

C

A

P

U

I



WorldFAIR

# Digital recipes for managing chemical data (Cookbook)

WorldFAIR Chemistry  
Stuart Chalk

ACS National Meeting, 27 March 2023



## Digital Recipes for Managing Chemical Data (Cookbook)

WorldFAIR Chemistry

ACS National Meeting, 27 March 2023



### Outline:

1. **Scope of the Project**
2. **Considerations for Content**
3. **Sample Content**
4. **Technical Deployment**
5. **Example Recipes**
6. **Creating a Community Submission Workflow**
7. **Give Us Feedback!**
8. **Contacts**

*Project Members: Ann-Christin Andres;  
Simon Coles; Jordi Cuadros; Sonja  
Herres-Pawlis; John Jolliffe; Sunghwan  
Kim; Nicola Knight; Kenneth Kroenlein;  
Ye Li; Vin Scalfani*

# Use Case: How can I reuse this set of $pK_a/pK_b$ data?

- Curated IUPAC dissociation constant data digitized and made FAIR  
<https://github.com/IUPAC/Dissociation-Constants>
- 21,000+ values, 8800+ unique chemicals
- Not practical to work with as a set because of the scale -> automate
- Chemist: Digitally what can I do with this set?
  - Ingest the data into Excel or a relational database (e.g. MySQL)
  - Search, sort, filter, clean/normalize data
  - Convert SMILES strings into other identifiers, IUPAC name
  - Get the DOI's of the source documents
  - Aggregate data about each compound and calculate the mean and uncertainty
- Data Scientist: How do I understand this set?
  - What is a SMILES? What are all these other identifiers? Why are there many?
  - What is  $pK_a/pK_b$ ? Why are there other properties? How do I model this data?

# WorldFAIR Chemistry D3.2: Training package

## How-to “cookbook”

### Hands-on routines, demonstrations, templates covering use of standards in:

- Preparing and depositing FAIR machine-enabled chemical data
- Identifying and extracting chemically relevant metadata
- Compiling and validating chemical data

### Objectives

- Smooth adoption of best practices (make it easier/lower barriers)
- Get practical tools and tips in the hands of practicing chemists
- Reusable education resources for working with FAIR chemical data

### Target audiences

- Information tool/service providers supporting FAIR data practices
- Chemists and others reusing machine readable data for data science
- Chemists and others preparing machine-readable chemical data for sharing

# Considerations for Content

## Demonstration of chemistry data standards

### Show why standards are important and how to use them

- Digital standards - InChI, JCAMP-DX, ThermoML, IUPAC Nomenclature
- Highlight online resources that enable IUPAC/community Digital Standards
- Exemplify reuse of open chemical datasets - IUPAC Dissociation Constants [1]

### Broadly Accessible, Easy-to-use Platform

- Use open software and tools - Python, JupyterBook [2], Jupyter Notebooks [3]
- Leverage online resources/systems - GitHub [4], Binder [5], Colab [6]
- Annotate content to enable reuse in different contexts

### Enable FAIR Data Principles

- Create content to exemplify what FAIR means for Chemistry
- Deploy content in a FAIR way for programmatic reuse
- Design a FAIR workflow for creation of content - with open review

