

Hands-on Python PDFs: Using the pypdf Library To Programmatically Design, Complete, Read, and Extract Data from PDF Forms Having Digital Signatures

Troy Martin Hughes

ABSTRACT

The pypdf library (<https://pypdf.readthedocs.io/en/stable/index.html>) facilitates the programmatic creation, editing, completion, and extraction of PDF forms. Using only pypdf, developers can design a PDF form, make the form available to end users, and programmatically collect user data. One notable limitation of pypdf is its lack of support for PDFs that require one or more digital signature blocks (<https://pypdf.readthedocs.io/en/stable/meta/scope-of-pypdf.html>), and similar Python libraries such as pyhanko acknowledge they are “in beta, and not yet production-ready.” (<https://pypi.org/project/pyHanko/>). This hands-on workshop introduces users to the powerful pypdf library, and demonstrates a solution that does enable PDF forms to be generated (and populated) dynamically—including forms that contain digital signature blocks. The real-life business case examines the permission-to-publish (aka copyright-release) forms that are regularly required to attend conferences like PharmaSUG. Attendees will learn how to build these forms dynamically using a data-driven software design approach. All Python code will be shared so that attendees can implement similar solutions within their respective organizations.

INTRODUCTION

This hands-on workshop introduces users to the powerful Python pypdf library, and demonstrates two use cases that dramatically increase the efficiency and accuracy of processing user data through PDF forms. These real-life use cases presage advancements showcased during the upcoming WUSS 2024 conference, with the author—the WUSS conference Academic Chair—demonstrating how he leveraged Python to improve a SAS conference.

The first use case overcomes the common practice of collecting user data haphazardly, such as through nonstructured formats such as email, or structured formats such as PDF forms—but without taking advantage of automatic text extraction therefrom. New at WUSS 2024, for the first time ever, Section Chairs were required to complete a PDF that solicited their personal information, contact information, and historical conference attendance. This HOW introduces the form, its underlying metadata, and how to extract these data using pypdf. Derivative data products, all of which are programmatically generated, include an Excel workbook and an HTML webpage that features the names of the WUSS Academic Team.

The second use case overcomes the unpopular practice of copyright release forms (aka permission to publish) that presenters are required to complete at conferences like PharmaSUG and WUSS. These forms require the presenter to type the presentation ID number, the presentation title, the author’s name, job title, and company or organization, and to digitally sign and date the PDF. Ridiculously, all of these metadata are already collected by conference organizers, so this error-prone, manual process is entirely unnecessary. Thus, this HOW demonstrates a programmatically generated copyright release form that is debuting at WUSS 2024, prepopulated by the conference, and securely posted online. Rather than presenters having to needlessly enter data, they now only need to download a specific form for a presentation, digitally sign it, and return it to the conference.

Throughout these two use cases are interwoven overarching best practices in software design and data management that WUSS is embracing to make processes more efficient and accurate. Master data management (MDM) requires that master data (such as a Section Chair’s name or email address) are entered once, maintained in one table, and used to feed derivative tables and data products. Thus, MDM overcomes data integrity issues such as when a user’s name is spelled correctly in one file yet incorrectly in another. Data-driven design is a second tenet that WUSS has embraced by using these tables to build processes that ingest control data, and interpret those data to produce data products. Finally, automation is a third focus, in which these processes—including both data extraction and data writing—are handled programmatically rather than through manual processes.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Troy Martin Hughes

E-mail: troymartinhughes@gmail.com



SEIZE THE DATA
SACRAMENTO, CA

WUSS

WESTERN USERS OF SAS SOFTWARE

SEPTEMBER 4TH - 6TH 2024 | WWW.WUSS.ORG

WUSS 2024 Copyright Release Form

Dear Author(s),

Congratulations once again for being accepted into the WUSS 2024 conference proceedings, held at the Hyatt Regency Sacramento, September 4 to 6, 2024. We look forward to seeing your presentation in person!

