

# Global Substance Registration System: consistent scientific descriptions for substances related to health

Tyler Peryea<sup>1,2</sup>, Noel Southall<sup>1b,2,\*</sup>, Mitch Miller<sup>2</sup>, Daniel Katzel<sup>2</sup>, Niko Anderson<sup>2</sup>, Jorge Neyra<sup>2</sup>, Sarah Stemann<sup>2</sup>, Đắc-Trung Nguyễn<sup>1b,2</sup>, Dammika Amugoda<sup>2</sup>, Archana Newatia<sup>1</sup>, Ramez Ghazzaoui<sup>1</sup>, Elaine Johanson<sup>1</sup>, Herman Diederik<sup>3</sup>, Larry Callahan<sup>1</sup> and Frank Switzer<sup>1</sup>

<sup>1</sup>Office of the Commissioner, US Food and Drug Administration, Silver Spring, MD 20993, USA, <sup>2</sup>Informatics, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD 20892, USA and <sup>3</sup>College ter Beoordeling van Geneesmiddelen, 3531 AH Utrecht, Netherlands

Received August 13, 2020; Revised October 05, 2020; Editorial Decision October 06, 2020; Accepted October 08, 2020

## ABSTRACT

The US Food and Drug Administration (FDA) and the National Center for Advancing Translational Sciences (NCATS) have collaborated to publish rigorous scientific descriptions of substances relevant to regulated products. The FDA has adopted the global ISO 11238 data standard for the identification of substances in medicinal products and has populated a database to organize the agency's regulatory submissions and marketed products data. NCATS has worked with FDA to develop the Global Substance Registration System (GSRS) and produce a non-proprietary version of the database for public benefit. In 2019, more than half of all new drugs in clinical development were proteins, nucleic acid therapeutics, polymer products, structurally diverse natural products or cellular therapies. While multiple databases of small molecule chemical structures are available, this resource is unique in its application of regulatory standards for the identification of medicinal substances and its robust support for other substances in addition to small molecules. This public, manually curated dataset provides unique ingredient identifiers (UNIs) and detailed descriptions for over 100 000 substances that are particularly relevant to medicine and translational research. The dataset can be accessed and queried at <https://gsrs.ncats.nih.gov/app/substances>.

## INTRODUCTION

The globalization of the pharmaceutical industry has made global data standardization essential for promoting the safety, availability and quality of health products through-

out the world (1). With increasing globalization, the supply chain spans many countries, raising issues about international standards and safety in regulated products and for their component substances. As fewer drugs are sourced entirely within one jurisdiction, cooperation between international regulatory bodies becomes critical. Interoperability and standardization of data based on international standards can remove the greatest barriers to such international coordination (1).

The ISO 11238 standard provides a stable structure and set of data elements for defining substances in a consistent, scientifically useful manner (2). The United States Food and Drug Administration (FDA) has adopted this standard to enhance the regulatory review of active and inactive substances in submissions and facilitate understanding of the relationships to other substances and products from a quality, safety and drug utilization perspective (3).

The complexity and enormous variety of health products currently marketed poses a significant challenge to systematic identification, yet it is vitally important for the sake of public safety. Effective regulation depends on the ability to answer complex queries that require leveraging data from multiple sources in many formats. As such, the need was for a system capable of representing substance data with rigorous definitions, supporting multiple scientific domains, such as nuclear chemistry, herbal plant varieties, autologous genetically transformed cell therapies, and medical air (4).

The ISO 11238 standard identifies key requirements for a global registration system: (i) the collection of defining properties for substances to enable their unambiguous definition, (ii) the creation of unique substance identifiers to reliably identify and trace the use of medicinal products and the materials within medicinal products and (iii) the centralized generation of unique identifiers and deposition of substance facts to both facilitate sponsor interactions with multiple regulators and harmonization amongst regulating

\*To whom correspondence should be addressed. Tel: +1 301 480 9836; Email: southalln@mail.nih.gov

agencies. Inherent in the ISO 11238 standard is the acknowledgement that existing systematic standards are often too rigid to accommodate all of the substances found in commerce. Market forces constantly drive innovation, pushing the boundaries of science itself and the creation of entirely new classes of products in ways that are difficult to anticipate. The standard is therefore accommodating of new and unusual materials in ways that have traditionally challenged other standards.

