# Structure-Data File from Sponsors' Submissions Promotes Rapid Chemical Registration and Computational Analyses

**Marlene Kim[1*], Tyler Peryea[2], Barbara Scott[3], David Green[3], and Naomi Kruhlak[1]**

[1]Office of Translational Science/Office of Clinical Pharmacology/Division of Applied Regulatory Science (DARS)

[2]Office of the Commissioner (OC)/Office of Chief Scientist/Office of Health Informatics

[3]Office of Pharmaceutical Quality/Office of New Drugs/Division of Lifecycle API (DLAPI)



## Background

- Chemical structures are submitted through the Electronic Common Technical Document (eCTD) within PDFs as embedded images.
- FDA chemists review thousands of chemicals a year.
- Cheminformaticians and assessors currently spend hours to days per application isolating and redrawing the chemical structures into computer-readable chemical formats for databasing efforts, computational analyses, and/or internal reviews.
- This step is prone to errors and is time consuming. Redrawing the structures is an unnecessary burden to FDA.
- The pharmaceutical industry already has chemical structures in a computer-readable format prior to regulatory submission; however, the chemical format and associated data (e.g., chemical name, CAS) is lost when the structures are converted to PDFs.
- In contrast, the Structure-Data File (SD File) format can encode computer-readable chemical structures and associated data.
- Therefore, a collaboration between OPQ, OTS and OC resulted in the inclusion of SD Files as an acceptable file format in the eCTD.
- The pharmaceutical industry may now submit structures to the Agency through the eCTD as a single SD File with standardized data fields meeting the Agency's cheminformatics and review needs[1].

## What is an SD File?

- An SD File is an extensible, portable text file encoding computer-readable chemical structures linked to associated data fields.
- Unlike MOL Files, SD Files can include more than one chemical structure per file.

**Sample SD File**

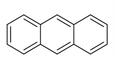
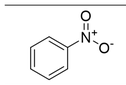
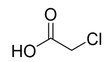
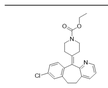
| Structure | ID | NAME | UNII | CAS | ROLE | APPLICATION_NUMBER |
|---|---|---|---|---|---|---|
|  | Structure 1 | anthracene | EH46A1TLD7 | 120-12-7 | process impurity | MF-012345 |
|  | Structure 2 | nitrobenzene | E57JCN6SSY | 98-95-3 | starting material | MF-012345 |
|  | Structure 3 | 2-chloroacetic acid | 5GD84Y125G | 79-11-8 | starting material | MF-012345 |
|  | Structure 4 | loratadine | 7AJO3BO7QN | 79794-75-5 | active ingredient | MF-012345 |

**Figure 1.** Example of an SD File suitable for DMF submissions. The file includes drug substance name, UNII, CAS, role, and application number. Freely available and/or commerical software tools can aid in creating or editing SD Files. A *Quick Guide for Creating a Structure Data File (SD File) for DMF Submissions* was created to help Industry with this new process[2].

[1] Please see companion poster by Scott et al., titled "Stop Re-drawing Chemicals! Implementation of Computer-Readable Chemical Structure Format for Drug Impurities"

[2] *https://www.fda.gov/drugs/forms-submission-requirements/drug-master-files-dmfs*

## Additional Benefits

- The SD File format facilitates efficient substance registration in the Global Substance Registration System (GSRS), computational analyses (e.g., (quantitative) structure-activity relationship, or (Q)SAR), model predictions), and regulatory reviews.
- Specialized systems like GSRS and (Q)SAR software cannot efficiently interpret images and can only accept chemical structure files, respectively. In contrast, SD Files are *computer-ready* and require little to no processing before moving forward to register or run (Q)SAR predictions.
- By expanding the eCTD to accept SD Files, the chemical structures retain their native file format and associated data, thereby eliminating unnecessary rework and increasing the accuracy of the entire review process.

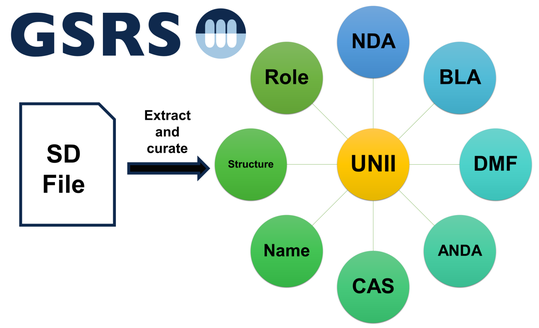



**Figure 2.** GSRS assigns a Unique Ingredient Identifier (UNII) to substances submitted to FDA. Industry can request UNIIs at FDA-SRS@fda.hhs.gov. Submitting SD Files and including UNII codes allows FDA to easily track the history of a chemical across different applications which promotes consistency in regulatory decisions.

## Old vs New Method of Processing and Registering Chemicals

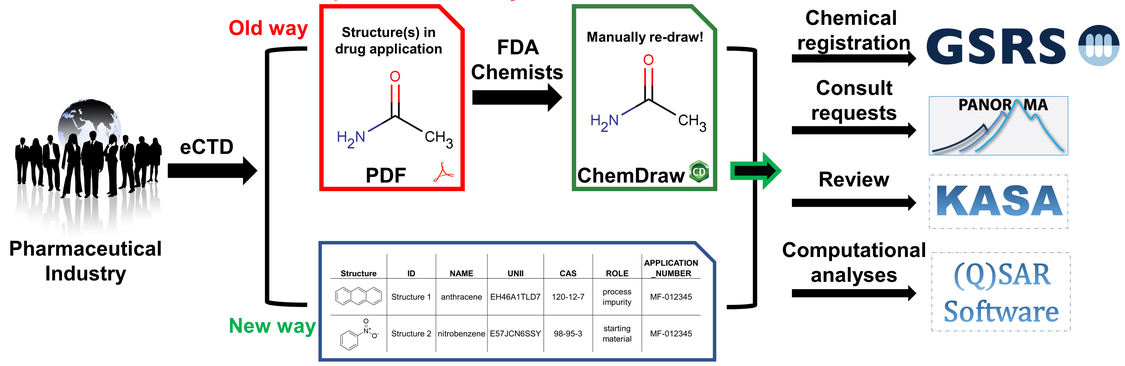



**Figure 3.** Currently, the pharmaceutical industry submits drug application files into the eCTD, where chemical structures are embedded in PDFs. FDA chemists spend hours to days per application isolating and redrawing chemical structures into an appropriate computer-ready format (e.g., ChemDraw, SD File, JSON). The file is then sent across the Agency where chemicals are registered in GSRS, sent for consultations, reviewed and/or undergo computational analyses. Allowing Industry to submit SD Files saves FDA time, where the chemist now only has to review the file or simply copy-paste the structures of interest for further analyses.

## Conclusions

Including the SD Files as an acceptable file format in the eCTD promotes efficiency across various CDER offices. Data associated with the chemical is also conserved and allows FDA to easily track the chemical across applications. Since the Agency receives thousands of chemical structures per year, this streamlined process increases accuracy and eliminates redundancy in GSRS registration efforts, computational analyses, and regulatory reviews.

## Acknowledgements

Thank you to Norman Schmuff, Jonathan Resnick, Jeff Florian, and Kirk Roy for supporting and helping us achieve this goal.