[1] <https://github.com/IUPAC/Dissociation-Constants>

[2] <https://jupyterbook.org>

[3] <https://jupyter.org>

[4] <https://github.com>

[5] <https://mybinder.org>

[6] <https://colab.research.google.com>

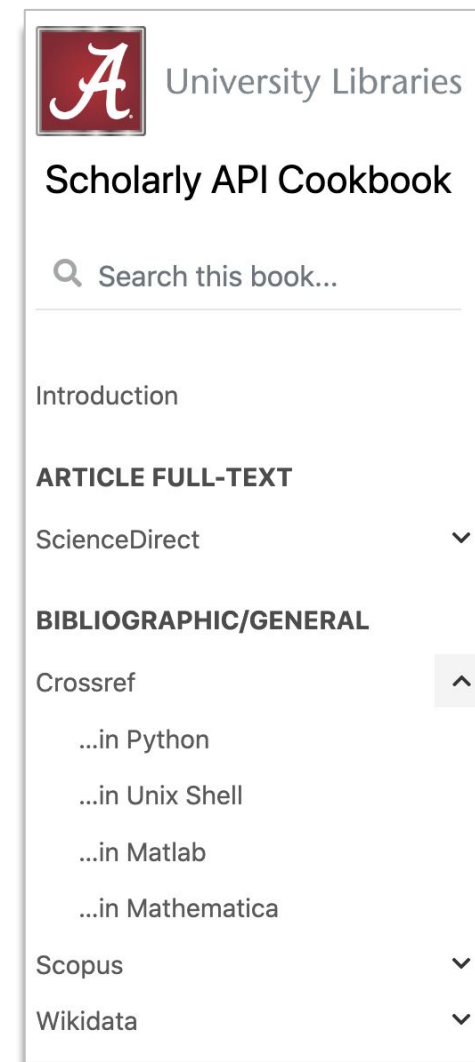
# Chemistry Cookbook Sample Content

## Development of a Cookbook of FAIR Recipes for Chemistry

- Inspired by:
  - University of Alabama: The Scholarly API Cookbook (right) [7]
  - Elixir Europe: FAIR Cookbook [8]
- Recipes: How to apply/implement/enable FAIR for Chemistry
- Built using the Python Jupyter Book package

## Current recipes are a sample of what will make up the cookbook


- Jupyter Notebooks used to make the recipes 'interactive'
- Stepwise explanation of Python code - enabling FAIR
- Introductory metadata that characterises the content



[7] [https://ualibweb.github.io/UALIB\\_ScholarlyAPI\\_Cookbook](https://ualibweb.github.io/UALIB_ScholarlyAPI_Cookbook)

[8] <https://faircookbook.elixir-europe.org>

# Example Recipe #1: Computing InChIs



IUPAC FAIR Chemistry Cookbook

Using the Cookbook

**SAMPLE RECIPES**

We need your feedback!

Accessing the IUPAC Gold Book API in Python

**Computing InChIs**

FAIR formatting of (multi, hyper)spectral data

Finding a Findable Dataset

Preparing Chemical Data for FAIR Sharing - A Minimalist Approach

PubChem PUG-REST API

Structured UV/Vis (meta)data for small molecule characterisation

Using the OPSIN API in Python

**CULINARY SCHOOL**

FAIR basics for Chemistry

## Computing InChIs

About this interactive recipe

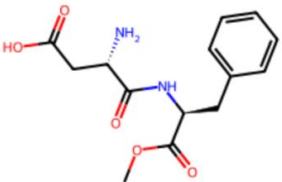
### 1. Using RDKit

#### 1.1 Import RDKit Modules

```
from rdkit import Chem
from rdkit.Chem import Draw
```

#### 1.2 Create a Molecular Object from SMILES

```
# PubChem CID: 134601
m = Chem.MolFromSmiles('COC(=O)[C@H](CC1=CC=CC=C1)NC(=O)[C@H](CC(=O)O)N')
m # to show image of molecule
```



```
# Internally, we have created an RDKit molecular object
print(m)
```

```
<rdkit.Chem.rdchem.Mol object at 0x1247d04a0>
```

Contents

- Using RDKit
- Using Open Babel

# Example Recipe #2: IUPAC Gold Book Definitions

The screenshot shows the IUPAC FAIR Chemistry Cookbook website. The left sidebar contains the IUPAC logo, the text 'WorldFAIR IUPAC FAIR Chemistry Cookbook', a search bar, and a list of sample recipes including 'Accessing the IUPAC Gold Book API in Python'. The main content area is titled 'Accessing the IUPAC Gold Book API in Python' and includes a 'recipe' icon. Below the title is a list of details: Author (Stuart Chalk), Reviewer, Topics (The IUPAC Gold Book, APIs, JSON), Format (Interactive Jupyter Notebook (Python)), Scenarios (Retrieve the definition of a chemical concept via code), and Skills (Application Programming Interfaces (APIs), The JavaScript Object Notation (JSON) file format, Introductory Python, [Regular expressions]). Learning outcomes include understanding Python functions, writing code to request data from a URL, and using variables to call APIs. The citation states that citations for recipes will be available in the full release. The recipe is licensed under CC-BY-4.0. The right sidebar shows a 'Contents' menu with six steps: 1. Import needed Python packages, 2. Add a Python function, 3. Download a JSON file, 4. Search for a term, 5. Use the term code to retrieve its definition, and 6. Try other terms.

