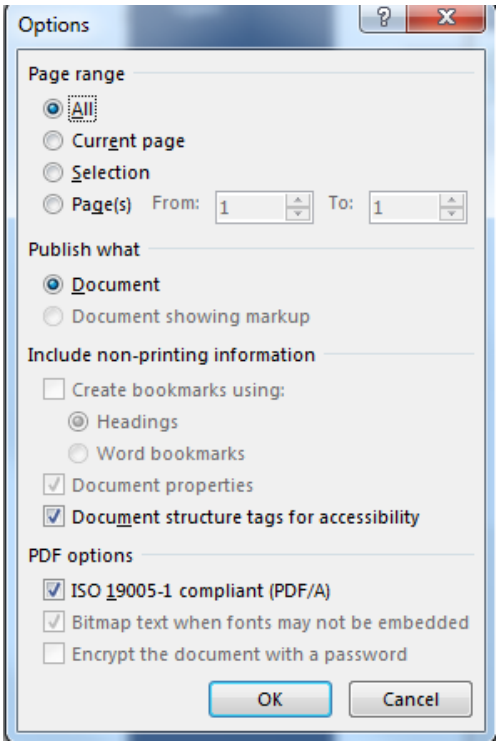
Full versions of the following documents are required (along with appropriate redactions):

1. Study Protocol
    - **A.** Defined in ClinicalTrials.gov regulations as "the written description of the clinical trial, including objective(s), design, and methods. It may also include relevant scientific background and statistical considerations." (42 CFR 11.10, page 65140)
2. Statistical Analysis Plan (SAP)
    - **A.** The SAP can be a stand-alone document
    - **B.** The SAP can be contained in the protocol, e.g., in Statistical Methods section
    - **C.** The ClinicalTrials.gov regulations do not define SAP. FDA Guidance for Industry E9 defines SAP as "a document that contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary variables and other data." (E9 Statistical Principles for Clinical Trials)
3. All amendments approved by a Human Subjects Review Board (applies to all locations for multi-site studies)
    - **A.** The final amended version that incorporates the original protocol and all amendments can be submitted
        - **i.** The version that is in effect at the data cut date for the results submission
        - **ii.** Submission of multiple documents (I.e., original version, previous amendments) is not required
        - **iii.** A clean version is sufficient, full track changes version is not required
    - **B.** "All amendments" is defined in ClinicalTrials.gov regulations as amendments "that have been approved by a human subjects protection review board (if applicable) before the time of submission that apply to all clinical trial Facility Locations." (42 CFR 11.48, page 65150)
    - **C.** Best practice: upload Protocol/SAP/Amendments at time of results submission to avoid multiple updates
        - **i.** See Appendix 3 for instructions in cases where multiple documents of the same type must be uploaded to the ClinicalTrials.gov system
4. NOT REQUIRED: Informed consent forms (ICF)

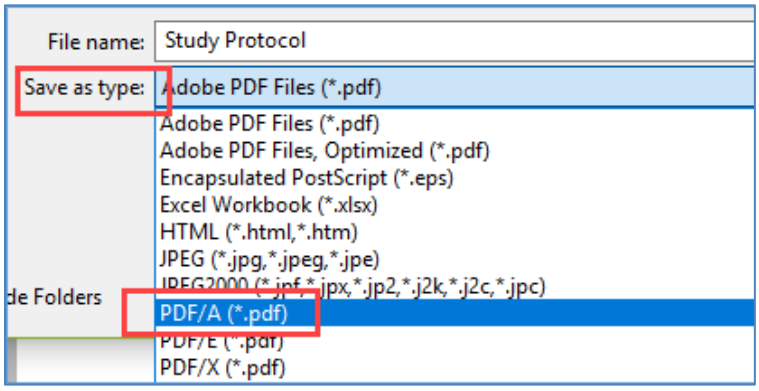
   A. Additionally, there is an option to provide copies of the blank informed consent forms (ICF), but it is not a requirement to do so. (42 CFR 11, preamble, pages 65001-2)
   B. There is a requirement for posting blank ICFs "on a publicly available Federal website" in the new Common Rule (6 CFR 46.116(h)) (scheduled to be effective on January 19, 2018 but now delayed until early 2019); this option is intended to provide a mechanism to fulfill that requirement.
5. Availability of Study Documents
   A. Assess for each study
      i. Are the required documents available?
      ii. If available, do they contain the necessary information?
      iii. If not available, organizations will need to contact ClinicalTrials.gov

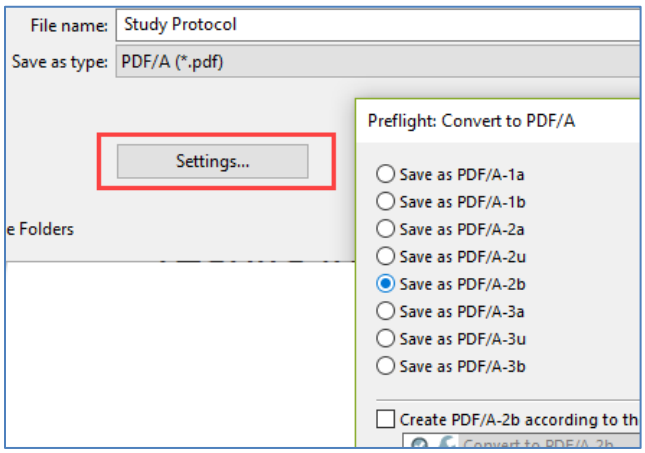
Document Formatting

1. Information about PDF/A format
   A. PDF/A is an ISO-standardized version of the Portable Document Format (PDF) (ISO 19005) specialized for use in the archiving and long-term preservation of electronic documents that has been adopted by federal agencies. PDF/A differs from PDF by prohibiting features ill-suited to long-term archiving, such as font linking (as opposed to font embedding) and encryption.
   B. PDF/A subtypes include PDF/A-1, PDF/A-2, and PDF/A-3
   C. ClinicalTrials.gov will accept any version of PDF/A.
   D. They also encourage that the file be consistent with the PDF Universal Accessibility (PDF/UA), but this is not required.
   E. PDF/A documents can be produced using the 'Save As' functions in Adobe Acrobat Professional or Microsoft Word. Other methods can be used (e.g., Acrobat Distiller, other Acrobat pro alternatives)
2. Each document must include a cover page with official title, NCT number (if available), and version date of the document
   A. The document date is the date on which the uploaded document was most recently updated and, if needed, approved by a human subjects protection review board.
3. Study documents must be included with all results submissions, both mandatory and voluntary
4. When using the "Save as PDF" function in Word, the format is regular PDF by default. To save as PDF/A, save as PDF, click "Options," and check the "ISO 19005-1 compliant (PDF/A)" box.