The Global Substance Registration System (GSRS) addresses these needs to uniquely identify, register and store substance-related information, consistent with the ISO 11238 standard. GSRS provides a system for the definition and identification of substances within medicinal products or substances used for medicinal purposes, including dietary supplements, foods and cosmetics and their official names across different languages, jurisdictions and domains. The system also captures relationships between substances and references all captured data to a definitive source of information. GSRS references existing nomenclatures but coins terms when necessary for nomenclature consistency. All the software developed is intended to be freely distributable to academic, government and commercial entities. The public reference database of substances is provided at <https://gsrs.ncats.nih.gov/app/substances>.

## MATERIALS AND METHODS

The information system is designed around the 6 types of substances referenced in the ISO 11238 standard: chemicals, nucleic acids, proteins, polymers, structurally diverse, mixtures. Further details of the substance data model and software architecture are provided in Supplementary Data. Each of these substance types, and all relevant data fields present in the ISO standard and its technical implementation guide (5) are included within the system's data model, including official names, common names, brand names, systematic names, company codes and other identifiers such as registry numbers. All names and identifiers are provided with references. References (including links to external sites) are used in the ISO 11238 data model to document evidence for specific aspects of a substance definition and associated data. The authors do not intend GSRS to be a comprehensive cross-referencing index of substance websites. Moreover, inclusion of a reference to an external site does not imply that the substance definition found at that external site is fully consistent with the ISO 11238 definition provided within GSRS. GSRS also has the capability to capture an extensive number of relationships between substances. The latest public data release was on 7 July 2020 (v2.5.1–20200707) and consists of 116 636 substance definitions.

The FDA supports health information technology initiatives by generating unique ingredient identifiers (UNIIs) for substances in drugs, biologics, foods, cosmetics, dietary supplements, tobacco products, and devices. The UNII is a non-proprietary, free to use, unique, unambiguous, non-semantic, alphanumeric identifier based on a substance's defining properties from the ISO 11238 data model. The UNII is permanently associated with a given substance definition and when corrections are made, the UNII re-

mains the same. GSRS is the software that generates UNIIs used in FDA electronic listing as seen in DailyMed website at <https://dailymed.nlm.nih.gov/dailymed/>. It is also used for other regulatory activities throughout product life cycles, encompassing clinical trial phases, product marketing and post-market surveillance. New UNII requests, data issues, or questions can be addressed by contacting FDA-SRS@fda.hhs.gov.

## RESULTS

### GSRS public substance dataset

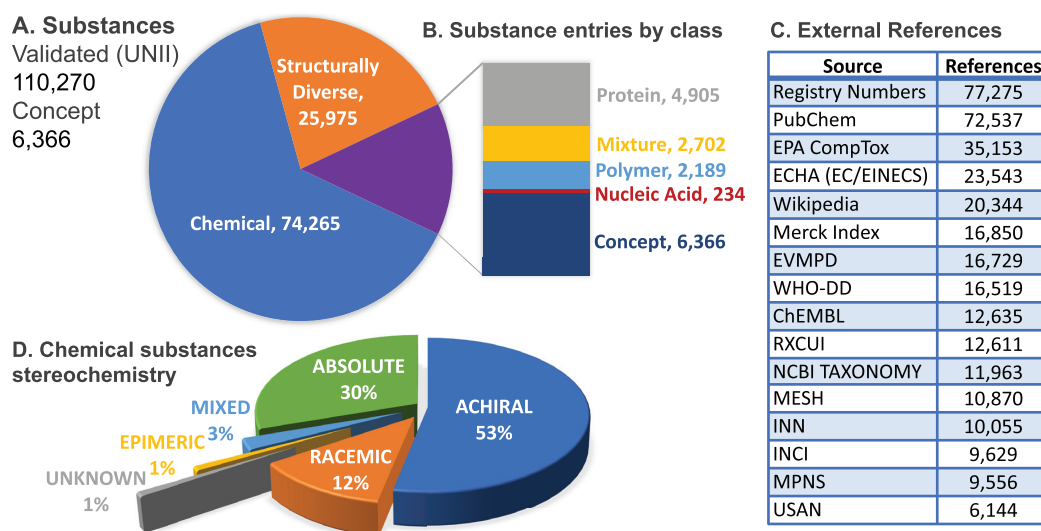
Included in the 7 July 2020 public release are 116 636 substance definitions and accompanying data on nomenclature, properties and relationships between substances (Figure 1). GSRS includes detailed examples of an enormous range of drug substances encountered in drug discovery and development including chemicals, proteins, nucleic acids, polymers, mixtures and structurally diverse substances. The system captures many substance identifiers—names, codes and structural keys to facilitate substance identification. However, none of these on their own are sufficient for regulatory use both because redundant identifiers often exist for the same substance and also because some identifiers are ambiguous and do not differentiate between related substances. We further explore some of the challenges of providing a unique and unambiguous identifier for substances below.

