



# There's a Snake in My Group:

Exploring Python's Capabilities

Kevin Miller,  
Gilead Sciences

Flat is better than nested. Sparse is better than dense. Readability counts. Special cases aren't special enough to break the rules. Although practicality beats purity.



# There's a Snake in My Group:

Exploring Python's Capabilities

Kevin Miller,  
Gilead Sciences

Flat is better than nested. Sparse is better than dense. Readability counts. Special cases aren't special enough to break the rules. Although practicality beats purity.

# AGENDA

Why Python?



SAS, Meet Python



Projects Leveraging Python



Thoughts and Considerations





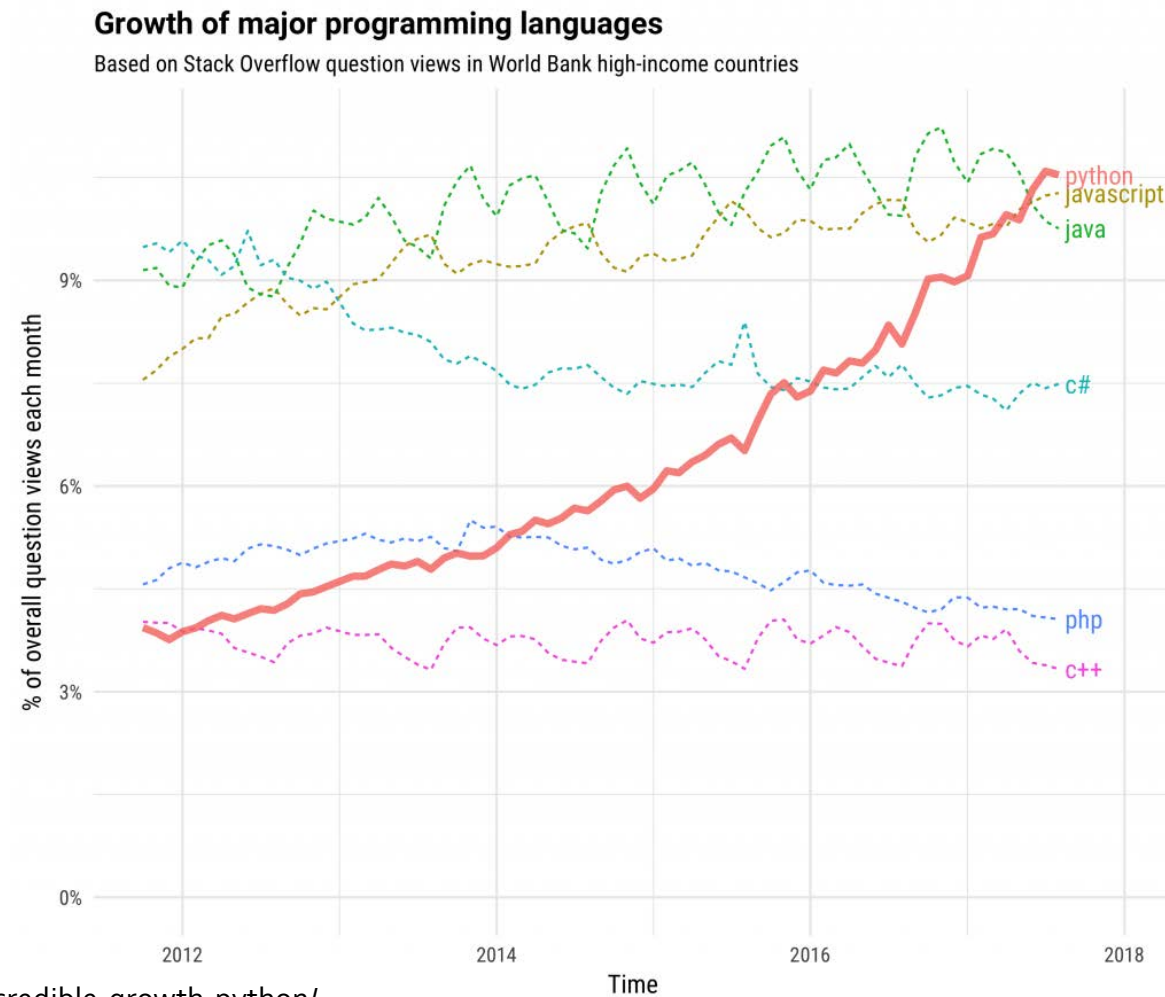
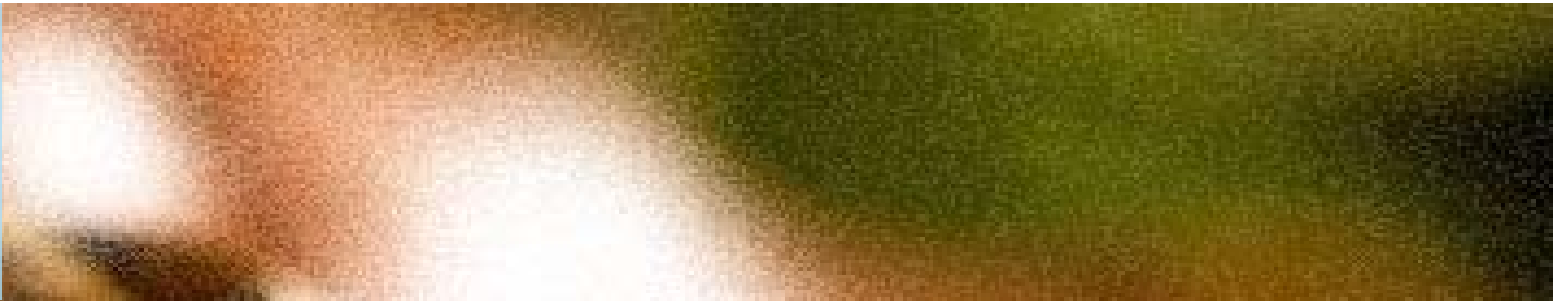
Why Python?



