

GSRS Development Update

16 Nov 2018

TYLER PERYEA
DANNY KATZEL
NIH/NCATS

NCATS

GSRS Status

High Level GSRS Core Status

- GSRS 2.3.1 in production at FDA (October 2018)
 - 2.3.1 public released
- GSRS 2.3.2 development (November 2018)
 - Pre-production at FDA
 - Pending public release soon (December 2018)
- GSRS 2.4 development (1st quarter 2019)
 - In development
 - Improved data exchange / communication

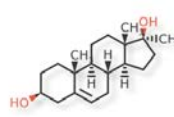

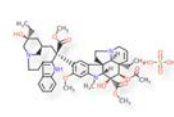
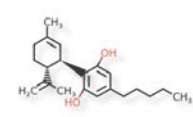

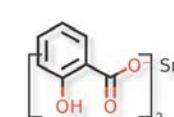
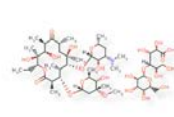
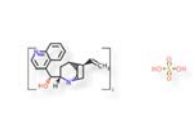
GSRS Status

▼ Substance Type

Search Substance Class...

- ☐ Chemical 64,971
- ☐ Structurally Diverse 24,833
- ☐ Concept 5,461
- ☐ Mixture 2,417
- ☐ Protein 2,146
- ☐ Polymer 2050
- ☐ Nucleic Acid 134

Show Less

<p>5730Z864KG</p> <p>ABSOLUTE</p>  <p>METHANDRIOL</p> <p>↗ Q</p>	<p>733QK35BCI</p> <p>ACHIRAL</p>  <p>GLYCERYL 1,3-DISTEA...</p> <p>↗ Q</p>	<p>N00W22YO2B</p> <p>ABSOLUTE</p>  <p>VINBLASTINE SULFATE</p> <p>↗ Q</p>	<p>19GBJ60SN5</p> <p>ABSOLUTE</p>  <p>CANNABIDIOL</p> <p>↗ Q</p>
<p>1LM0WK173J</p> <p>ACHIRAL</p>  <p>COPPER UNDECYLEN...</p> <p>↗ Q</p>	<p>OGF9EV99TC</p> <p>ACHIRAL</p>  <p>STRONTIUM SALICYLATE</p> <p>↗ Q</p>	<p>33H58I7GLQ</p> <p>ABSOLUTE</p>  <p>ERYTHROMYCIN LACT...</p> <p>↗ Q</p>	<p>5JT77A1M4W</p> <p>ABSOLUTE</p>  <p>CINCHONIDINE SULFATE</p> <p>↗ Q</p>

GSRS 2.3.1 Status

- **GSRS 2.3.1 Improvements**

- Substance “Hierarchy” views
- Additional Validation Rules
- Improved Relationship Handling
- Improved Structure Resolver
- Configurable / modular Validation Engine
- Porting of some FDA-specific enhancements to main code base
- Bug fixes
- Improved UI
- New Guided Search

GSRS 2.3.1 New Search Guide

Search ...

[Show help](#) [Hide guide](#)

I would like to search

in

every field of the record

for the following exact phrase, which must match completely (no partial words)

and that's it

Run query

☐ Concept

10

☐ Mixture

10

☐ Polymer

10

[More ...](#)

▼ Source Tag

Search GHSAS Tag...

☐ INN

36

EPOETIN BETA

CONCEPT



Names: EPOETIN BETA
1-165-ERYTHROPOIETIN (BM 06.019)
EPOCH
EPOETIN BETA RDNA

Codes: CAS: [122312-54-3](#)
WHO-ATC: [B03XA01](#)
EVMPD: [SUB06576MIG](#)

GSRS 2.3.1 New Search Guide

I would like to search

in

✓ every field of the record

Any Name - Any name (official, common, systematic, etc) found in a record

Any Code - Any code (database link, classification code, etc) found in a record

Approval ID (UNII) - The approval ID (UNII) of a record

Molecular Weight - The molecular weight of the substance

Last Edited Date - The date that the record was last edited

another field ...