Each substance class uses a different definitional data model, which reflects how the substances are often produced as well as the common types of substance heterogeneity that are encountered. For example, small molecule chemical definitions include a stereochemistry status field because many are marketed as chiral mixtures. Heterogeneity in protein samples typically arises from variations in glycosylation and other chemical modifications made after protein synthesis. Structurally diverse materials are inherently heterogeneous preparations from natural materials. Common sources of variability that can be defining for these substances include the part of an organism from which is was prepared (leaves, roots, etc.) and even the time of harvest.

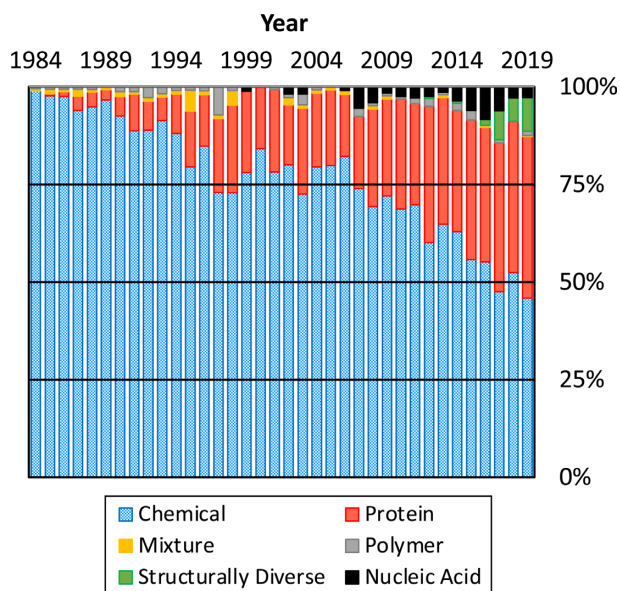
### Trends in product materials

During clinical development, drug sponsors request non-proprietary names for active ingredients from the United States Adopted Name (USAN) and/or International Non-proprietary Name (INN) committees, often disclosing development candidates for the first time in the process of doing so. These public requests are typically made midway through clinical development, and the set of proposed names from a given year provides a useful snapshot of the types of products currently in development. As seen in Figure 2, medicines have historically been dominated by synthetic organic small molecules, but in recent times chemical substances represent a minority of all of the therapeutics in clinical development. Robust support for the registration of non-small molecules is therefore required to support regulatory needs.

## GSRs v2.5.1-20200707 Substance Dataset



**Figure 1.** Overview of information provided in GSRs v2.5.1-20200707 dataset. (A) Total number of ingredient entries. (B) Number of ingredient records provided by substance class. (C) Commonly used public substance information sources referenced by the database. Only a partial list of information sources is provided. (D) For chemical substances in particular, a breakdown of type of stereochemical annotations is provided. The ‘Mixed’ type denotes where more complicated annotations are provided. ‘Unknown’ type denotes where drug substances have chiral specificity, for example demonstrate rotation of light, but absolute stereochemistry has never been assigned experimentally. Abbreviations provided in Supplementary Data.