5. To save as PDF/A in Acrobat PRO, in the "Save as document type" drop-down menu, select PDF/A. Click settings to choose the type of PDF/A, but default (PDF/A-2b) should be fine.





Document Retention

1. Redaction is 2-step: first, text is marked for redaction (proposed) then the redactions are applied. Once applied the redactions cannot be removed.
2. Retain all available versions; clean, proposed redactions (unapplied), and redacted PDF/A version.
3. Institutions should plan for long-term storage of multiple versions of the documents in the event that the redactions are not acceptable to ClinicalTrials.gov.
    a. ClinicalTrials.gov "may contact a responsible party if it appears that the responsible party has redacted information that is otherwise required to be submitted under these regulations." (42 CFR 11, page 65001)
    b. In any redaction system, once the redactions are applied they cannot be removed, so users will need to return to clean versions, or versions with redactions proposed but not applied, if new versions are required.
4. Allowable Redactions per the Regulations (42 CFR 11, §11.48(a)(5), p. 65150)
    a. Names, addresses, and other personally identifiable information
    b. Trade secret and/or confidential commercial information (e.g., exploratory endpoints)
    c. Any information that is "otherwise required to be submitted" under the regulations may NOT be redacted.

Redaction Software

Generally, protocol files will be either MS Word files or PDF files. With both file types, redaction can be performed and the *Save As* function can be used to create a PDF/A version. Acrobat Pro or a similar alternative product is needed to save a PDF as a PDF/A.

The main features needed in a redaction tool are the ability to *select and redact* text, and *find and redact* text. The tool should also be able to remove metadata in a similar fashion to the methods described in Appendix 1 and Appendix 2. In general, unless a user has specific redaction software, that user does not have the tools necessary to perform redaction.

Legal departments at institutions may have software deployed for redaction, so it may be useful to inquire whether their solutions might somehow be extended or adapted to fit this need.

1. **Microsoft Word**
    a. MS Word add-ins are available for purchase (Word does not have an out of the box redaction feature)

i. Redact Assistant, $4.99 per seat when purchased from store.office.com, or when purchased as an add-in using the Word Online platform, provides the necessary redaction tools.
ii. Free 30-day trial available
iii. The add-in slows the performance of the application, but it can be disabled and re-enabled easily; when disabled it doesn't impact performance.
iv. Do not overwrite the unredacted version of the Word document after running Redact Assistant; once the file is saved, the redactions cannot be removed.
v. Retain 2 versions: the clean Word file and the redacted PDF/A version.

b. Foxit Redactor for Office ($39.95) – Plug in that can redact documents in Microsoft Office and publish them as PDFs. PDFs can then be saved as PDF/A format.
c. There may be other redaction solutions utilizing MS Word.

2. **Adobe Acrobat Pro** ($15/Month, price varies)
   a. Adobe Acrobat is available in three versions: Reader, Standard, and Pro. Of the different versions of Adobe Acrobat (Reader, Standard, and Pro), only Acrobat Pro has the necessary PDF redaction tools. It is an ideal program if available, but the price per seat is somewhat expensive (pricing varies).
3. **Alternatives to Acrobat Pro** (This is not a complete list of alternatives)
   a. NitroPDF ($159 per license) – Similar ability to Acrobat pro to redact and convert PDFs to PDF/A but with a cheaper price tag. Discounts for buying multiple licenses
   b. Nuance Power PDF Advanced ($149 per license): Standard version does not have reaction feature but the 'Advanced' version does. Another alternative to Acrobat Pro.
   c. Foxit PhantomPDF Business ($139 or $9/month): Create and edit PDFs including PDA/A and redacted documents. PhantomPDF standard does not include the ability to convert to PDF/A.
   d. pdfelement (starting at $99 per seat).
   e. iSkysoft (starting at $99 per seat).
4. **Other PDF Redaction Tools**
   a. Third-party vendors offer redaction software that can be expensive, but provide added benefits. They are good for large projects (~250k pp). These solutions may allow for use of customized scripts, and they have many advanced features. Users will not be familiar with it as an interface, so training would be needed and likely only a fraction of the software functionality will be utilized.

Redaction Practices: Personally Identifiable Information (PII) / Protected Health Information (PHI)

While protocols generally do not contain PII or PHI aside from things like study team members' names and contact information, it is important to ensure that this information is not accidentally disclosed without an individual's consent.

In the appendix of OMB M-10-23 (Guidance for Agency Use of Third-Party Website and Applications) the definition of PII was updated to include the following:

> Personally Identifiable Information (PII). The term "PII," as defined in OMB Memorandum M-07-1616 refers to information that can be used to distinguish or trace an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified. In performing this assessment, it is important for an agency to recognize that non-PII can become PII whenever additional information is made publicly available — in any medium and from any source — that, when combined with other available information, could be used to identify an individual

The U.S. Department of Health & Human Services (HHS) issued Guidance "Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule" that can be used to help identify what constitutes PII or PHI.

The 'Safe Harbor Method' is one of the most commonly used guidelines for identifying and removing PII/PHI. The safe harbor outlines 18 individual identifiers that could be used to identify a specific individual either alone or in conjunction. This method was developed for de-identifying data sets, but the same principles apply to a protocol.

In §164.514(b), the Safe Harbor method for de-identification is defined as follows:

The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of

the ZIP code if, according to the current publicly available data from the Bureau of the Census:

   a. The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and
   b. The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000

3. All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
4. Telephone numbers
5. Fax numbers
6. Email addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) addresses
16. Biometric identifiers, including finger and voice prints
17. Full-face photographs and any comparable images
18. Any other unique identifying number, characteristic, or code, except for as detailed below

    a. *Implementation specifications*: reidentification. A covered entity may assign a code or other means of record identification to allow information deidentified under this section to be reidentified by the covered entity, provided that:
        i. *Derivation*. The code or other means of record identification is not derived from or related to information about the individual and is not otherwise capable of being translated so as to identify the individual; and
        ii. *Security.* The covered entity does not use or disclose the code or other means of record identification for any other purpose, and does not disclose the mechanism for re-identification.