# Why Python

- Increasingly Complex Tasks
  - Increased ownership outside of traditional programming role
- Current Method Has Drawbacks
  - Multiple tools to accomplish task
- Need for Automation
  - API Interaction
  - XML Modification and Creation

# Why Python



Source: <https://stackoverflow.blog/2017/09/06/incredible-growth-python/>

# Benefits of Using Python

- Python has a very simple and consistent syntax
  - "There should be one—and preferably only one—obvious way to do it"
- Lots of resources:
  - 5th Largest StackOverflow Community
  - 4th Most-Used Language at GitHub
  - 125,000+ 3<sup>rd</sup> Party Libraries
  - Built in Libraries
- Forces good programming practices
- Its free!\*

# Considerations of Using Python

- Our organization relies heavily on SAS proprietary data formats
  - From export of edc to final analysis
- Python can be considered slow
  - Large memory consumption
- Python programs must be consistent



SAS, Meet Python



# SAS, Meet Python

- Replicating a utility in Python
  - Good intro to new language
  - Compare results
  - Familiar subject matter
- Python Modules Used
  - argparse
  - pandas
  - xlsxwriter



# Methodology: Comparing Excel Files



Import as Dataframes, use Pandas library methods to evaluate differences, and output excel file of results

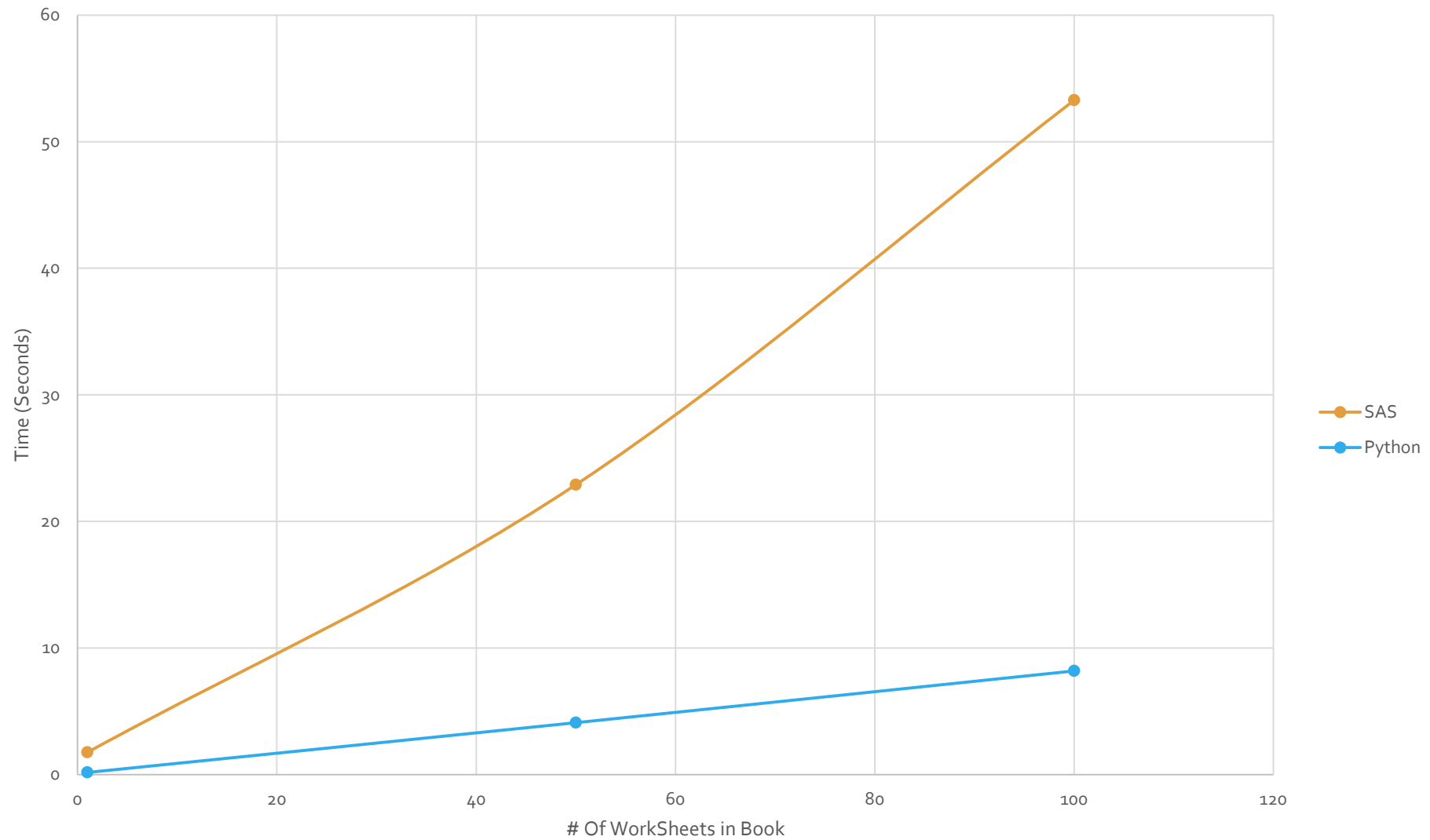


Import as SAS Datasets, use proc compare to evaluate differences, and output excel file of results