GSRS 2.3.1 New Search Guide

Any Name - Any name (official, common, systematic, etc) found in a record

- ✓ for the following exact phrase, which must match completely (no partial words)
- for the following contained phrase, which must be found as written (no partial words)
- for ANY of the following words in any order or position
- for ALL of the following words in any order or position
- for a WORD that starts with
- for a WORD that ends with
- for a WORD that contains
- for a value that starts with with the word(s)
- for a value that ends with the word(s)



Concept

10



Mixture

10

EPOETIN BETA



GSRS 2.3.1 New Search Guide

Any Name Any Name (official, common, systematic, etc.) found in a record

- ✓ for the following exact phrase, which must match completely (no partial words)
- for the following contained phrase, which must be found as written (no partial words)
- for ANY of the following words in any order or position
- for ALL of the following words in any order or position
- for a WORD that starts with
- for a WORD that ends with
- for a WORD that contains
- for a value that starts with with the word(s)
- for a value that ends with the word(s)

Required

☐ Concept

10



GSRS 2.3.1 New Search Guide

[Show help](#) [Hide guide](#)

I would like to search

in

every field of the record

for the following exact phrase, which must match completely (no partial words)

aspirin|

and that's it

Run query



GSRS 2.3.1 New Search Guide

✓ and that's it

also make sure to search
if the above is not found, search

Run query

also make sure to search

▼

in

every field of the record

for the following exact phrase, which must match completely (no partial words)

▼

and that's it

▼

Run query

EPOCH
EPOETIN B

GSRS 2.3.1 New Search Guide

root_names_name:"^aspirin\$" OR "^foo\$"

[Show help](#) [Hide guide](#)

I would like to search

in

Any Name - Any name (official, common, systematic, etc) found in a record

for the following exact phrase, which must match completely (no partial words)

aspirin

if the above is not found, search

in

every field of the record

for the following exact phrase, which must match completely (no partial words)

foo

and that's it

Run query

▼ Source Tag

Search GlnAS Tag

INN

38



Codes:

EPOCH

EPOETIN BETA RDNA

CAS: 122312-54-3

WHO-ATC: B03XA01

EVMPD: SUB06576MIG

EMA ASSESSMENT RE

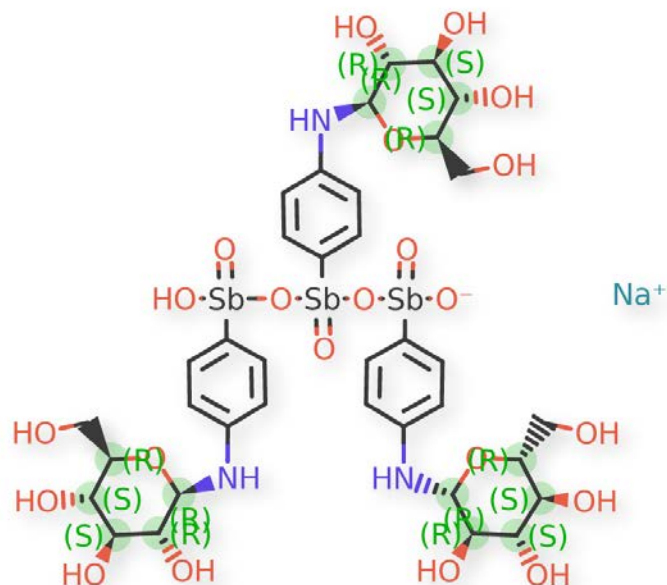
GSRS 2.3.2 Status

- **GSRS 2.3.2 Improvements**

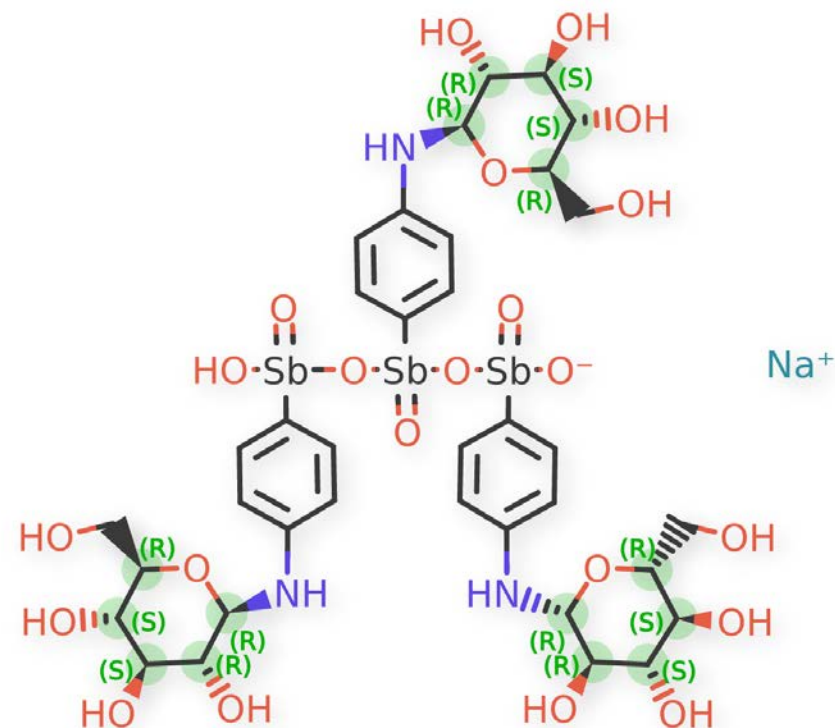
- Improvements to rendering of molecules
- Further improvements to guided search
- Further improvements to duplicate check / sequence searching
- Ability to easily revert records for authorized users
- Additional stability and data integrity checks
- Improvements to hierarchy view

GSRS 2.3.2: Rendering

GSRS 2.3.1 (old)



GSRS 2.3.2 (new)



Systematic Names

SMILES

InChi

SMILES:

```
[Na+] . OC[C@H]1O[C@@H](NC2=CC=C(C=C2)[Sb](O)(=O)O[Sb](=O)(O[Sb]([O-])
(=O)C3=CC=C(N[C@@H]4O[C@@H](CO)[C@@H](O)[C@H](O)
[C@H]4O)C=C3)C5=CC=C(N[C@@H]6O[C@@H](CO)[C@@H](O)[C@H](O)[C@H]6O)C=C5)[C@H]
(O)[C@@H](O)[C@@H]1O
```

GSRS 2.3.2: Guided Searches

GSRS 2.3.1 (old)

(a little opaque and inaccurate)

Searched for:
root_structure_mwt :
[100 TO 10E50] AND
aspirin (contains)

GSRS 2.3.2 (new)

(more informative and accurate)

You searched for the **Molecular Weight** field > **100** AND for any field = ...**aspirin**... [show long explanation](#)

You searched in the **Molecular Weight** field, for a value or values that must greater than **100**. Also making sure to search in any field, for a value or values that contains **aspirin** [show short explanation](#)

GSRS 2.3.2: Restoring Old Versions

GSRS 2.3.1 (old)

History				
Version	Change Reason	Editor	Change Date	
1	Changed stereochemistry	admin	Tue Mar 06 17:15:50 EST 2018	View

Super users can **view** old versions, but involved process to revert

GSRS 2.3.2 (new)

History				
Version	Change Reason	Editor	Change Date	
View 1	Changed Preferred Name to "STIBAMINE GLUCOSIDE"	Tyler Peryea	Thu Nov 15 10:10:26 EST 2018	Restore