**Figure 2.** Trends in time for product materials: analysis of INN proposed lists categorized by substance class. In 2019, more than half of all new drugs in clinical development were substances other than small molecule chemicals.

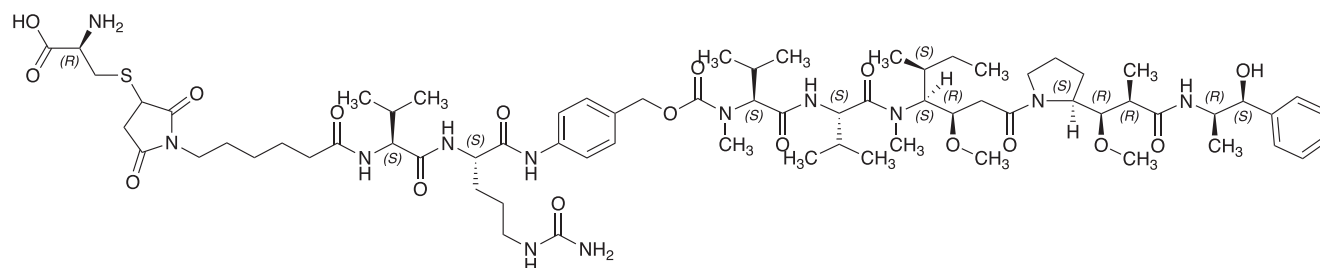
### Challenging substances

*Antibody-drug-conjugates.* Analyzing these substance trends by comparing high-level substances classes obscures some important additional therapeutic innovations such as antibody-drug-conjugates which also must be supported by GSRs. Brentuximab vedotin is a cancer drug that delivers the toxin monomethyl auristatin E to the cancer cell upon

internalization of the antibody by binding to CD-30, which is also known as tumor necrosis factor receptor superfamily member 8. This antibody-drug-conjugate, UNII: 7XL5ISS668, is registered as a protein with structural modifications. In addition to the full protein sequence for its four subunits, details of disulfide links, glycosylation and two structural modifications are provided. The first modification indicates the tendency of N-terminal glutamic acids to form the lactam pidolic acid, which is commonly seen in many proteins. The second indicates the partial conjugation of a toxin-linker moiety to available cysteines on the protein. Instead of specifying the reactants, the substance definition registers the replacement of protein cysteines with the product of the reaction, whose full details are provided in UNII: 6603L01WUR (Figure 3).

*Vaccines.* One notable product class excluded from the INN list analysis in Figure 2 but which are supported by the software are vaccines. Live vaccines are registered as structurally diverse substances. One example is a recently developed ebola virus vaccine candidate (6). This vaccine is registered as UNII: Y9VG7O3KTT. The vaccine substance is a live, attenuated, genetically-modified vesicular stomatitis Indiana virus (rVSV), engineered to express Zaire ebolavirus strain Kikwit-95 envelope glycoprotein (ZEBOV-GP). Multiple copies of glycoprotein are expressed and assembled into the viral envelope responsible for inducing protective immunity. The chimeric virus vaccine is attenuated by deletion of the principal virulence factor of VSV (the G protein), which also removes the primary target for anti-vector immunity. This is described within the substance definition by including the structural modifications provided in Table 1.





**Figure 3.** Chemical structure of the vedotin conjugate UNII: 6603L01WUR employed in definition of brentuximab vedotin, UNII: 7XL5ISS668. This moiety replaces defined cysteine residues in the protein sequence. It is a cysteine derivative with a maleimide-caproic acid attachment group, cathepsin cleavable linker (valine-citrulline), and para-aminobenzylcarbamate spacer attached to the toxin monomethyl auristatin E. Atoms and bonds are depicted as stated in Annex B of the ISO 11238 standard.