Redaction Practices: Trade Secrets/ Confidential Commercial Information Guidance*

General guidance on what is generally considered to be disclosable under freedom of information act and what may be exempt: based on the FDA's 2008 "Guidance for Industry Advisory Committee Meetings — Preparation and Public Availability of Information Given to Advisory Committee Members"

1. **Information in Briefing Materials That Typically Will Be Disclosable Under FOIA** We generally will consider the following information in advisory committee briefing materials to be disclosable without redaction, unless the sponsor demonstrates that disclosure of the information is likely to cause substantial competitive harm:
   - Summaries of clinical safety and effectiveness data;
   - Summaries of non-clinical safety and effectiveness data;
   - Summaries of adverse drug reaction data;
   - Written discussion or analysis of safety or effectiveness data relevant to the topic of the meeting;
   - A general description (such as that which would typically be included in product labeling) of product functions, mechanics, and/or engineering;
   - A general description of physical characteristics and performance parameters;
   - Clinical or preclinical protocols or summaries of protocols;
   - Statistical protocols and analyses;
   - Information that is proposed to be included in product labeling, such as indications and usage, dosage and administration, and safety information such as warnings and precautions;
   - Literature references;[1]
   - Any other information that has been previously publicly disclosed by the sponsor;
   - Copies of the sponsor's slides to be presented at the advisory committee meeting, if included in the briefing materials; and
   - Guidance documents.

The above list is neither exhaustive nor absolute.

2. **Information in Briefing Materials That Will Typically Be Exempt from Disclosure** We generally will consider the following types of information to be exempt from disclosure under FOIA:
   - Information about product functions, mechanics, engineering, and schematic drawings not in the proposed labeling and not within the scope of the agenda for the meeting;
   - Proprietary physical characteristics and performance parameters not in the proposed labeling and not within the scope of the agenda for the meeting;
   - Manufacturing process information;
   - Manufacturing quality control information;

- Clinical raw data; [2]
- Non-clinical raw data;
- Supplier names, customer lists, production costs, inventory information, failure rates of products, production quality control information;
- Information for which the release would constitute an unwarranted invasion of personal privacy; and
- Product formulation information not in the labeling.

The above list is neither exhaustive nor absolute.

[1] FDA does not post copyrighted materials on its website. If sponsors do wish to submit copyrighted materials, they should provide a bibliography of the copyrighted materials that can be posted.

[2] For the purposes of this guidance, FDA considers "raw data" to be a complete data set of case report forms, case report tabulations, or line listings. Data that summarize individual or multiple subject outcomes or results are considered summaries. Summaries may include examples of specific findings.

*****The above information is targeted specifically at advisory committee meetings and is not specific to ClinicalTrials.gov requirements. It is the responsibility of the responsible party and sponsor to ensure that only appropriate information is redacted or disclosed.

Administrative Considerations

1. Contact Contracts team if applicable
2. PI should review the protocol to make sure protected information is not disclosed (confidential commercial information; personally identifiable information; 42 CFR 11.48, page 65150)
3. When should protocols be uploaded?
    a. Define best practice
    b. If provided at the time of results reporting, less is revealed during the conduct of the study. That may be important to prevent replication by others while the study is still ongoing. Waiting to post the protocol until the time of results reporting could also reduce the administrative burden of having to repeatedly post updated versions of the protocol due to amendments.
4. What is your institution's stance on intellectual property (IP)? Who needs to be part of that conversation (tech transfer, legal)?
5. Who reviews redacted versions? - PI, IP legal, central office/PRS administrator, industry partner?
    a. Early phase studies more likely to contain confidential commercial information – by later phases much may already have been published.

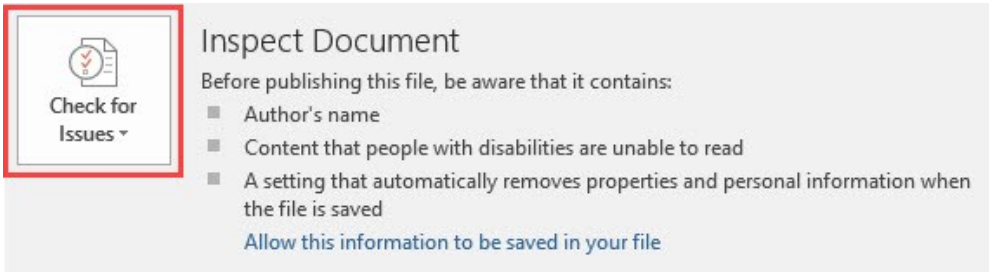
b. Is IP held at your institution? Determine if IP/legal review is needed.
c. Is IP held by industry collaborators? Determine if industry review is needed.

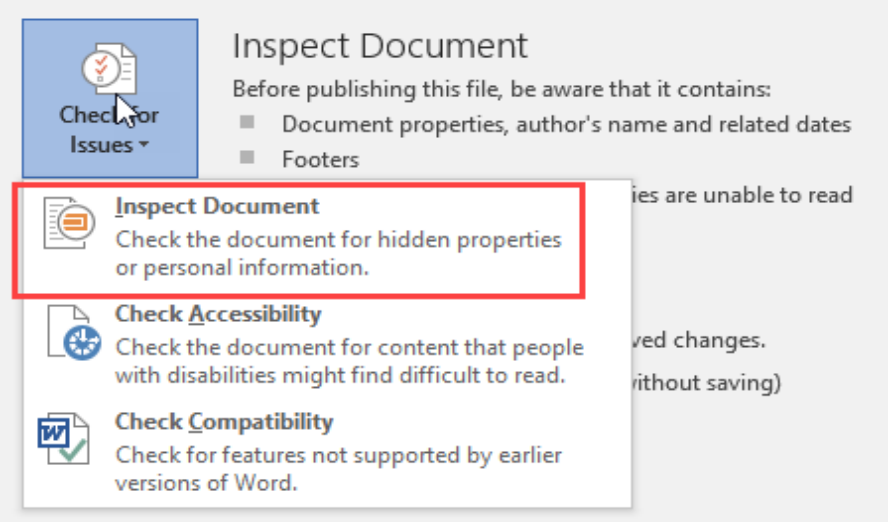
6. Who maintains original and redacted versions? Where will the electronic files be stored and for how long?
7. Does your institution need to invest in redaction software?
8. Plan for education on the new requirements – get the word out
    a. Do you need to provide training on common redaction errors? (For example, just deleting/blacking out info doesn't mean it can't be recovered)
    b. Are formal protocol documents required at your institution? If not, investigators need to know and plan for how to provide documents acceptable to ClinicalTrials.gov and the institution.
9. In the early implementation of this requirement, organizations may wish to monitor releases to determine whether study documents are included, and determine if adequate safeguards were considered in their release.