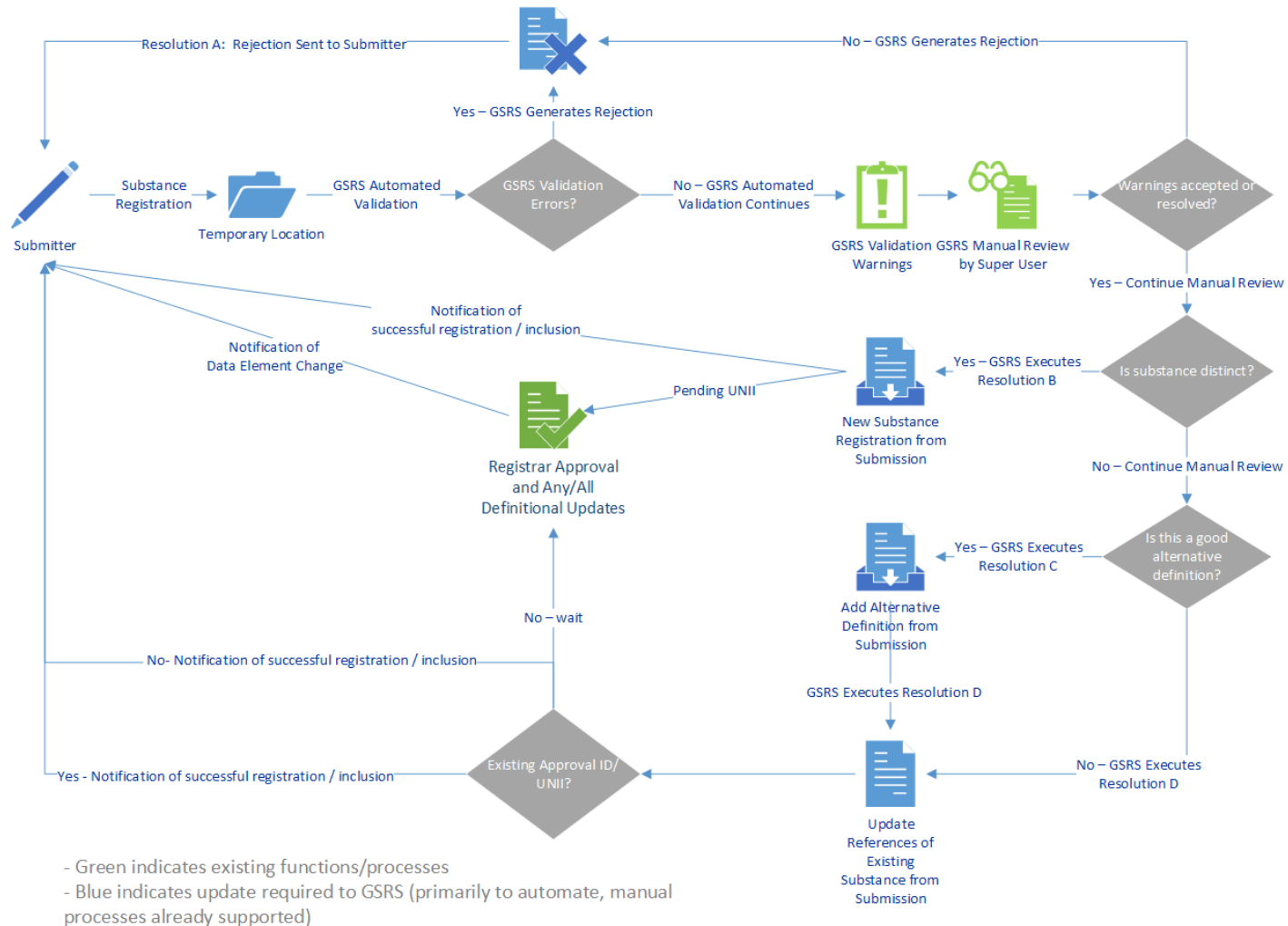
Super users can quickly **revert** records, after review

GSRS Status

On the horizon

- GSRS 2.4
 - Submissions / System-to-system communication
 - Establishment of requirements document (*complete*)
 - Basic support for accepting / staging / processing new and modified records from external sources (*1st quarter 2019*)
 - Simplified installation / launch
 - High-level feasibility report (*complete*)
 - Basic support (*1st quarter 2019*)

System Submission and Exchange Mechanism



GSRS Status

On the horizon

- GSRS 2.5+
 - Improvements to REST API / Systematic access
 - Upgrade of UI to modern web framework for client-side UI
 - Angular 6+
 - Websocket support

GSRS Tools

- **GSRS Find excel tool** (*currently in alpha release*)
 - Search, create, update, retrieve records and substance information in batches
 - Support for loading SD files
- **Stand-alone Java API library** (*1st quarter 2019*)
 - Easier programmatic access to data for integration with other applications / stand-alone tools
- **Stand-alone Javascript API library** (*2nd quarter 2019*)
 - Provides simple communication and data operations with running instance