# Project Example: Comparing Excel Files

- Comparing Increasing # of Sheets
  - Team wanted a quick and easy method of comparing SDTM/ADaM Specification sheets across studies and tasks
  - Small but varied sheets with 30ish Variables

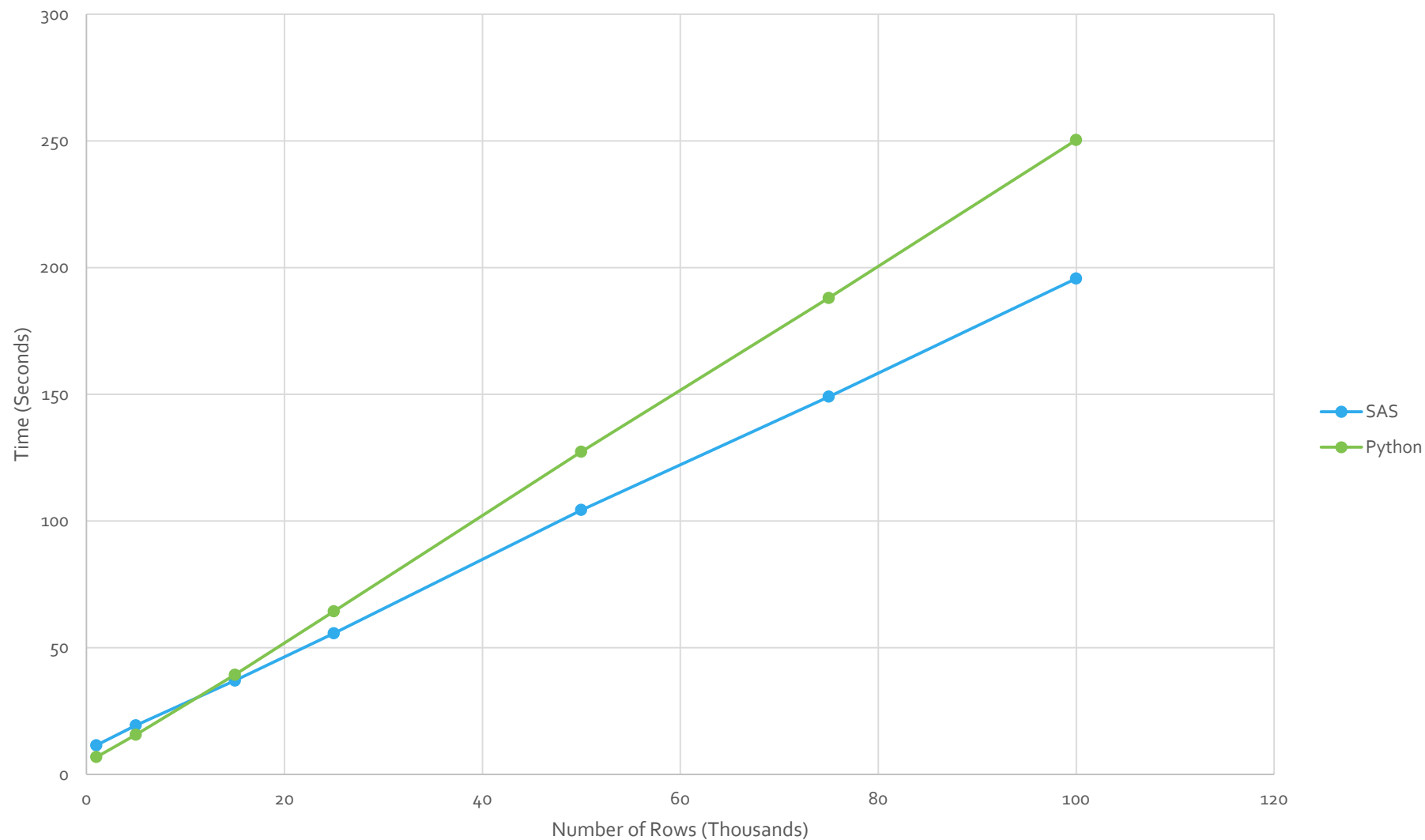
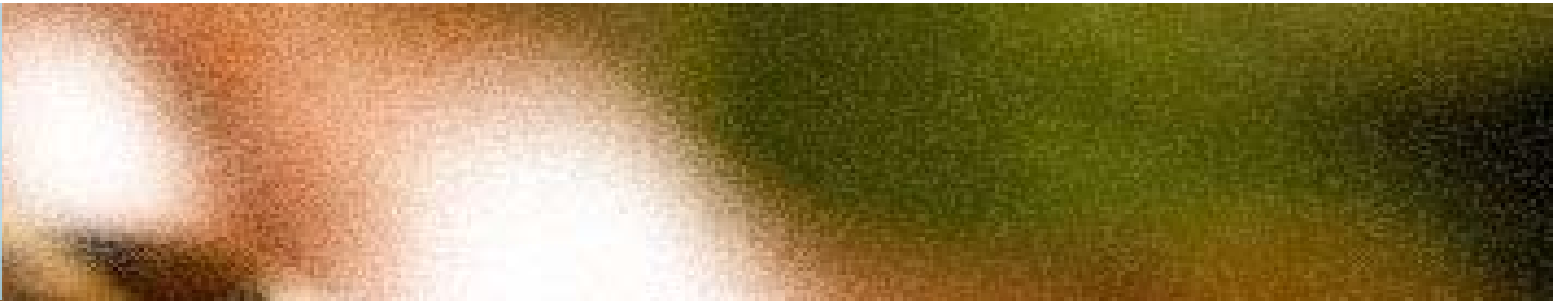
# Comparison Based on Worksheet Count