This Copyright Release Form (CRF) applies **only** to the following paper/presentation, hereafter known as “**the Work**”:

Presentation ID:

White Paper:

Slide Presentation:

Title:

Author(s):

Note that the preceding title must match the Work title that is submitted to WUSS for publication, and that the preceding author name(s) (and order thereof) must match the author name(s) that appear in the Work. If you detect a discrepancy in the presentation track (white paper vs. slide presentation), in the title, or in the author(s), please immediately notify your Section Chair, who can modify these data in the WUSS Submission System.

Your presentation will not be scheduled until and unless this CRF is completed, and any delay in submission of this PDF could result in your presentation being removed from the conference proceedings. Only one CRF should be returned per presentation, so all author(s) must digitally sign the same form, after which the completed CRF should be uploaded to the WUSS Submission System (<https://softconf.com/n/wuss2024>) no later than

Adobe Acrobat Reader is required to digitally sign this PDF, and free software can be obtained here: <https://get.adobe.com/reader/>. If you do not already possess a digital signature, you can create a “self-signed digital ID” for free. Instructions for creating your digital ID are here: <https://helpx.adobe.com/acrobat/using/digital-ids.html>. **Only digitally signed documents will be accepted by WUSS.**

Section A.

WUSS is the **Publisher** of this Work.

The **Copyright Owner** of this Work is:

The undersigned, being all of the authors of the above captioned Work, hereby grant(s) a perpetual, worldwide, royalty-free, nonexclusive right under the copyright in the Work to the Publisher to publish the Work in the WUSS 2024 Conference Proceedings and to take the Work, as published, and reproduce it in sales and marketing literature produced by the Publisher, effective if and when the Work is accepted for publication by the Publisher. A copyright notice in the name of the Copyright Owner is to accompany every reproduction of the Work if Publisher's publication or reprint does not include an overall copyright notice protecting the Work. This grant is subject to the reservation of the Copyright Owner's rights to (1) copyright ownership, including use or publication of all or any part of the contents of the Work in the course of their business or in any publication authored or published by or for WUSS or its designee; (2) reproduction and/or distribution of the Work but not the Publication as a whole as published by Publisher, and (3) approval of any changes in the Work by the Publisher. No proprietary rights other than this grant under copyright are included, and no transfer of any rights in the Work may be made by Publisher to a third party without the prior written approval of the Copyright Owner.

First Author

By signing, the author acknowledges that no further permissions are required by the author's company or organization.

Author Name:	Authorized Representative Name:
Job Title:	Job Title:
Company/Organization:	Company/Organization:

Second Author

By signing, the author acknowledges that no further permissions are required by the author's company or organization.

Author Name:	Authorized Representative Name:
Job Title:	Job Title:
Company/Organization:	Company/Organization:



WUSS

WESTERN USERS OF SAS SOFTWARE

SEPTEMBER 4TH - 6TH 2024 | WWW.WUSS.ORG

WUSS 2024 Deputy Section Chair Application

WUSS invites you to apply to be a Deputy Section Chair! Deputies are invaluable volunteers who work in concert with a single Section Chair, as well as the WUSS Academic Team, and who help recruit, facilitate, and moderate conference content, including presentations and papers.

Deputy Section Chair benefits include:

- Half price registration (a \$337.50 discount) for the WUSS 2024 conference, held at the Hyatt Regency Sacramento, California, September 4 to 6, 2024.
- Three days of rousing presentations, workshops, and demonstrations on SAS software development, software design best practices, statistics and data analytics, data visualization, open-source programming, and so much more.
- Invitation to the volunteer appreciation dinner, in addition to other conference social events and activities.
- Friendship, fun, and future leadership opportunities in the WUSS family!

Deputies (as well as all volunteers and attendees) are required to read the WUSS Code of Conduct (<https://www.wuss.org/code-of-conduct/>), and are required to adhere to it both in communications preceding the conference and during the conference proceedings.