**Table 1.** Structural modifications defining rVSVΔG-ZEBOV-GP (UNII: Y9VG7O3KTT), an ebola vaccine

Modification Type	Extent	Modification Name	UNII
Gene expression vector <sup>1</sup>		Vesicular stomatitis Indiana virus	KTI7RPW4I0
Gene deletion	Complete	Vesicular stomatitis Indiana virus vsivgp4 glycoprotein (g protein) precursor	E6TJ0Z0ZE8
Gene fragment replacement <sup>2</sup>	Complete	ebola virus/h. sapiens-tc/cod/1995/Kikwit-9510622 envelope glycoprotein gene (viral negative strand)	P7ZRG1LJ3A
Vector expressed protein <sup>3</sup>	Complete	Zaire ebolavirus strain Kikwit-95 envelope glycoprotein	XH5V2SQ5FI

(1) Parent organism: vesicular stomatitis Indiana virus (UNII: KTI7RPW410). (2) Reference: GenBank: KU182909.1 Ebola virus isolate Ebola virus/H. sapiens-tc/COD/1995/Kikwit-9510622, complete genome. (3) Virus particle incorporation (23) of ebolavirus glycoprotein (UNII: XH5V2SQ5FI) occurs via expression of the gene fragment replacement (UNII: P7ZRG1LJ3A) and further cellular processing including a number of complex idiosyncratic steps (24). The vector expressed protein registered reflects the final heteromultimeric protein product.

**Other cases.** The database also contains records for lesser defined substances using the concept class as well as some further specified substances which are defined in the ISO 11238 standard as group 1 specified substances. An example of a substance concept is fish, unspecified (UNII: 1PIO77PW2X). Such concept terms do not sufficiently define the ingredient used for regulatory purposes as producers must specify the type of fish actually used in a product, but the term may be useful for describing a class effect. For example, ‘fish, unspecified’ is used to refer to fish allergies. On the other hand, it can be equally important to further define aspects of a substance used in a product, using a group 1 specified substance definition. One recently published substance is air polymer type A from ExEm<sup>®</sup> Foam (UNII: WLT3PF2KX0) which is indicated for sonohysterosalpingography to assess fallopian tube patency in women with known or suspected infertility (7). This UNII refers to the foam ingredient created by a specific combination of water, air, glycerin, and hydroxyethyl cellulose (5500 MPA.S AT 2%). Other unique cases of substances can be accommodated within the existing substance model as in the case for the atropisomer BMS-986142 (UNII: PJX9GH268R) (8).

In addition to substance definitions, the database provides many relationships between substances that provide additional biological and manufacturing context. For example, the record for neratinib (UNII: JH94R3PWB) contains 37 relationships with other substances reported in the product New Drug Application including its salt forms, links to a variety of tyrosine kinases it is known to target, cytochrome P450s that it interacts with and transporters.

## Uniqueness and ambiguity of identifiers

**Substance names.** The literature usually refers to a given substance by its name, but names are not always unique and unambiguous. On average, each substance record includes 6 synonyms – including systematic, common, official and code names. *Capsella bursa-pastoris* L. (UNII: W0X9457M59) which is one of the most common weeds in the world (9) and has been the subject of clinical study (10) has an astounding 240 different names included in its record.

The dataset also contains over 1000 examples where the same name can refer to two different substances. For these cases of homographs, one needs additional information or context of use to distinguish which substance the name refers to. Four representative examples are given in Table 2. The first example, alpha-tocopherol, reflects ambivalence by naming authorities in distinguishing between the all *R* isomer of alpha-tocopherol purified from natural sources and the industrially-produced *R,S* mixture often used as a vitamin supplement in foodstuffs. Both substances have separate existences and are captured in the database, along with their absolute stereochemistry, and this shared synonym asserted by different references or sources. The second example reflects differences in naming conventions within the United States and outside of it, where the name must include the explicit hydration state or excludes the hydration state from the name of the predominant form currently marketed. In a similar vein, ‘scientific’ names are often appropriated as shorthand (third example) to refer to a specific part or useful component from a whole organism. Finally, we see the not infrequent occurrence of a word

**Table 2.** Representative examples of homographs from the public dataset

Homograph	UNII	Description	UNII	Description	Case
alpha-tocopherol azithromycin	H4N855PNZ1 5FD1131I7S	Synthetic vitamin E Azithromycin (trihydrate)	N9PR3490H9	Natural extract vitamin E Azithromycin (anhydrous)	Stereochemical ambiguity Implicit versus explicit hydration
lobelia lime	7QFT17RLRG C7X2M0VVNH	Indian tobacco leaf Lime (calcium oxide)	J2KLZ20U1M 9PPI13TC5U 8CZS546954	<i>lobelia inflata</i> L. plant Lime (citrus)	Whole versus part Language ambiguity

in common use having multiple, distinct meanings. Names, unfortunately, are inadequate for the purpose of uniquely identifying ingredients.