# Project Example: Comparing Excel Files

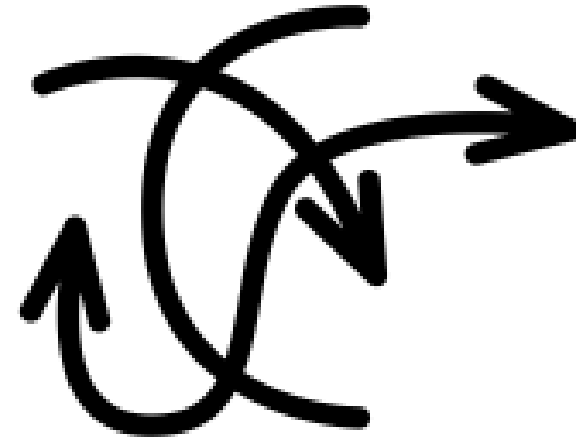
- Comparing Increasing # of Obs
  - Utilizing the same utility for large excel data dumps
  - Thousands of records & around 100 Variables

# Comparison Based on # Observations



# Takeaways: Comparing Excel Files

- Flexibility
  - SAS Cares about Spaces
  - Datasets have to follow rules



- Complexity
  - Python: 300~ lines of code
  - SAS: 1000~ lines of code



## Projects Leveraging Python



# Project Example : Updating ClinicalTrials.Gov

## ClinicalTrials.gov

A service of the U.S. National Institutes of Health

ClinicalTrials.gov is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world. [Learn more about clinical studies](#) and [about this site](#), including relevant [history](#), [policies](#), and [laws](#).

[Find Studies](#) [About Clinical Studies](#) [Submit Studies](#) [Resources](#) [About This Site](#)

ClinicalTrials.gov currently lists **209,194 studies** with locations in all 50 States and in **192 countries**. [Text Size](#)

### Search for Studies

Example: "Heart attack" AND "Los Angeles"

[Advanced Search](#) | [See Studies by Topic](#)  
[See Studies on Map](#)

#### Search Help

- [How to search](#)
- [How to find results of studies](#)
- [How to read a study record](#)

#### Locations of Recruiting Studies

| Location               | Percentage |
|------------------------|------------|
| Non-U.S. only          | 54%        |
| U.S. only              | 41%        |
| Both U.S. and non-U.S. | 6%         |

Total N = 37,436 studies  
(Data as of February 23, 2016)

- [See more trends, charts, and maps](#)

#### For Patients and Families

- [How to find studies](#)
- [See studies by topic](#)
- [Learn about clinical studies](#)
- [Learn more](#)

#### For Researchers

- [How to submit studies](#)
- [Download content for analysis](#)
- [About the results database](#)
- [Learn more](#)

#### For Study Record Managers

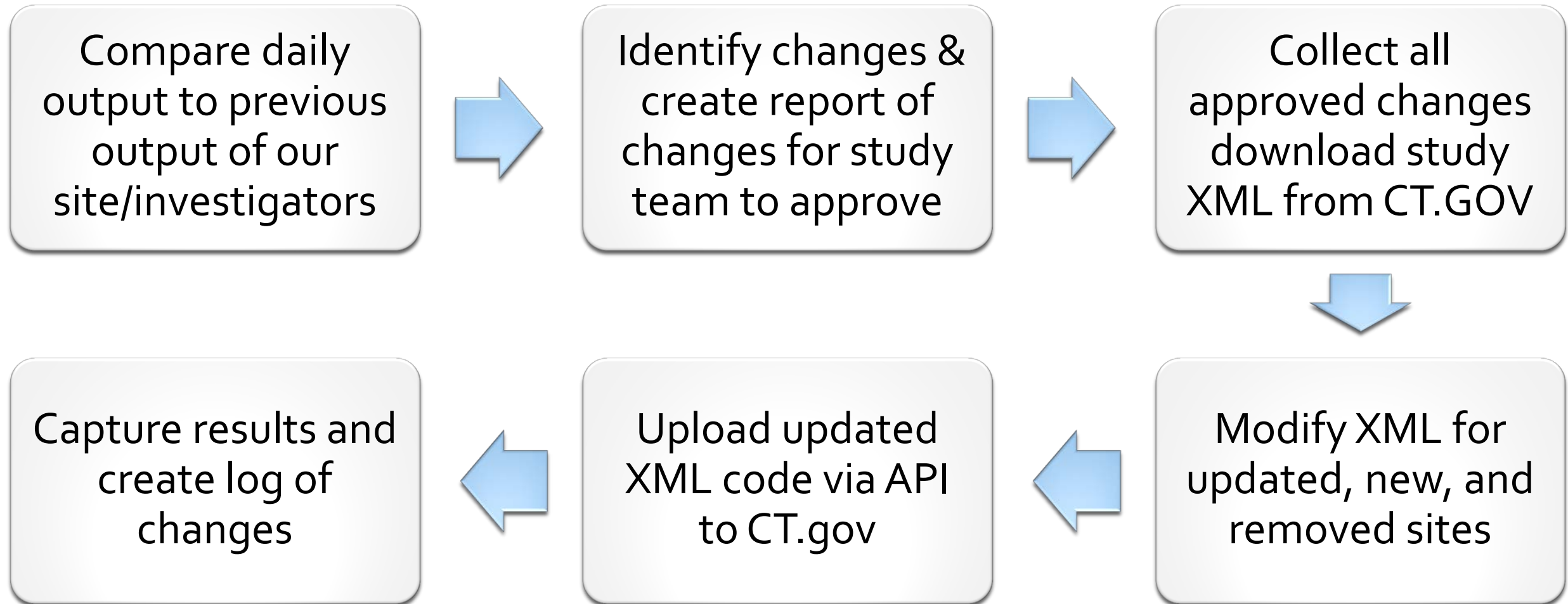
- [Why register?](#)
- [How to register your study](#)
- [FDAAA 801 requirements](#)
- [Learn more](#)