APPENDICES

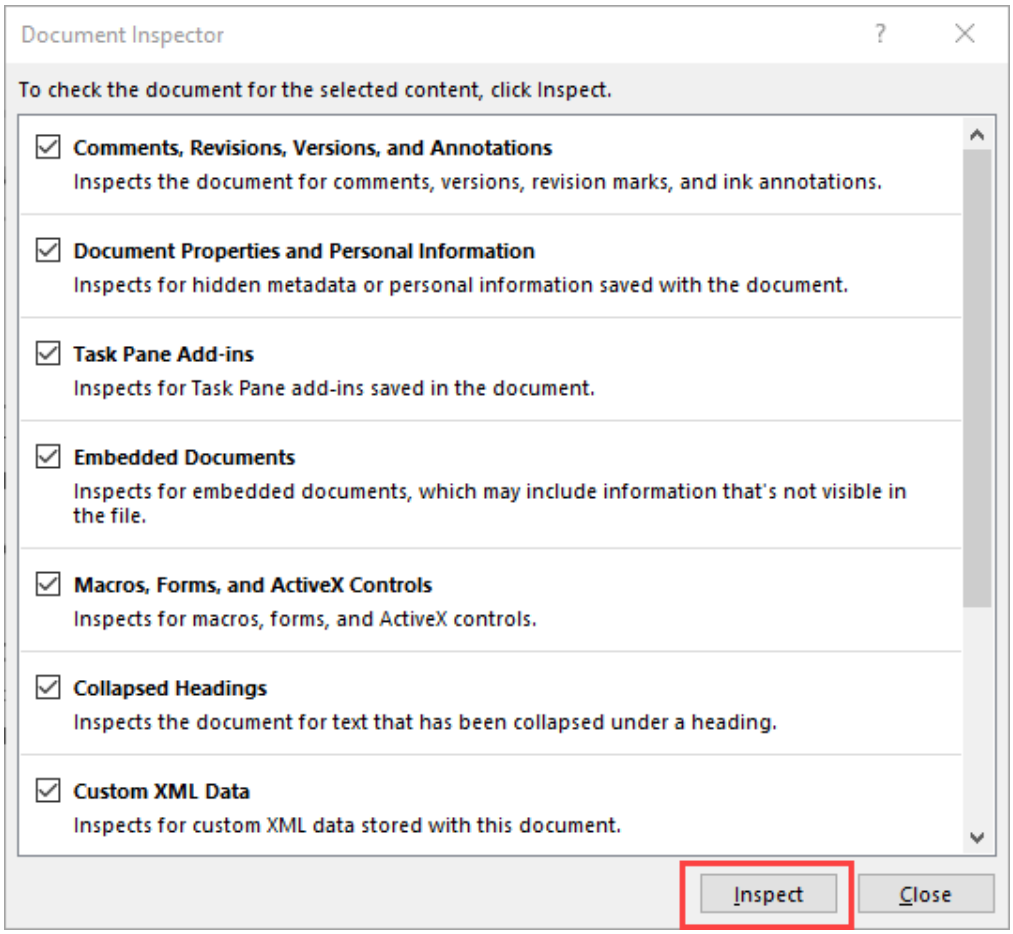
Appendix 1, Removing Metadata Using MS Word:

1. Go to File>Info and click 'Check for Issues', then 'Inspect Document'

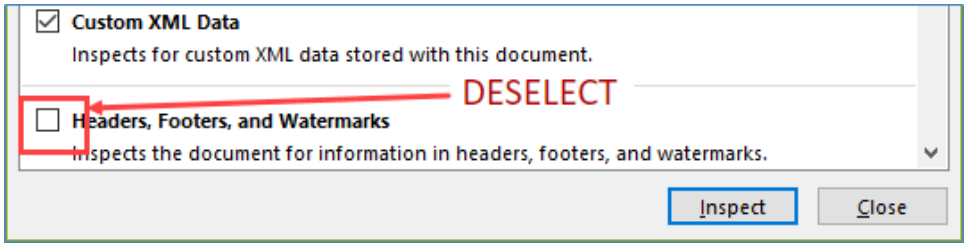




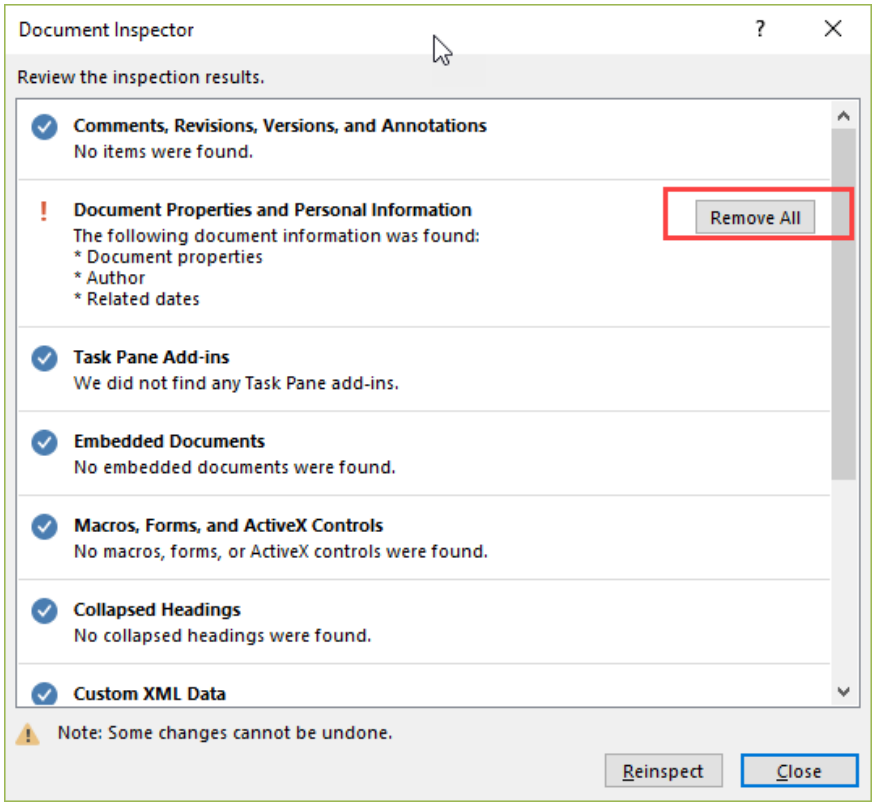
2. An interface window appears with the various items to be inspected. Leave all checked and click Inspect