**Accessing the IUPAC Gold Book API in Python**

About this interactive recipe

- Author: [Stuart Chalk](#)
- Reviewer:
- Topics: The IUPAC Gold Book, APIs, JSON
- Format: Interactive Jupyter Notebook (Python)
- Scenarios: Retrieve the definition of a chemical concept via code
- Skills: You should be familiar with
  - [Application Programming Interfaces \(APIs\)](#)
  - [The JavaScript Object Notation \(JSON\) file format](#)
  - [Introductory Python](#)
  - [Regular expressions]
- Learning outcomes: After completing this example you should understand:
  - What is a Python function (def)
  - How to write Python code to request data from a URL (typically an API)
  - How to use a Python variable to call an API and download data
- Citation: We will not be providing citations for the sampler, but citations for recipes will be available (with copy button) in the full release of the Cookbook and will include DOI's.
- Reuse: This notebook is made available under a [CC-BY-4.0](#) license.

## Step 1: Import needed Python packages

Python has a lot of functionality that can be imported using the 'import' function

<https://iupac.github.io/WFChemCookbook/samples/goldbook.html>



# Example Recipe #3: Structured UV/Vis Metadata

The screenshot shows the IUPAC WorldFAIR Cookbook interface. On the left is a navigation sidebar with the IUPAC logo and a search bar. The main content area displays the title 'Structured UV/Vis (meta)data for small molecule characterisation' and a dropdown menu for 'About this interactive recipe'. The page is divided into three sections: '1. Objectives', '2. Target audience', and '3. Structuring the (meta)data'. A right-hand sidebar contains a 'Contents' table of contents.

**IUPAC**  
WorldFAIR  
IUPAC FAIR Chemistry Cookbook

Search this book...

Using the Cookbook

**SAMPLE RECIPES**

We need your feedback!

Accessing the IUPAC Gold Book API in Python

Computing InChIs

FAIR formatting of (multi, hyper) spectral data

Finding a Findable Dataset

Preparing Chemical Data for FAIR Sharing - A Minimalist Approach

PubChem PUG-REST API

**Structured UV/Vis (meta)data for small molecule characterisation**

Using the OPSIN API in Python

**CULINARY SCHOOL**

FAIR basics for Chemistry

## Structured UV/Vis (meta)data for small molecule characterisation

About this interactive recipe

### 1. Objectives

To provide guidance on the minimum set of metadata required for publishing UV/Vis spectroscopic data as supplementary information to support compound characterisation using the Beer-Lambert Law. To apply the FAIR principles for data management to produce digital data objects that contain raw, spectral and metadata to enable current and future researchers in the reanalysis and visualisation of UV/Vis spectroscopic data.

### 2. Target audience

Researchers who routinely work with a user interface (spreadsheet) for preparation of supplementary material to support compound characterisation for publication: authors who aim to improve the FAIRness of their published supporting data.

FAIR data management refers to the [IUPAC FAIRSpec standard](#): IUPAC specifications for the findability, accessibility, interoperability and reusability of spectroscopic data in chemistry.

### 3. Structuring the (meta)data

#### 01 File format

- 1. Objectives
- 2. Target audience
- 3. Structuring the (meta)data
- 4. Data reporting using Jupyter Notebook (Python)
- 5. Glossary of Python terms

# Technical Deployment

- Website
  - Layout
  - Table of contents
  - Links to repo/issue
- Repository
  - Version control
  - Issues
  - Pull requests
  - Discussions/Wiki
- Coding environment
  - PyCharm (shown)
  - Visual Studio Code
  - Python IDLE
  - Many more...