#### Learn More

- [Tutorials for using ClinicalTrials.gov](#)
- [Glossary of common site terms](#)
- [For the press](#)
- [Using our RSS feeds](#)

[HOME](#) [RSS FEEDS](#) [SITE MAP](#) [TERMS AND CONDITIONS](#) [DISCLAIMER](#) [CONTACT NLM HELP DESK](#)

# Steps: Updating ClinicalTrials.Gov



# Project Example : Updating ClinicalTrials.Gov

- Import & Compare CSV Files Using Pandas
  - Today's File vs Yesterday's File
- Using Unique Identifier Per Site
  - Changes per site
  - New Sites
  - Removed Sites



# Project Example : Updating ClinicalTrials.gov

- Generate Change Report
  - Using library `xlsxwriter`

| Protocol       | Site Number    | Facility Name | # Subjects Enrolled | PI First Name | PI Last Name | City Value   | State Value | Zipcode Value | Region Value | Changed Field         | Changed Value                                        | Status         |
|----------------|----------------|---------------|---------------------|---------------|--------------|--------------|-------------|---------------|--------------|-----------------------|------------------------------------------------------|----------------|
| PY-PL-123-3210 | 123-3210-12616 | Facility 1    | 1234                | First Name 1  | Last Name 1  | City Name 1  |             | 123456        | Region 1     | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12699 | Facility 2    | 1234                | First Name 2  | Last Name 2  | City Name 2  |             | 123456        | Region 2     | Site Zipcode Changed  | Old Value: Zipcode Old    New Value: Zipcode New     | Review Pending |
| PY-PL-123-3210 | 123-3210-13284 | Facility 3    | 1234                | First Name 3  | Last Name 3  | City Name 3  |             | 123456        | Region 3     | Site Zipcode Changed  | Old Value: Zipcode Old    New Value: Zipcode New     | Review Pending |
| PY-PL-123-3210 | 123-3210-12616 | Facility 4    | 1234                | First Name 4  | Last Name 4  | City Name 4  |             | 123456        | Region 4     | Site Zipcode Changed  | Old Value: Zipcode Old    New Value: Zipcode New     | Review Pending |
| PY-PL-123-3210 | 123-3210-12682 | Facility 5    | 1234                | First Name 5  | Last Name 5  | City Name 5  |             | 123456        | Region 5     | Site City Changed     | Old Value: City Name Old    New Value: City Name New | Review Pending |
| PY-PL-123-3210 | 123-3210-12454 | Facility 6    | 1234                | First Name 6  | Last Name 6  | City Name 6  |             | 123456        | Region 6     | Site City Changed     | Old Value: City Name Old    New Value: City Name New | Review Pending |
| PY-PL-123-3210 | 123-3210-12699 | Facility 7    | 1234                | First Name 7  | Last Name 7  | City Name 7  |             | 123456        | Region 7     | Site City Changed     | Old Value: City Name Old    New Value: City Name New | Review Pending |
| PY-PL-123-3210 | 123-3210-12680 | Facility 8    | 1234                | First Name 8  | Last Name 8  | City Name 8  |             | 123456        | Region 8     | Site City Changed     | Old Value: City Name Old    New Value: City Name New | Review Pending |
| PY-PL-123-3210 | 123-3210-12622 | Facility 9    | 1234                | First Name 9  | Last Name 9  | City Name 9  |             | 123456        | Region 9     | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12685 | Facility 10   | 1234                | First Name 10 | Last Name 10 | City Name 10 |             | 123456        | Region 10    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12780 | Facility 11   | 1234                | First Name 11 | Last Name 11 | City Name 11 |             | 123456        | Region 11    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12683 | Facility 12   | 1234                | First Name 12 | Last Name 12 | City Name 12 |             | 123456        | Region 12    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12710 | Facility 13   | 1234                | First Name 13 | Last Name 13 | City Name 13 |             | 123456        | Region 13    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12693 | Facility 14   | 1234                | First Name 14 | Last Name 14 | City Name 14 |             | 123456        | Region 14    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12454 | Facility 15   | 1234                | First Name 15 | Last Name 15 | City Name 15 |             | 123456        | Region 15    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12628 | Facility 16   | 1234                | First Name 16 | Last Name 16 | City Name 16 |             | 123456        | Region 16    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12692 | Facility 17   | 1234                | First Name 17 | Last Name 17 | City Name 17 |             | 123456        | Region 17    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12615 | Facility 18   | 1234                | First Name 18 | Last Name 18 | City Name 18 |             | 123456        | Region 18    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12690 | Facility 19   | 1234                | First Name 19 | Last Name 19 | City Name 19 |             | 123456        | Region 19    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12617 | Facility 20   | 1234                | First Name 20 | Last Name 20 | City Name 20 |             | 123456        | Region 20    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12704 | Facility 21   | 1234                | First Name 21 | Last Name 21 | City Name 21 |             | 123456        | Region 21    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12613 | Facility 22   | 1234                | First Name 22 | Last Name 22 | City Name 22 |             | 123456        | Region 22    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12680 | Facility 23   | 1234                | First Name 23 | Last Name 23 | City Name 23 |             | 123456        | Region 23    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12618 | Facility 24   | 1234                | First Name 24 | Last Name 24 | City Name 24 |             | 123456        | Region 24    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12682 | Facility 25   | 1234                | First Name 25 | Last Name 25 | City Name 25 |             | 123456        | Region 25    | Site Zipcode Changed  | Old Value: Zipcode Old    New Value: Zipcode New     | Review Pending |
| PY-PL-123-3210 | 123-3210-12828 | Facility 26   | 1234                | First Name 26 | Last Name 26 | City Name 26 |             | 123456        | Region 26    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12619 | Facility 27   | 1234                | First Name 27 | Last Name 27 | City Name 27 |             | 123456        | Region 27    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12457 | Facility 28   | 1234                | First Name 28 | Last Name 28 | City Name 28 |             | 123456        | Region 28    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12627 | Facility 29   | 1234                | First Name 29 | Last Name 29 | City Name 29 |             | 123456        | Region 29    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12699 | Facility 30   | 1234                | First Name 30 | Last Name 30 | City Name 30 |             | 123456        | Region 30    | All                   | New Site                                             | Review Pending |
| PY-PL-123-3210 | 123-3210-12681 | Facility 31   | 1234                | First Name 31 | Last Name 31 | City Name 31 |             | 123456        | Region 31    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12456 | Facility 32   | 1234                | First Name 32 | Last Name 32 | City Name 32 |             | 123456        | Region 32    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12626 | Facility 33   | 1234                | First Name 33 | Last Name 33 | City Name 33 |             | 123456        | Region 33    | Facility Name Changed | Old Value: Facility Name Old    New Value: Facility  | Review Pending |
| PY-PL-123-3210 | 123-3210-12685 | Facility 34   | 1234                | First Name 34 | Last Name 34 | City Name 34 |             | 123456        | Region 34    | Site Zipcode Changed  | Old Value: Facility Name Old    New Value: Facility  | Review Pending |