GSRS Tools: GSRS Find

The screenshot displays the GSRS Find tool interface. On the left, an Excel spreadsheet titled 'smalldataset.xlsx' is open, showing a table of substances. A red arrow labeled '1. Input names/codes' points to the 'METHANDRIOL' entry in the first column. On the right, the 'Data Retriever' dialog box is open, showing various options for retrieving data. A red arrow labeled '2. Choose what you want' points to the 'Molecular Formula' checkbox, which is checked. A red arrow labeled '3. Results' points to the results table in the Excel spreadsheet.

	A	B	C	D
1				
2	METHANDRIOL	C[C@]1(O C20H32O2	304.4669	
3	GLYCERIN 1,3-DISTEARATE	CCCCCCCC C39H76O5	625.0177	
4	P-T-BUTYL-M-CRESOL	CC1=CC(O C11H16O	164.2441	
5	METOCURINE CHLORIDE	[Cl-].[Cl-].C40H48N2	723.725	
6	DOMINE	CCCCCCCC C17H38N2	270.497	
7	SELENOASPIRINE	CC(=O)[Se C9H8O3Se	243.12	
8	ASPIRIN POTASSIUM	[K+].CC(=C9H7O4.K	218.2478	
9	ZENIPLATIN	[Pt++].NCI C6H6O4.C	471.37	

1. Input names/codes

2. Choose what you want

3. Results

Data Retriever

Substance

- ☐ Active Moiety PT
- ☐ All Names
- ☐ Bracket Terms
- ☐ Active Moiety ID
- ☐ ATC Code
- ☐ Substance Class
- ☐ Preferred Term

Chemical

- ☒ SMILES
- ☐ Full Lychi
- ☒ Molecular Weight
- ☐ InChIKey
- ☒ Molecular Formula
- ☐ Equivalence Factor
- ☐ Lychi L1
- ☐ Molfile
- ☐ Stereo Type

Protein

- ☐ Protein Sequence

Identifiers

Processing complete!

☒ Save diagnostic info when finished? ☐ Resolve to new sheet?

GSRS Tools: Stand-alone Java API Library

```
try(GsrsRestClient client = DefaultGsrsRestClient.create("https://ginas.ncats.nih.gov/ginas/app")){
```

```
    GsrsSearchResult result = client.searcher().substructure("C1=CC2=CC=CC=C2C=C1");
```

```
        System.out.println("search results :");
        System.out.println("total : " + result.getTotal());
        System.out.println("top 10 substance names:");
```

```
        result.getUids()
            .stream()
                .limit(10)
                .forEach(uuid->{
try {
    Substance sub = client.fetchSubstance(uuid);

    System.out.println(sub.getName());
} catch (IOException e) {};
```

```
}
```

```
search results :
total : 1973
top 10 substance names:
NAPHTHALENE
POLYCYCLIC AROMATIC
HYDROCARBONS
PHENANTHRENE
ANTHRACENE
BENZ(A)ANTHRACENE
TRIPHENYLENE
CHRYSENE
BENZO(C)PHENANTHRENE
TETRACENE
DIBENZ(A,H)ANTHRACENE
```

Acknowledgements

NCATS

Tyler Peryea
Noel Southall
Dac-Trung
Nguyen
Ivan Grishagin
Dammika
Amugoda
Jorge Neyra
Niko Anderson
Mark Williams
Tim Sheils
Chris LeClair
Paul Shinn
Qian Zhu
Christine Colvis
Mark Williams

FDA

Larry Callahan
Frank Switzer
Yulia Borodina
Ramez Ghazzaoui
Elaine Johansen
Ta-Jen Chen
Archana Newatia
Mary-Ann Slack
Alex Welsch
Sarah Stemann
Yoshiyuki Tokiwa

Medicines Evaluation Board

Herman Diederik
Marcel Hoefnagel
Joris Kampmeijer

Ciska Matai
Burt Kroes

US Pharmacopeial Convention

Fouad Atouf
Andrej Wilk
Tina Morris

Questions