Note: Do not remove headers and footers. These will come up on the inspection if they exist, however, it removes headers and footers completely, not just metadata.



3. Results are returned, click Remove All (document properties and personal information).

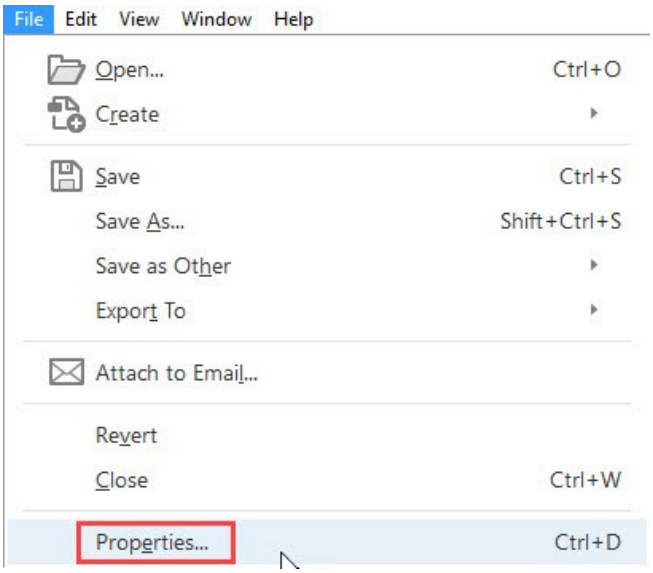


4. Close and Save

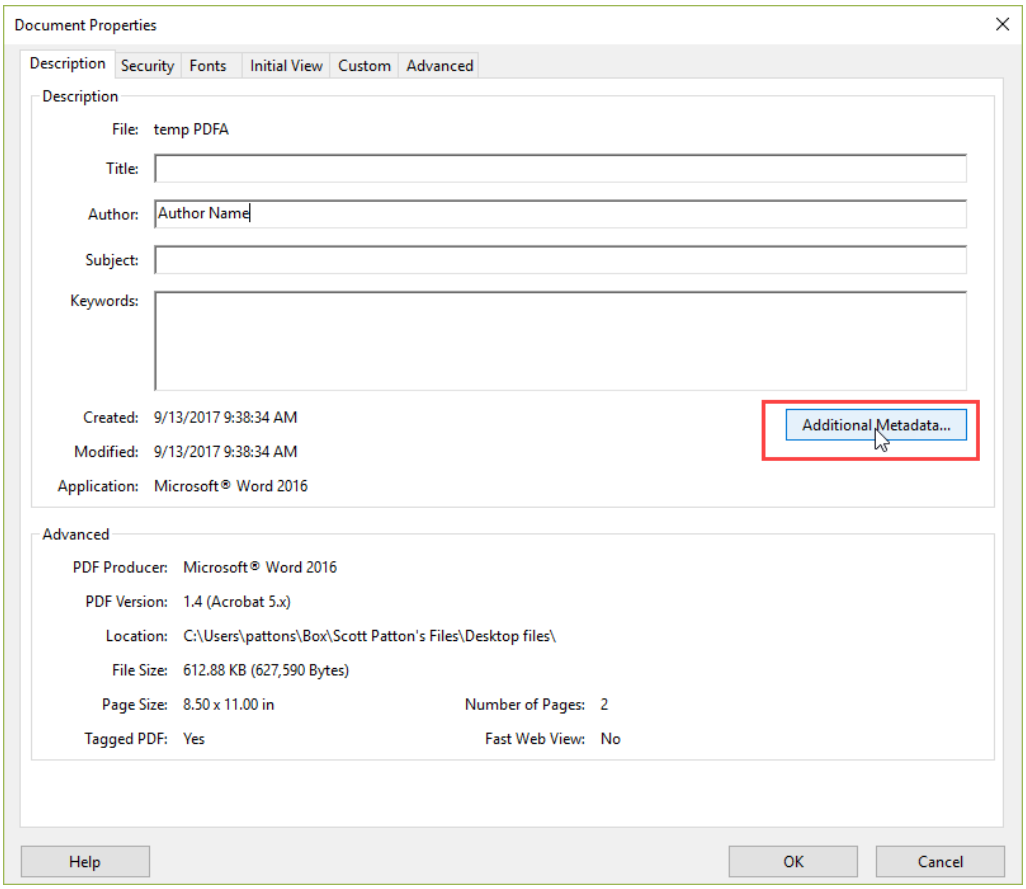
Appendix 2, Removing Metadata using Adobe Acrobat Pro