# Project Example : Updating ClinicalTrials.Gov

- Download XML via API
  - Use Library requests
- Modify XML
  - Use Library lxml

```
<location>
  <facility>
    <name>Kent Innovation Centre</name>
    <address>
      <city>Broadstairs, Kent</city>
      <state>None</state>
      <country>UK</country>
      <zip>CT10 2QQ</zip>
    </address>
  </facility>
  <status>Enrolling By Invitation</status>
  <contact>
    <first_name/>
    <middle_name/>
    <last_name/>
    <degrees/>
    <phone/>
    <phone_ext/>
    <email/>
  </contact>
  <contact_backup>
    <first_name/>
    <middle_name/>
    <last_name/>
    <degrees/>
    <phone/>
    <phone_ext/>
    <email/>
  </contact_backup>
</location>
```

# Project Example : Updating ClinicalTrials.Gov

- Upload XML via API
  - Use Library requests
  - Grab Response from API

```
<?xml version="1.0" encoding="UTF-8" ?>
<externalUploadReport>
  <uploadStatus>CHANGED_STUDY</uploadStatus>
  <validationMessages>
    <validationMessage>
      <severity>WARNING</severity>
      <source>property:Study.locations[0].Facility.Addr.City</source>
      <message>BLANK is not a recognized city for the specified state/country.</message>
    </validationMessage>
    <validationMessage>
      <severity>WARNING</severity>
      <source>property:Study.locations[44].Facility.Addr.City</source>
      <message>BLANK is not a recognized city for the specified state/country.</message>
    </validationMessage>
    <validationMessage>

```

# Project Example : Updating ClinicalTrials.Gov

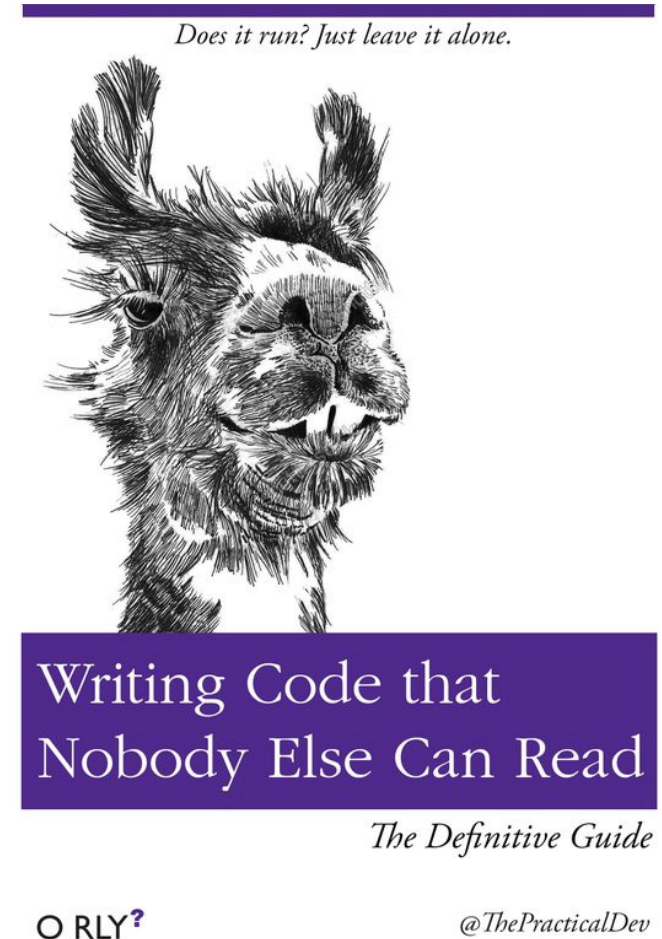
- Create Excel Log File to Track Status of Changes
  - Using library `xlsxwriter`