**Structure-based identifiers.** Structural identifiers and keys are also often not unique and unambiguous identifiers of therapeutic ingredients. Of the 73 122 chemicals included in the database, 5617 have two or more IUPAC International Chemical Identifier (InChI) keys (11) referring to the same substance and 1244 InChI keys are shared by two or more substances. For example, different tautomer forms of the same compound can produce different InChI key values. In addition, some problematic substances such as a chiral substance of unknown absolute chirality can appear to have the same InChI as the racemic mixture, however such cases are really out of scope for the InChI approach. Chemical Abstracts Services (CAS) (12) and other registry numbers are also widely used to index and inventory chemicals. In the GSRS database release, 4296 substances have reference to more than one registry number. Carnitine chloride (UNII: F64264D63N) has ten. And 503 registry numbers point to multiple substances. This most often occurs when more specific substances refer to a more general concept registry number such as 100403-19-8 which can refer to any of seven different ceramides or the generic registry number 25322-68-3 for polyethylene glycol which is linked to 52 different substances of specific chain length and polydispersity.

## DISCUSSION

The goal of this resource is to benefit public health, translational research and facilitate the transfer of regulatory information into the public domain and provide industry with the means to both obtain a global identifier and deliver information related to substance identification.

GSRS supports the registration of new substances by regulators, providing easy access to existing substance information and a framework to validate information integrity and systematize regulators' expert opinion on what defines a new substance. The system links information and identifiers from different domains and jurisdictions together into a single database. The system also incorporates other data elements of the Identification of Medical Products ISO standard along with biological, chemical and physical data relevant to drug safety, quality and development.

This database is unique in its semantic approach to defining product ingredients and its support for the enormous range of substances encountered in medical products. Regulating the sale of food and medicines and reviewing their health claims requires an integrated review infrastructure, where product ingredients are cross indexed across applica-

tions and information on the safety and efficacy of related substances is easily retrievable.

Historically, FDA has approached this challenge through the development of several different databases and products. The Drug Registration and Listing System (13) was one of the first systems developed to support a specific aspect of manufacturer listing with the agency. Subsequently, the 'Orange Book' (14) and Inactive Ingredient Guide (15) were published, all of which focused on organizing agency information by ingredient name. This was followed by substantial efforts to develop cheminformatics capabilities through the initial development of a substance registration system (16) and eventually expanding in scope to provide a framework proposal for the further development of the ISO 11238 standard and to meet the needs of the Structured Product Labeling (17) standards. The labeling standard requires GSRS-generated UNIIIs as the primary identifier for ingredients in medical products and includes UNIIIs into the product labels of all marketed products regulated by the agency. Adoption of GSRS by other agencies will help to improve international harmonization, pharmacovigilance efforts and understanding of global supply chains by enabling data exchange based on a common standard for product ingredients.

Inherent in the ISO 11238 standard is the acknowledgement that existing systematic standards for organizing substance information are incomplete and will continue to be so. Market forces constantly drive innovation, pushing the boundaries of science itself and the creation of entirely new classes of products. For example, the problem of chemical registration is one of the most mature areas of research and many systems exist, but still provide incomplete support for unusual stereochemistry (18–20), metal-organic structures (21), and metastable isotopes (22). The ISO 11238 standard addresses this problem both by the use of an accommodating data model and use of expert review to enforce consistent use of that data model. While it is desirable to determine uniqueness via automated computational methods, in practice a process built with expert review at its center is necessary to handle the scope and challenges of regulated products. Approaches and tools to address the challenge of discerning uniqueness and removing ambiguity in the systematic definition of substances can be incorporated into this software in the future.