The screenshot shows the PyCharm IDE interface for the WFChefCookbook repository. The left sidebar displays the project structure, including folders like .github, book, \_build, howtos, images, recipes, repositories, resources, samples, and static, along with various files such as \_config.yml, \_toc.yml, \_tocold.yml, about.md, bibliography.md, cookware.md, glossary.md, howtos.md, ingredients.md, intro.md, kitchen.md, logo.png, markdown.md, markdown-notebooks.md, menus.md, pantry.md, recipes.md, references.bib, sampler.md, school.md, .gitignore, CONTRIBUTING.md, LICENSE, LICENSE-CODE, Pipfile, and README.md. The main editor window displays the content of the intro.md file, which includes a heading "# Using the Cookbook", a paragraph describing the IUPAC FAIR Chemistry Cookbook, a link to "About this project", a heading "## Recipes", and a paragraph explaining the use of Jupyter Notebooks. The bottom status bar shows the current file is "intro.md" and the Python interpreter is "Pipenv (WFChefCookbook) [Python 3.10.10]".

```
WFChefCookbook - intro.md
WFChefCookbook > book > intro.md
Project
  WFChefCookbook ~/PycharmProjects/WFChefCookbook
  > .github
  > book
  > _build
  > howtos
  > images
  > recipes
  > repositories
  > resources
  > samples
  > static
  _config.yml
  _toc.yml
  _tocold.yml
  about.md
  bibliography.md
  cookware.md
  glossary.md
  howtos.md
  ingredients.md
  intro.md
  kitchen.md
  logo.png
  markdown.md
  markdown-notebooks.md
  menus.md
  pantry.md
  recipes.md
  references.bib
  sampler.md
  school.md
  .gitignore
  CONTRIBUTING.md
  LICENSE
  LICENSE-CODE
  Pipfile
  README.md
  External Libraries
  Scratches and Consoles
Commit
Pull Requests
Database
SolView
Notifications
1 # Using the Cookbook
2
3 The IUPAC FAIR Chemistry Cookbook is intended to be an open, collaborative, community focused resource on working with
4 FAIR data in chemistry. FAIR data are findable, accessible, interoperable, and reusable for machine processing
5 {cite:p}`Wilkinson2016`, the new paradigm for reporting research data. This cookbook aims to support best practices
6 using chemical data standards through practical, interactive tutorials, provided by the International Union of Pure
7 and Applied Chemistry (IUPAC) through the EU funded WorldFAIR project
8 (see \[About this project\]\(https://iupac.github.io/WFChefCookbook/about.html\)).
9
10 ## Recipes
11 - Many of the recipes on this site will take advantage of Jupyter Notebooks to run Python code
12 in the browser for an interactive (and educational) feel for the user.
13 - We have tried to provide useful information about each recipe in the collapsible 'header'
14 that is below the title of the recipe. We are hoping this provides users with some sense
15 of the how, what and when a recipe might be useful.
16 - As part of the header we have included bullets for skills and learning objectives but have also
17 thought to add use cases. If you have any ideas about how we should characterize recipes we would love to
18 hear them!
19
20 ## Contributing to the Cookbook
21
22 We welcome \[feedback & suggestions\]\(https://github.com/IUPAC/WFChefCookbook/issues/new/choose\) for additional content.
23
24 *WorldFAIR "Global cooperation on FAIR data policy and practice" is funded by the EC HORIZON-WIDERA-2021-ERA-01-41
25 Coordination and Support Action under Grant Agreement No. 101058393.*
26
```

# Expanding the Cookbook Content

## Ideas for other content areas

- *Culinary School* - FAIR Basics for Chemistry
- *Ingredients* - Data standards and formats for Chemistry
- *Test Kitchen* - Checking your chemical data/metadata
- *Pantry* - Finding and sharing FAIR chemical data
- *Cookware* - Finding and sharing FAIR chemical data; Use cases for FAIR chemical data

# Community Submission Workflow

## After initial development by the project we will invite community contributions

- Develop submission guidelines
- Develop a submission process using GitHub repository forking and pull-requests
- Outline an open review process



## Digital Recipes for Managing Chemical Data (Cookbook)

WorldFAIR Chemistry



ACS National Meeting, 27 March 2023

# Discussion

FAIR Chemistry  
Training Cookbook

Shared notes link:

<https://bit.ly/ACSChemNotes>



<https://bit.ly/CbookFeedback>



[bit.ly/CookFAIR](https://bit.ly/CookFAIR)



# Digital Recipes for Managing Chemical Data (Cookbook) WorldFAIR Chemistry

ACS National Meeting, 27 March 2023



## Contact & Follow US



[iupac.org](http://iupac.org)



@FAIRChemistry  
@iupac



[tinyurl.com/iupacchan](https://tinyurl.com/iupacchan)



[FAIRChemistry@iupac.org](mailto:FAIRChemistry@iupac.org)



@iupac.org



FAIRChemistry Community