| Protocol       | Change Type | Change Value | Program Change Comment | Program Change Status | XML Download Comment | XML Download Status | XML Upload Comment | XML Upload Status |
|----------------|-------------|--------------|------------------------|-----------------------|----------------------|---------------------|--------------------|-------------------|
| PY-PL-123-4321 | All         | New Site     | Added New Facility     | Success               | XML Downloaded       | Success             | CHANGED_STUDY      | Success           |
|                |             |              |                        |                       |                      |                     |                    |                   |

- Grabs response from each stage
  - Python Program
  - XML Download
  - XML Upload

# Project Example : Updating ClinicalTrials.Gov

- Benefits of using Python For Task
  - Customizable Excel Output
  - Straight Forward XML Modification
  - Requests automatically encodes
  - Easy to receive/examine response from API
  - Many Complex parts split into concise manageable pieces
    - “Simple is better than complex. Complex is better than complicated”  
- Tim Peters, Zen of Python



# Project Example: Anonymization of Documents

- EMA Has Introduced Policy 0070
  - Requiring Anonymization of Study Documents
- Need for automated method to identify potentially identifying information
- Transformation vs. Redaction:
  - Delivery of higher-value output with greater utility yet still with strong measures to protect patient privacy



# Project Example: Anonymization of Documents

- Still at Early Stages and Exploring Best Method to Accomplish Task
- Distinct Needs:
  - ID Personal Data
  - Create method of viewing these items
  - Allow user to approve changes
  - Apply changes to documents



```
<bound method Tree.label of Tree> 'GPE', [['American', 'NNP']]> American
<bound method Tree.label of Tree> 'ORGANIZATION', [['Society', 'NNP']]> Society
<bound method Tree.label of Tree> 'GPE', [['Hematology', 'NNP']]> Hematology
<bound method Tree.label of Tree> 'ORGANIZATION', [['ASH', 'NNP']]> ASH
<bound method Tree.label of Tree> 'PERSON', [['Annual', 'NNP'], ['Meeting', 'NNP']]> Annual Meeting
<bound method Tree.label of Tree> 'GPE', [['Exposition', 'NNP']]> Exposition
<bound method Tree.label of Tree> 'GPE', [['Orlando', 'NNP']]> Orlando
<bound method Tree.label of Tree> 'ORGANIZATION', [['FL', 'NNP']]> FL
<bound method Tree.label of Tree> 'PERSON', [['Barrientos', 'NNP']]> Barrientos
<bound method Tree.label of Tree> 'GPE', [['JC', 'NNP']]> JC
<bound method Tree.label of Tree> 'PERSON', [['Hirani', 'NNP'], ['JR', 'NNP']]> Hirani JR
<bound method Tree.label of Tree> 'PERSON', [['Mendamin', 'NNP']]> Mendamin
<bound method Tree.label of Tree> 'PERSON', [['Rituximab', 'NNP']]> Rituximab
<bound method Tree.label of Tree> 'GPE', [['Patients', 'NNP']]> Patients
<bound method Tree.label of Tree> 'PERSON', [['Adverse', 'NNP'], ['Prognostic', 'NNP'], ['Features', 'NNP']]> Adverse Prognostic Features
<bound method Tree.label of Tree> 'GPE', [['American', 'NNP']]> American
<bound method Tree.label of Tree> 'ORGANIZATION', [['Society', 'NNP']]> Society
<bound method Tree.label of Tree> 'PERSON', [['Clinical', 'NNP'], ['Oncology', 'NNP']]> Clinical Oncology
<bound method Tree.label of Tree> 'ORGANIZATION', [['ASCO', 'NNP']]> ASCO
<bound method Tree.label of Tree> 'PERSON', [['Annual', 'JJ'], ['Meeting', 'NNP']]> Annual Meeting
<bound method Tree.label of Tree> 'GPE', [['Chicago', 'NNP']]> Chicago
<bound method Tree.label of Tree> 'ORGANIZATION', [['IL', 'NNP']]> IL
<bound method Tree.label of Tree> 'GPE', [['Hillman', 'NNP']]> Hillman
<bound method Tree.label of Tree> 'PERSON', [['P', 'NNP']]> P
<bound method Tree.label of Tree> 'PERSON', [['Ferra', 'NNP'], ['C', 'NNP']]> Ferra C
<bound method Tree.label of Tree> 'GPE', [['Idelalisib', 'NNP']]> Idelalisib
<bound method Tree.label of Tree> 'PERSON', [['Mendamin', 'NNP']]> Mendamin
<bound method Tree.label of Tree> 'PERSON', [['Hirani', 'NNP'], ['Improves', 'NNP'], ['Overall', 'NNP'], ['Survival', 'NNP']]> Hirani Improves Overall Survival
<bound method Tree.label of Tree> 'GPE', [['Patients', 'NNP']]> Patients
<bound method Tree.label of Tree> 'ORGANIZATION', [['Randomized', 'NNP']]> Randomized
<bound method Tree.label of Tree> 'GPE', [['European', 'JJ']]> European
<bound method Tree.label of Tree> 'ORGANIZATION', [['Hematology', 'NNP'], ['Association', 'NNP']]> Hematology Association
<bound method Tree.label of Tree> 'ORGANIZATION', [['EHA', 'NNP']]> EHA
<bound method Tree.label of Tree> 'PERSON', [['Annual', 'JJ'], ['Meeting', 'NNP']]> Annual Meeting
<bound method Tree.label of Tree> 'GPE', [['Copenhagen', 'NNP']]> Copenhagen
<bound method Tree.label of Tree> 'PERSON', [['Denmark', 'NNP']]> Denmark
<bound method Tree.label of Tree> 'PERSON', [['Study', 'NNP']]> Study
<bound method Tree.label of Tree> 'PERSON', [['Period', 'NNP']]> Period
<bound method Tree.label of Tree> 'ORGANIZATION', [['First', 'NNP'], ['Subject', 'NNP']]> First Subject
<bound method Tree.label of Tree> 'PERSON', [['Safety', 'NNP']]> Safety
<bound method Tree.label of Tree> 'PERSON', [['Overall', 'NNP'], ['Survival', 'NNP']]> Overall Survival
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree-label of Tree> 'PERSON', [['P-S', 'NNP']]> P-S
<bound method Tree.label of Tree> 'ORGANIZATION', [['CLL', 'NNP']]> CLL
<bound method Tree.label of Tree> 'PERSON', [['Idelalisib', 'NNP'], ['Study', 'NNP']]> Idelalisib Study
<bound method Tree.label of Tree> 'PERSON', [['Clinical', 'NNP'], ['Report', 'NNP'], ['Final', 'NNP']]> Clinical Study Report Final
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['OS', 'NNP']]> OS
<bound method Tree.label of Tree> 'ORGANIZATION', [['PRQL', 'NNP']]> PRQL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['BR', 'NNP']]> BR
<bound method Tree.label of Tree> 'ORGANIZATION', [['TD', 'NNP']]> TD
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IDL', 'NNP']]> IDL
<bound method Tree.label of Tree> 'ORGANIZATION', [['BR', 'NNP']]> BR
<bound method Tree.label of Tree> 'PERSON', [['Study', 'NNP']]> Study
<bound method Tree.label of Tree> 'ORGANIZATION', [['TP53', 'NNP']]> TP53
<bound method Tree.label of Tree> 'ORGANIZATION', [['CLL', 'NNP']]> CLL
<bound method Tree.label of Tree> 'ORGANIZATION', [['IG4V', 'NNP']]> IG4V
<bound method Tree.label of Tree> 'PERSON', [['Bendamun Line', 'NNP']]> Bendamin Line
```