Note: metadata are not editable in Acrobat Reader

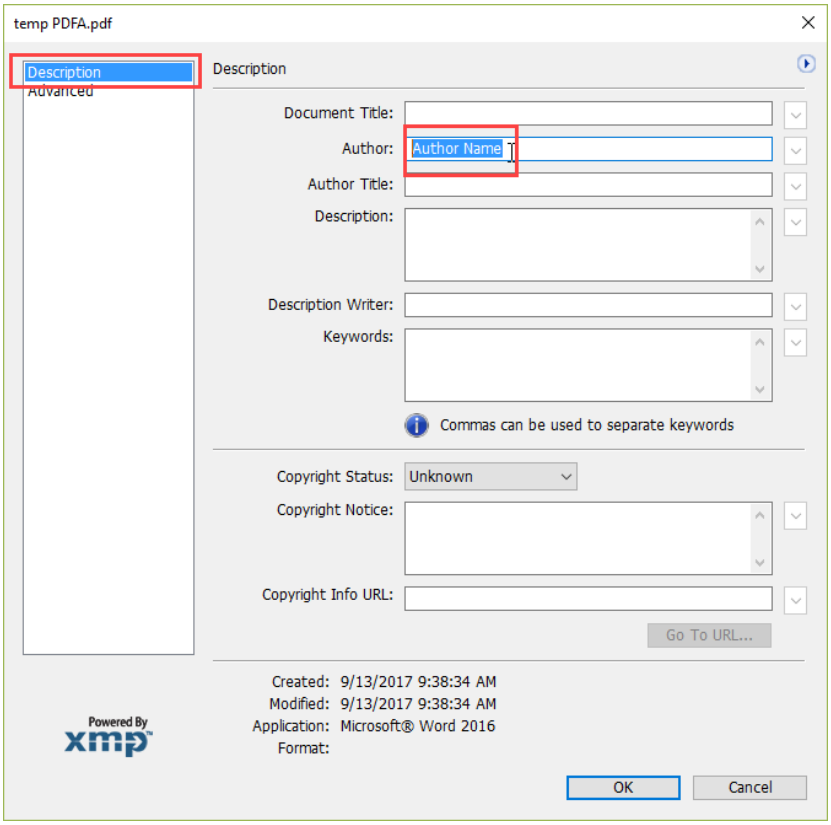
1. Go to File>Properties

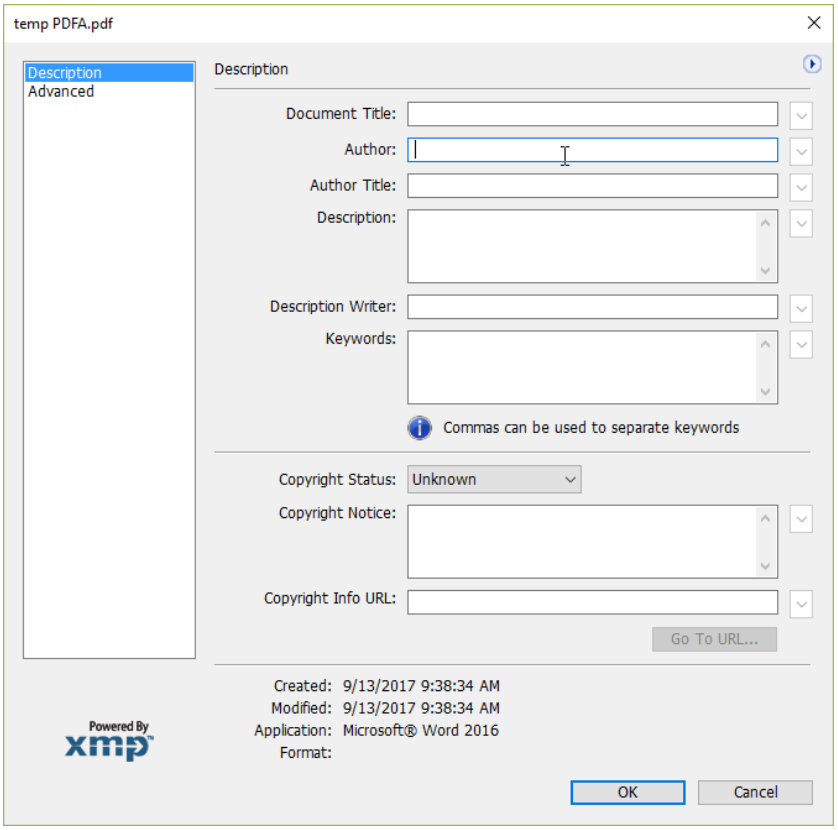


2. Click Additional Metadata

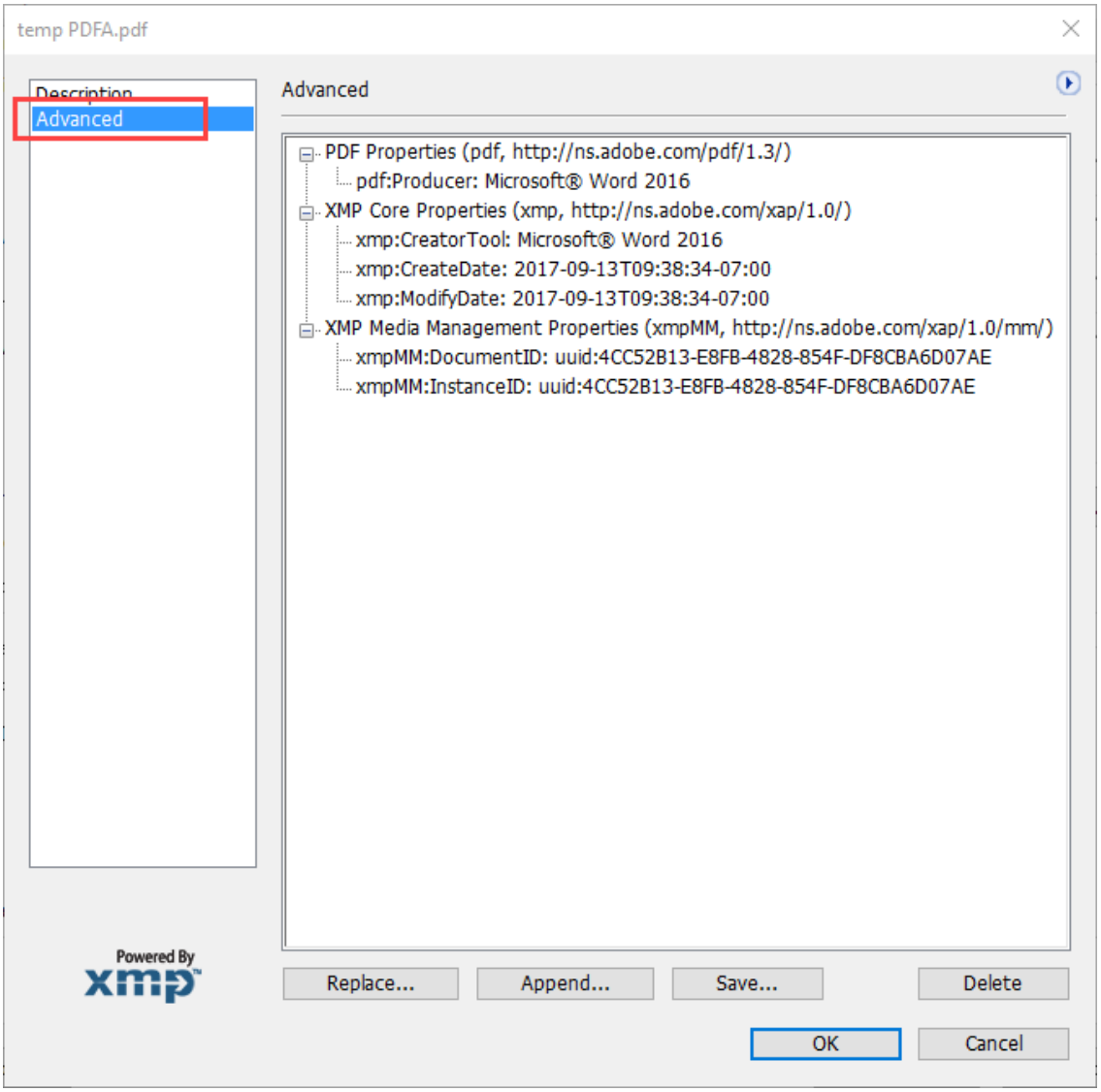


3. In the Description section, clear metadata that should be removed