The Global Substance Registration System provides a public, manually curated dataset of the ingredients in medicinal products and their scientific definitions for regulatory and translational research. GSRS is the first database to provide ingredient definitions using the global ISO 11238 data standard for the identification of substances in medicinal products. Especially important is its robust support for

substances other than small molecules and its curation process. Use of this data and the UNII will help to improve international harmonization and pharmacovigilance efforts as well as support knowledge diffusion within the translational research community.

## DATA AVAILABILITY

Software, public domain data and important documentation are available from: <https://gsrs.ncats.nih.gov>. Source code is available on GitHub at: <https://github.com/ncats/gsrs-play>. The software is provided under an Apache 2.0 license. The latest production release of the software is v2.5.1. The latest public data release was on 7 July 2020 (v2.5.1-20200707) and consists of 116 636 substance definitions. In accordance with FAIR data standards, the UNII is the globally unique and persistent identifier can be searched at <https://gsrs.ncats.nih.gov/app/substances> and <https://fdasis.nlm.nih.gov/srs/srs.jsp> and is also included within many other online repositories. Data objects are provided in JavaScript Object Notation (JSON) and substances, e.g. UNII: 5Y3NBK9IS7, can be retrieved through a request to, for example, [https://gsrs.ncats.nih.gov/app/api/v1/substances\(5Y3NBK9IS7\)](https://gsrs.ncats.nih.gov/app/api/v1/substances(5Y3NBK9IS7)). The process of extracting and transforming arbitrary exports can be resource intense and such functionality is not currently accessible from the public site. Local installation of the software allows users to download arbitrary sets of selected records in a variety of formats, including full JSON, TSV, SDF, etc. The database is accessible but not optimized for phone and tablet screens owing to the complexity of the data model and certain features such as chemical structure search; users may prefer to request the desktop version of the site from their mobile browser.

## SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

## ACKNOWLEDGEMENTS

The current project was developed in close collaboration with multiple regulatory authorities to ensure a robust substance registration product that supports the needs of national and regional authorities. In particular, we gratefully acknowledge the early support of Bob Allkin, Christopher Austin, Thomas Balzer, Marcel Hoefnagel, Panagiotis Telonis, Vada Perkins, Mary-Ann Slack and Philipp Weyermann toward this initiative and all of their feedback.

## FUNDING

Intramural research program of the National Center for Advancing Translational Sciences' National Institutes of Health; Office of the Commissioner, US Food and Drug Administration. Funding for open access charge: National Institutes of Health.

*Conflict of interest statement.* None declared.

