CvT Template Curation Workflow

# Summary of Steps

1. Template submitted by curator
   1. Curation approaches
      1. Option 1: Curated in Excel file with Excel (or software equivalent)
      2. Option 2: Curated in RShiny CvTdb Extraction application (**pending deployment**)
   2. Storage
      1. Stored in ExpoCast folder on L-Drive under “Format QA” subfolder
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\0\_to\_qa\_format"
2. Format QA Check ([**I-SCDCD-DEQEB-SOP-4604-0**](https://qatrack.epa.gov/sop_tab/show/4604/))
   1. Submitted template is reviewed for any formatting issues
      1. Reviewer may opt to use the Format QA RShiny application to assist in review
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\RShiny QA Tool"
      2. Template and document names are standardized and stored in ExpoCast folder on L-Drive (see Format QA SOP)
      3. Reviewer adds an entry for the template to the Format QA Log
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\qa\_log.xlsx"
         2. **Note:** be sure to check for duplicate logs and combine or reconcile to one version. Sometimes happens in this collaborative environment
   2. Following review, template is moved from “0\_to\_qa\_format” folder to another subfolder based on the result of the review (see Format QA SOP)
   3. TO ADD
      1. Determining if they are actually CvT data relevant
         1. Checklist or something added to RShiny application
3. Template Normalization (CvT Load Data SOP” {**drafting**})
   1. Templates in “1\_qa\_format\_complete” are run through the “normalization” R Script workflow
      1. [CCTE Bitbucket for CvTdb](https://ccte-bitbucket.epa.gov/projects/CVTDB)
   2. Dictionary updates (output sent to SME if needed for update help)
      1. Administration Route
      2. Species
      3. Concentration Medium
      4. TO ADD
         1. Organization/sorting of templates for weird or invalid dictionary entries
         2. Also add a check for conc\_units
   3. See “CvT Load Data SOP” for full normalization details
   4. Normalized templates are stored in ExpoCast folder on the L-Drive
      1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\normalized\_templates"
4. SME Template Normalization Review
   1. Spot check review of normalized templates
      1. Consultation with normalization log for flagged issues
      2. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\template\_normalization\_log.xlsx"
   2. Ignores the “missing\_required\_fields” folder of templates
   3. Log Series ID values to load or not load
      1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\template\_metadata\_qa.xlsx"
   4. TO ADD
      1. Which flags are hard stops vs. soft stop warnings. How to kick them back to curation? Go back to original curator?
         1. Filter out the obvious ones
      2. Split:
         1. Template is ready to go (all)
         2. Template is partial (some are ready, some not)
         3. Template overall needs more
5. Chemical Curation
   1. Follows after SME for final spot checks for which series are valid and for chemical info checks
   2. Jira ticket submissions to Chemical Curation Team
6. Template Loading
   1. Templates are loaded to CvT using the updated dictionaries and the Series ID values identified as ready to load by the SME review
   2. See “CvT Load Data SOP” for full loading details
   3. Loaded templates are moved from “1\_qa\_format\_complete” subfolder to “2\_pushed\_to\_database” subfolder of ExpoCast folder on the L-Drive
7. Loaded template assets uploaded to Clowder
   1. Template Assets
      1. Original template
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Original Extracted Templates"
      2. Format QA template
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\2\_pushed\_to\_database"
      3. Normalized template
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\Format QA\normalized\_templates"
      4. Loadable Templates
         1. Final phase of a template
      5. PDF
         1. "L:\Lab\NCCT\_ExpoCast\ExpoCast2022\CvT-CompletedTemplates\CvT PDFs"
   2. Logs uploaded to Clowder

# To Discuss

1. Schema changes
   1. Improved template provenance with curator LAN ID
   2. Chemicals table (or even for all dictionary style tables) – align with Toxval and others
      1. Add “raw chemical” information to chemicals table
      2. Only have chemical ID in series table and studies table
   3. TK Params
2. Partial Uploads
   1. Only way we track and know what’s been uploaded is by the normalization review log that lists series to load or not
      1. At least without having to try to match back to everything between a database entry and a template
      2. Concern for how to maintain a partially uploaded template and how they’ll eventually be edited to include the remaining series ID values
   2. Do we place this in the database as well, so it’s not a floating file somewhere?
3. Eventually move towards a special curation process using an application. Do not allow users to manually change partially uploaded templates in Excel! Has special logic to guard it.
   1. Have records on a series basis instead of a document basis (with some field immutable)
   2. Display in wide form for curation instead of by sheet?