# Project Example: Anonymization of Documents

- Convert Word Document to XML

```
<w:p w:rsidP="00BB0A0B" w:rsidR="00BB0A0B" w:rsidRDefault="00BB0A0B">
  <w:pPr>
    <w:autoSpaceDE w:val="0"/>
    <w:autoSpaceDN w:val="0"/>
    <w:adjustRightInd w:val="0"/>
    <w:spacing w:after="0" w:line="240" w:lineRule="auto"/>
    <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:cs="Times New Roman" w:hAnsi="Times New Roman"/>
      <w:color w:val="000000"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
    </w:rPr>
  </w:pPr>
  <w:r>
    <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:cs="Times New Roman" w:hAnsi="Times New Roman"/>
      <w:color w:val="000000"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
    </w:rPr>
    <w:t>This analysis set was used in the analyses of subject characteristics, PFS, ORR, lymph node</w:t>
  </w:r>
</w:p>
```



Thoughts and Considerations



# Future Projects

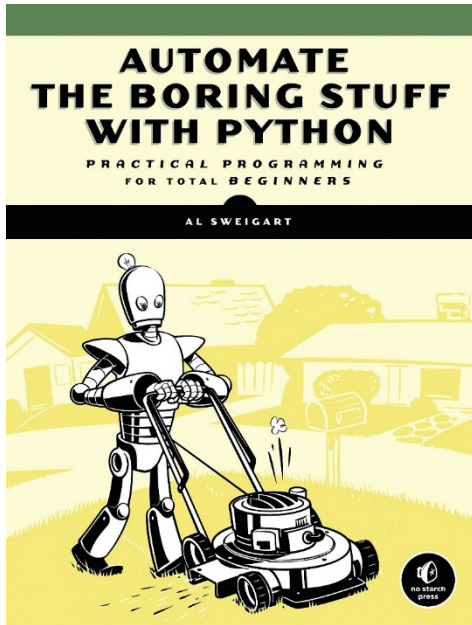
- Creation of Define XML
- Merging PDF Files
- Submission Checking Utility



# Growing Pains

- How to control packages
  - Do we want any user to have the ability to install packages
  - How do we want to ensure packages are trustworthy?
- How to validate scripts
  - Capitalize on python's validation/testing abilities
- Ensuring functionality when updating modules/python
  - Scripts behave the same when new versions of packages are released/installed
  - New versions of python

# Resources for Learning



PEP 8 Style Guide:

<https://www.python.org/dev/peps/pep-0008/>

PEP 20 The Zen of Python:

<https://www.python.org/dev/peps/pep-0020/>

Questions?