Deputies are required to maintain the following accounts, all of which are free:

- **Adobe Acrobat Reader** is required to view and digitally sign PDF documents (including this one). Free software can be obtained here: <https://get.adobe.com/reader/>. If you do not already possess a digital signature, you can create a “self-signed digital ID” for free. Instructions for creating your digital ID are here: <https://helpx.adobe.com/acrobat/using/digital-ids.html>. Only digitally signed documents will be accepted!
- **SoftConf/START** is the application in which the conference proceedings (e.g., papers, presentations, abstracts, speaker biographies) are maintained and in which Section Chairs and Deputies will review and update speaker submissions. A free account can be obtained by selecting “new user”: <https://softconf.com/n/wuss2024/>. WUSS will provide SoftConf training to all Section Chairs and Deputies.
- **Gmail** is *recommended* to access and edit documents on the WUSS Google Drive, and to communicate with prospective and accepted speakers; however, any email account is acceptable, as long as it can be distributed to other volunteers and speakers. A free Gmail account can be obtained here: <https://support.google.com/mail/answer/56256?hl=en>. Note that only one email account should be used for WUSS business, and this must be registered with SoftConf and registered for WUSS Google Drive access.


Pre-conference responsibilities include:

- Attend four 60- to 90-minute WUSS Academic Team virtual meetings between April and July, scheduled approximately every four weeks.
- Receive mentoring from one assigned Section Chair and the WUSS Academic Chair, and backup your Section Chair as needed.
- Help recruit authors and speakers to increase conference content and attendees, and recommend (and prioritize) proposed content for inclusion in WUSS conference proceedings.
- Help review papers and presentations for content, clarity, and adherence to all WUSS publication guidelines.
- Provide proposal feedback to authors and speakers, and work with them to ensure final papers and presentations incorporate all required modifications and are uploaded on time.

Conference responsibilities include:

- Attend the WUSS conference from September 4 to 6, 2024, in Sacramento, California.
- Help prepare classroom, including: loading presentations on WUSS laptop, verifying A/V equipment functionality, and interfacing with facilities and A/V personnel to remedy issues, as necessary.
- Help facilitate and monitor section presentations for approximately eight hours, including: introducing speakers, moderating audience questions, and assisting presenters, as necessary.
- Have fun, visit with and make friends, and enjoy the phenomenal content presented at WUSS!

First Name:	Middle Name (optional):	Last Name:
Company or Organization:		Job Title:
City of Residence:	State:	ZIP Code:



Email Address:	SoftConf/START Username:
Biography (five to six sentences describing your interest in volunteering and your SAS or other experience):	

Your phone number will be shared among WUSS leadership, and will be used sparingly to contact you only 1) at the conference to facilitate real-time communication, and 2) prior to the conference if attempts to contact you via email have been unsuccessful. You are not expected to share your phone number with speakers, but may choose to do so.

Cell Phone Number:

The following information is shared only with the WUSS Executive Committee, and is used to evaluate the recruitment of volunteers as well as the experience level of the conference team.

Which WUSS conference years have you attended?

Which regional and/or national conferences have you attended?

How many times have you been a Section Chair at any conference?

At which conferences have you been a Section Chair?

The following statements acknowledge your participation and your understanding of the commitment:

I have read and agree to abide by the WUSS Code of Conduct (<https://www.wuss.org/code-of-conduct/>).

I have read the foregoing Deputy Section Chair role, and accept the pre-conference and conference responsibilities as outlined; I understand that the discounted WUSS registration is conditional upon fulfilling these responsibilities.

If at any point I realize that I cannot complete my conference or pre-conference responsibilities, or that I cannot attend the WUSS 2024 conference, I will immediately notify the Academic Chair so a replacement can be selected.

Digitally sign this document using your Adobe digital ID, and save the locked form using your full name in the filename. Email this form to the Section Chair (or WUSS Academic Chair) recruiting you:

If you would like your photo displayed publicly on the WUSS website (Academic Leadership page), please email your 432w x 508h JPG image to the WUSS Academic Chair: academic@wuss.org.

The **Section Chair (or WUSS Academic Chair)** recruiting the Deputy applicant must digitally sign to acknowledge review of the applicant's qualifications and of this document, and to endorse the applicant's participation in WUSS leadership as a Deputy Section Chair:

The **WUSS Academic Chair** must digitally sign to acknowledge review of this document of the applicant's qualifications, and to endorse the applicant's participation in WUSS leadership as a Deputy Section Chair.