Excel tools for g-SRS

MITCH MILLER

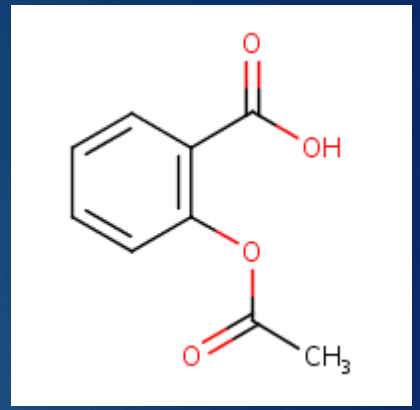
SCIENTIFIC THINKING, LLC

OCTOBER 11, 2017

Whys and wherefores

- ginas provides interactive data entry in the web application
  - Fine for a handful of records or data that requires a lot user attention
- Suppose you have a lot of data to load
  - 10, 000 records from a legacy system to load?
  - 1000 records to which you want to add a new synonym?
  - 2000 records whose code URLs have changed formats?
- Ginas bulk data loading tools provide a solution

Ginas API



- Users of new software systems usually look for an API…
- https://tripod.nih.gov/ginas/#/gsrs/api
  - Application Programmer Interface
  - A RESTful web service that provides key functions of the application in form that can be integrated with other applications
- Ginas API provides search, retrieval and update of a substance
- Swagger page with API information: https://tripod.nih.gov/ginas/#/gsrs/api

Constraints on data loading

- MS Excel as data entry ‘client’
- Single sign-on
- VBA webbrowser control
  - Older JavaScript interpreter
  - Tyler has done amazing work to make recent JavaScript functionality available within an older environment

Top-level Client Components
Contained within the sheets of a ginas Excel file

- Sheet with data to load
- Configuration sheet
  - Which server to use
- HTML definition sheet
  - Provides a basic UI
- JavaScript sheet
  - Process data and send to server
- Controlled vocabularies
- VBA Macros
  - Provide locate and invoke the JavaScript, passing va
  - Add processing results back to the sheet

What you can do with the data loading tools

- Load data from a spreadsheet into ginas
- Retrieve data from ginas into a sheet

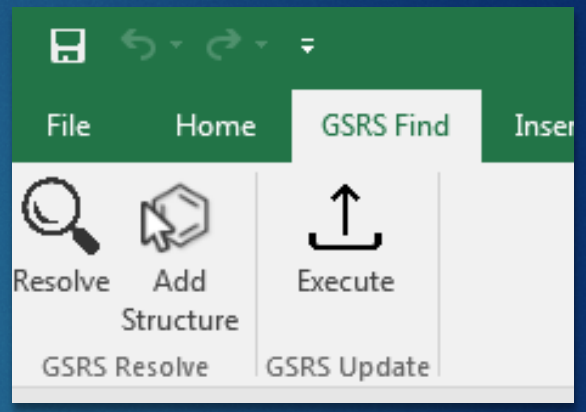
Data loading tasks (examples)

- Add names to existing substances
- Add codes to existing substances
- Create relationships between substances
- Substitute codes
  - Changed code values
  - Changed URLs
- Replace full JSON for an existing substance
- *Create brand new substances*
  - Under construction

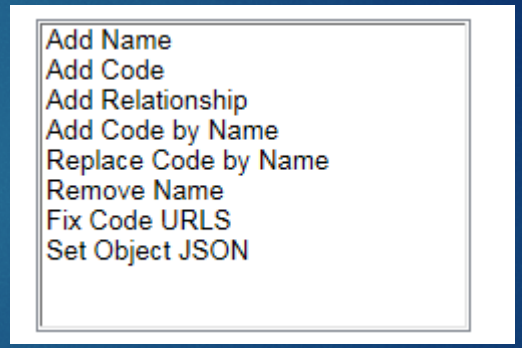
Basics provided by the code

- *These are the base operations provided within the JavaScript and used by Script developers to provide functionality for user*
- Locate a substance by name, UUID, or CAS
- Manipulate top-level properties
- Iterate through its collections
  - Names
  - Codes
  - Relationships
  - Properties
  - References
  - ...
- Add/remove items from collections

Setting up a sheet for data loading (one easy way)



- Select a blank sheet
- Click the GSRS Find ‘Execute’ button
- Select an operation from the list
- Click the ‘Add Batch Sheet’ button



You now have a new sheet with the columns needed to enter data for upload

Processing your data sheet

- Fill in data in the appropriate columns
- Select one or more rows
- Click the (same) Execute button
- Data are loaded and a status column is updated to let you know if the load was successful

Limitations

- Number of records that can be processed at one time: 1-2k
  - Subtle server errors begin to occur

Resolver tasks

- Retrieve data from Ginas into an existing spreadsheet to match existing substance
  - For example, fetch structures to match names

Acknowledgements

- Tyler Peryea for
  - supplying version 1 of the spreadsheet tools
  - Answering a lot of questions
- Frank Switzer, Larry Calllahan
  - Defining the tasks
  - Testing the data

Thank you for your attention!

- **Mitch Miller**
- Scientific Thinking, LLC
- +1 802 242 9017
- mitch.miller@thinkscience.us
- www.thinkscience.us