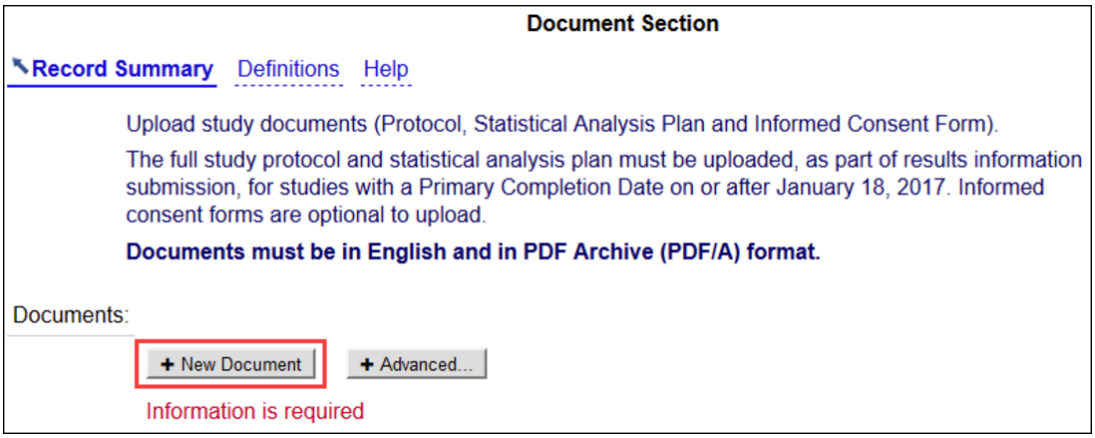
4. The Advanced section contains file properties that should be left as-is



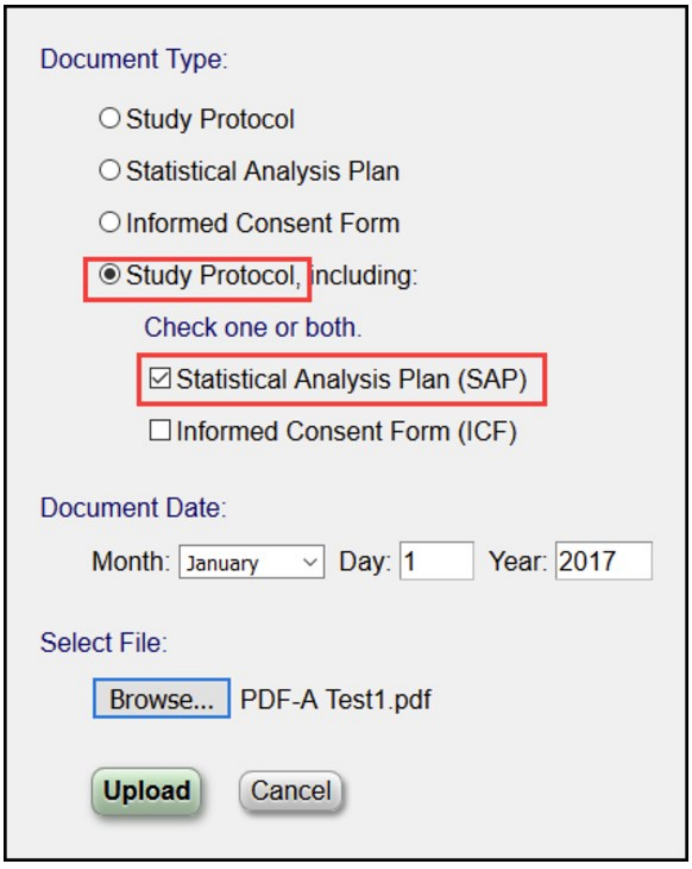
5. Click OK and Save

Appendix 3, Uploading Multiple Documents of the Same Type on ClinicalTrials.gov:

