# Workflow Manager Documentation

Jeshuran Pandian Thangaraj

Below are the tools used and some tips and tricks for a workflow manager.

**Tools**

- Microsoft Conference Management Tool (CMT) (https://cmt3.research.microsoft.com)
- Table Capture (https://chrome.google.com/webstore/detail/tablecapture/iebpjdmgckacbodjpijphcplhebcmeop?hl=en)
- Awesome Screenshot (https://chrome.google.com/webstore/detail/awesome-screenshot-screen/nlipoenfbbikpbjkfpfillcgkoblgpmj?hl=en)
- JSONView (https://chrome.google.com/webstore/detail/jsonview/chklaanhfefbnpoihckbnefhakgolnmc)

**Technical Skills for workflow manager**

Most workflow tasks will follow the same pattern of techniques
Extract Data from CMT -> Manipulate Data -> Create Report/ Send Email

**Extracting Data**

Data extraction from CMT can be done in multiple ways

- Bulk exports from CMT

Bulk exports of most data can be done from CMT in the options tab. Some of the exports take into consideration filters in the data while others do not.
The file formats of these are a little problematic. The “Export to Excel” exports in .xls format which is archaic, and python struggles to read it.

**Suggestion:** Beware of converting .xls to other formats using Microsoft excel as this can create data discrepancies.

XML and Tab Limited work well with python. The bulk of data is exported in excel and also it is filterable. To fix this problem, we use the second method.

- Scrape the table for data

Once we filter the data with the criteria we want. We can scrape the table to obtain all the data. The way I do it is by using the Table Capture extension for chrome.

Make sure you click on view all data before using the extension. The extension creates a google sheet which can then be exported as Tab separated.

- Using the API

There is an API provided by CMT3. It is still under development but seems to be quite stable. Code to login can be found here.
https://github.com/jpthanga/Conference-Workflow/tree/master/CMT%20API%20Login

This is a OData service and the start point
https://cmt3.research.microsoft.com/api/odata/<ConferenceName>/

This will also give you a list of services. For more details on which are active and the exact syntax to access them, contact "CMT Support" <support@msr-cmt.org>

**Manipulate Data**

Manipulating data is done in python. Depending on requirement every library from numpy to matplotlib to pandas is used.

Find some functions used for manipulating data imported from CMT here.

https://github.com/jpthanga/Conference-Workflow/tree/master/Data%20Processing%20Scripts

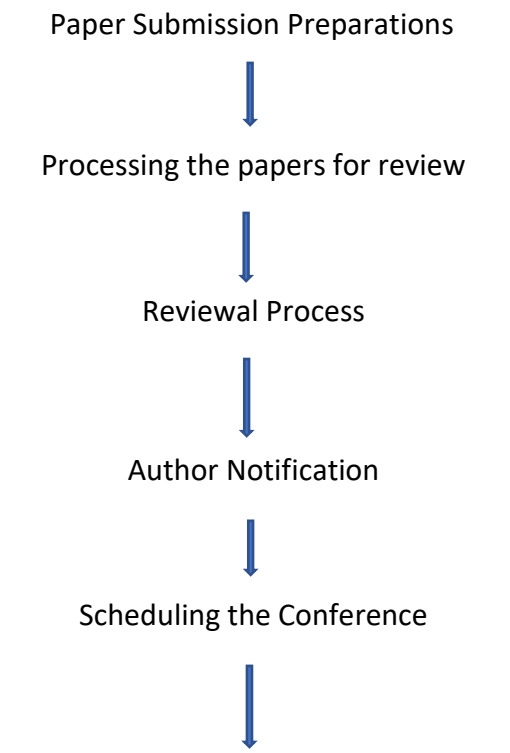
**Create Report / Send Email**

Depending on the requirement, we might create a report with the data or send an email to a subset of area chairs, reviewers or authors.

Reports are created on a one of basis and sending emails is discussed below.

Here, I document the different hats a workflow manager must wear and also document a bit of code that could be helpful.

**General lifecycle of a Conference:**




I will go through each of these steps and discuss what is required from the workflow perspective. Also, I will add my finding on risks and what can be done better in future conferences.

**Paper Submission Preparations:**

The weeks before the call for papers go out is critical to the success of the workflow. This period gives the program chairs and workflow chairs the time to streamline processes before the chaos of submission and reviewal processes begin.

Things to work on during this period may include:

- **Streamlining communication between program chairs and workflow chairs**

One of the challenges for the workflow manager is the fact that they will be working remotely from the program chairs. Therefore, setting expectations and well-established lines of communication are very important.

For NeurIPS 2018, we had a slack channel with all program chairs and workflow chairs which facilitated quick responses and doubt clearing. We also had one or two video conferences per week depending on urgency and requirements of that week.

Program Chairs do conferences on top of their day jobs and therefore are very busy. Hence, setting expectations of when and how much everyone is available is also required.

- **Invitation of Reviewers**

As meetings grow, there is a need for more and more reviewers. The call for reviewers phase is pretty busy for the workflow manager. One of the responsibilities of the workflow manager begins at this point.

Workflow managers are expected to send and monitor emails to Authors, Reviewers, Meta-Reviewers and everyone else required.

For NeurIPS 2018, outgoing e-mails were done using a script () which used a personal email. Limitations of this was the fact that Gmail allows only 1000 emails from a certain account in a 24-hour period.

https://github.com/jpthanga/Conference-Workflow/tree/master/Email%20Script

**Suggestion:** Investing in a g suite account will fix this problem.

For incoming inquiries, all authors and reviewers wanting to contact a program chair were asked to reply to a JIRA where the emails are converted to tickets for easy processing. This worked well and could be used again even though it has a few quirks.

- **Any Scripts that can be done**

Any scripts from the other steps that can be completed before the submissions could be done as this will make time management very easy later on.

**Processing Paper for Review:**

Once the paper deadline is complete. There are a few important steps that need to be taken.

- **Duplicate Submissions**

Finding duplicate submissions is a step done to prevent authors from submitting multiple papers with the same idea and preventing submissions that were already accepted in other conferences. Some code for this can be found here.

https://github.com/chisingh/nips/blob/master/notebooks/Identify%20Shotgun%20Submissions.ipynb
https://github.com/chisingh/nips/blob/master/scripts/duplicates.py
https://github.com/chisingh/nips/blob/master/scripts/duplicates_abstract.py
https://github.com/chisingh/nips/blob/master/scripts/dual_submissions.py

Contact Chinmay Singh <chinmay.singh@gmail.com> for access and information.

- **Assigning Submissions to Reviewers**

Submissions are assigned to reviewers using a combination of TPMS (Toronto Paper Matching System) Scores and Bids on CMT.

There is a script written by Samy Bengio to do this. Program Chairs should be able to point you to this.

**Review Process:**

Bulk of the emails go out during the review process.

This phase is most Program chair intensive and the workflow supports the program chairs by sending bulk emails and generating reports of late reviewers etc.

**Scheduling the Conference:**

Once the Authors have been notified of accepted papers. The schedule of the conference needs to be prepared.

For NeurIPS 2018, this was done using a script which can be found here.

https://github.com/jpthanga/Conference-Workflow/tree/master/Scheduler

**Reporting statistics for the conference:**

The final step is to get some statistics reports for the conference. This could include things like number of reviewers, total number of submissions etc.

At the end of all this, the workflow manager gets to go to the conference and help out and enjoy the conference.