## REFERENCES

- Gronning, N. (2017) Data management in a regulatory context. *Front. Med. (Lausanne)*, **4**, 114.
- International Standards Organization. (2020) *ISO 11238: 2018. Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances*. <https://www.iso.org/standard/69697.html> (7 October 2020, date last accessed).
- United States Food and Drug Administration. (2020) *Substance Identification*. <https://www.fda.gov/industry/fda-resources-data-standards/substance-identification> (7 October 2020, date last accessed).
- Edwards, P., Therriault, P.A. and Katz, I. (2018) Onsite production of medical air: is purity a problem? *Multidiscip. Respir. Med.*, **13**, 12.
- International Standards Organization. (2020) *ISO/TS 19844: 2018. Implementation guidelines for ISO 11238 for data elements and structures for the unique identification and exchange of regulated information on substances*. <https://www.iso.org/standard/71965.html> (7 October 2020, date last accessed).
- Monath, T.P., Fast, P.E., Modjarrad, K., Clarke, D.K., Martin, B.K., Fusco, J., Nichols, R., Heppner, D.G., Simon, J.K., Dubey, S. *et al.* (2019) rVSVDeltaG-ZEBOV-GP (also designated V920) recombinant vesicular stomatitis virus pseudotyped with Ebola Zaire Glycoprotein: Standardized template with key considerations for a risk/benefit assessment. *Vaccine X*, **1**, 100009.
- US Food and Drug Administration. EXEM FOAM (air polymer-type A) intrauterine foam product label. Available from: [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2019/2122791bl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/2122791bl.pdf).
- Beutner, G., Carrasquillo, R., Geng, P., Hsiao, Y., Huang, E.C., Janey, J., Katipally, K., Kolotuchin, S., La Porte, T., Lee, A. *et al.* (2018) Adventures in Atropisomerism: Total Synthesis of a Complex Active Pharmaceutical Ingredient with Two Chirality Axes. *Org. Lett.*, **20**, 3736–3740.
- Cornille, A., Salcedo, A., Kryvokhyzha, D., Glémin, S., Holm, K., Wright, S.I. and Lascoux, M. (2016) Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Mol. Ecol.*, **25**, 616–629.
- Naafe, M., Kariman, N., Keshavarz, Z., Khademi, N., Mojab, F. and Mohammadbeigi, A. (2018) Effect of hydroalcoholic extracts of capsella bursa-pastoris on heavy menstrual bleeding: a randomized clinical trial. *J. Altern. Complement. Med.*, **24**, 694–700.
- Heller, S.R., McNaught, A., Pletnev, I., Stein, S. and Tchekhovskoi, D. (2015) InChI, the IUPAC international chemical identifier. *J. Cheminform.*, **7**, 23.
- Dittmar, P.G., Stobaugh, R.E. and Watson, C.E. (1976) The chemical abstracts service chemical registry system. I. General Design. *J. Chem. Inf. Comput. Sci.*, **16**, 111–121.
- Slavin, M. (1975) The Food and Drug Administration drug registration and listing system. *Drug Inf. J.*, **9**, 239–240.
- Knoben, J.E., Scott, G.R. and Tonelli, R.J. (1990) An overview of the FDA publication approved drug products with therapeutic equivalence evaluations. *Am. J. Hosp. Pharm.*, **47**, 2696–2700.
- Nema, S., Washkuhn, R.J. and Brendel, R.J. (1997) Excipients and their use in injectable products. *PDA J. Pharm. Sci. Technol.*, **51**, 166–171.
- United States Food and Drug Administration. (2007) *Substance Registration System Standard Operating Procedure*. <https://www.fda.gov/media/75274/download> (7 October 2020, date last accessed).
- Schadow, G. (2005) HL7 Structured Product Labeling - electronic prescribing information for provider order entry decision support. *AMIA Annu. Symp. Proc.*, **2005**, 1108.
- Canfield, P.J., Blake, I.M., Cai, Z.-Li, Luck, I.J., Krausz, E., Kobayashi, R., Reimers, J.R. and Crossley, M.J. (2018) A new fundamental type of conformational isomerism. *Nat. Chem.*, **10**, 615–624.
- Laplante, S.R., Fader, L.D., Fandrick, K.R., Fandrick, D.R., Hucke, O., Kemper, R., Miller, S.P.F. and Edwards, P.J. (2011) Assessing atropisomer axial chirality in drug discovery and development. *J. Med. Chem.*, **54**, 7005–7022.
- Chandrasekhar, J., Dick, R., Veldhuizen, J.V., Koditek, D., Lepist, E.-I., McGrath, M.E., Patel, L., Phillips, G., Sedillo, K., Somoza, J.R. *et al.* (2018) Atropisomerism by Design: Discovery of a selective and stable phosphoinositide 3-Kinase (PI3K) beta inhibitor. *J. Med. Chem.*, **61**, 6858–6868.

21. Jurgens,S., Kuhn,F.E. and Casini,A. (2018) Cyclometalated complexes of platinum and gold with biological properties: state-of-the-art and future perspectives. *Curr. Med. Chem.*, **25**, 437–461.
22. Kharissova,O.V., Méndez-Rojas,MA., Kharisov,BI., Méndez,U.O. and Martínez,P.E. (2014) Metal complexes containing natural and artificial radioactive elements and their applications. *Molecules*, **19**, 10755–10802.
23. Maruyama,J., Miyamoto,H., Kajihara,M., Ogawa,H., Maeda,K., Sakoda,Y., Yoshida,R. and Takada,A. (2014) Characterization of the envelope glycoprotein of a novel filovirus, Ilovu virus. *J. Virol.*, **88**, 99–109.
24. Lee,J.E. and Saphire,E.O. (2009) Ebola virus glycoprotein structure and mechanism of entry. *Future Virol.*, **4**, 621–635.