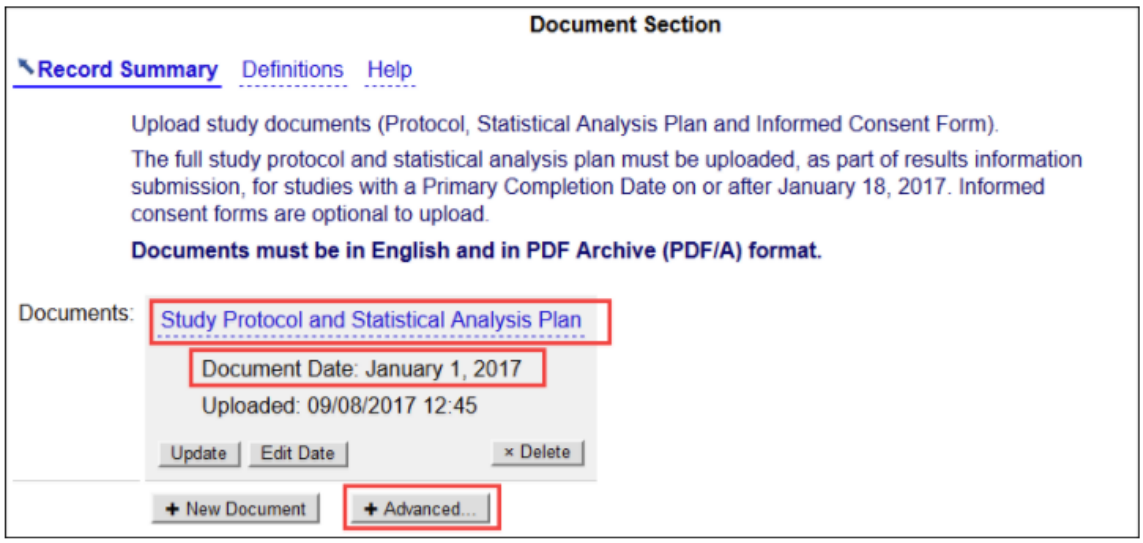
1. When the user enters the first document in the *Study Document* section, the *New Document* link is used:



2. Enter the document details, in this example, the protocol with statistical analysis plan (SAP), document date, browse for the file, and upload

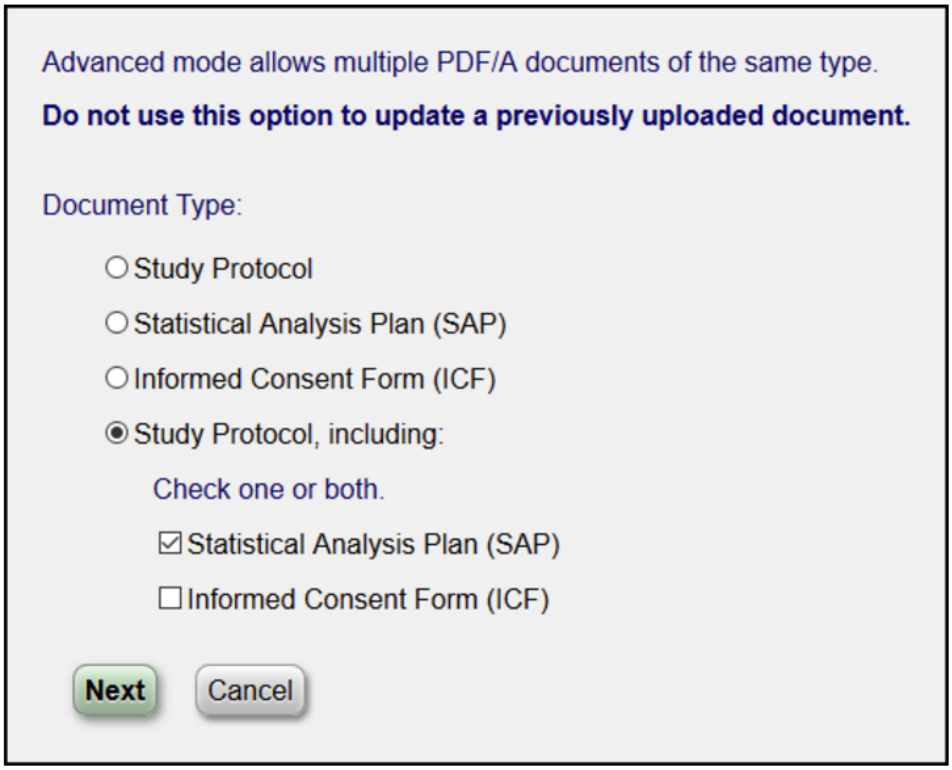


3. Now in the *Study Document* section, the file details and a link to view the uploaded file are visible, as well as several links

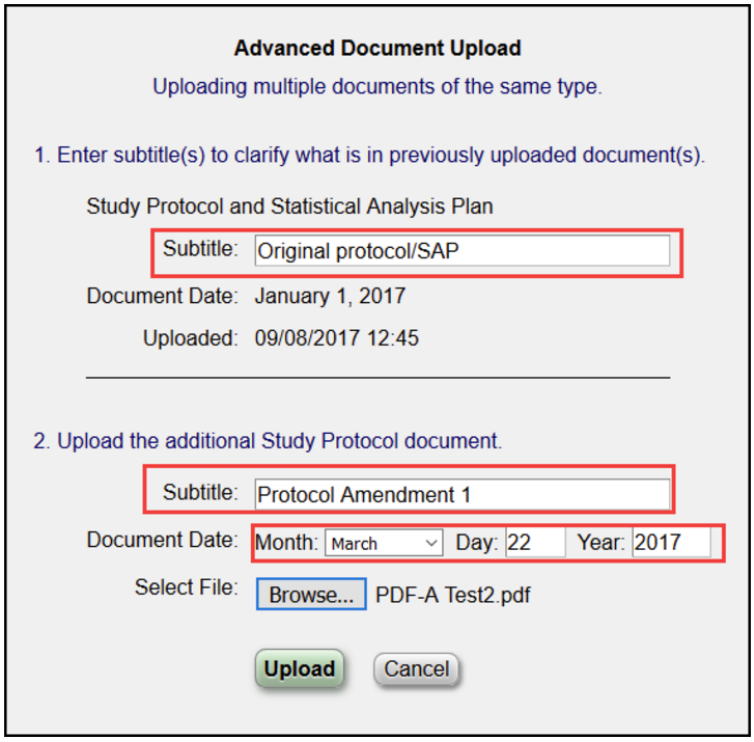


4. Click *Advanced* to upload a new document that is the same *Document Type*, in this example, the original protocol and the protocol amendment to be uploaded are the same *Document Type*

5. Enter the document details, in this case a protocol/SAP amendment



6. Click *Next*, then name the first document, the new document with details, browse for the file, and upload



Now in the *Study Document* section, the file details and links to view the uploaded files are visible