Towards improved FAIRness of the ThermoML Archive

Demian Riccardi[1], Zachary Trautt[2], Ala Bazyleva[1], Eugene Paulechka[1], Vladimir Diky[1], Joseph W Magee[1], Andrei F Kazakov[1], Scott A Townsend[1], and Chris Muzny[1]

[1]Thermodynamics Research Center, Applied Chemicals and Materials Division, NIST

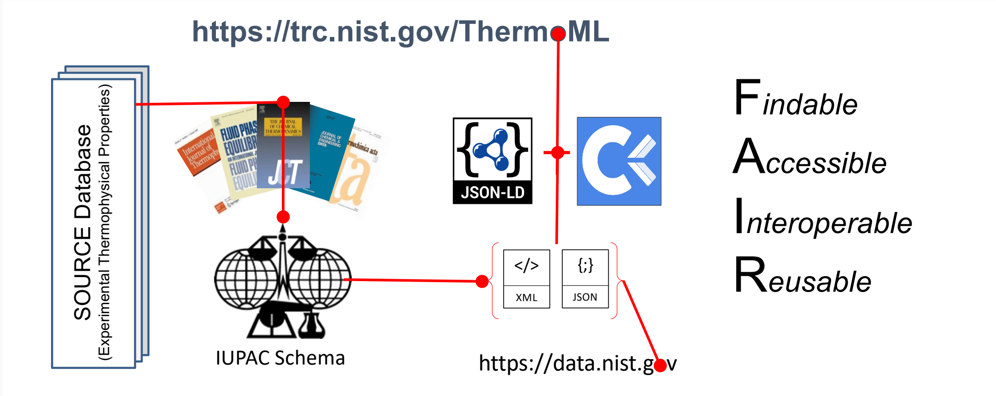
[2]Materials for Energy and Sustainable Development Group, Materials Measurement Science Division, NIST

February 18, 2022

Abstract

The ThermoML archive is a subset of Thermodynamics Research Center (TRC) data holdings corresponding to cooperation between NIST TRC and five journals: Journal of Chemical Engineering and Data (ISSN: 1520-5134), The Journal of Chemical Thermodynamics (ISSN: 1096-3626), Fluid Phase Equilibria (ISSN: 0378-3812), Thermochimica Acta (ISSN: 0040-6031), and International Journal of Thermophysics (ISSN: 1572-9567). Data from initial cooperation (around 2003) through the 2019 calendar year are included. The archive has undergone a major update with the goal of improving the FAIRness and user experience of the data provided by the service. The web application provides comprehensive property browsing and searching capabilities; searching relies on a RESTful API provided by the Cordra software for managing digital objects. JSON files with a schema derived from ThermoML are provided as an additional serialization to lower the barrier to programmatic consumption of the information, for stakeholders who may have a preference of JSON over XML. The ThermoML and JSON files for all available entries can be downloaded from data.nist.gov (https://data.nist.gov/od/id/mds2-2422)

Keywords: Thermophysical Properties, Thermochemical Properties, FAIR Data, ThermoML, JSON, JSON-LD, Schema ■



The NIST Thermodynamics Research Center critically evaluates experimental thermophysical and thermochemical property data. This work uses the ThermoML XSD established by IUPAC projects to transform a large slice (corresponding to five journals that cooperate with NIST TRC during the publication process) of highly-structured database (SOURCE) information into XML and JSON files contained in a web application that makes the resource more Findable, Accessible, Interoperable, and Reusable. All json and xml files are available for download at `https://data.nist.gov/od/id/mds2-2422`.

Introduction

Thermophysical and thermochemical properties provide a measurable framework that describes how molecular systems exist and transfer energy given a set of variable conditions. This information may be transformed, using thermodynamics principles, or extrapolated using models to resolve properties at conditions relevant to industrial processes and academic research. A competently measured value that is carefully reported may be used forever.[1] Such measurements have been carried out at an increasing rate since the 1800s (Fig. 1).

The NIST Thermodynamics Research Center (TRC) provides critically evaluated thermophysical and thermochemical property data complete with provenance and uncertainty. NIST TRC develops software tools to enable the transformation of information from primary sources into digital resources. The conventional scope of experimental data considered includes thermodynamic/thermophysical, phase equilibria, and transport properties of pure, binary, and ternary systems of organic compounds; reactions are also considered but with historically limited scope. TRC has recently expanded the scope of targeted data to include metals, alloys, and intermetallic compound systems,[2] which is not considered in the present work. TRC typically captures 300k-500k data points (k= 10^3) per year and currently has about 8M data points stored in an SQL database (SOURCE).[3] The SOURCE database is consumed by the ThermoData Engine,[4] software that enables users to critically evaluate property data involving all captured experimental data, property predictions, and thermodynamic consistency enforcement. A significant slice of the SOURCE database may be accessed in the form of ThermoML files deposited in the ThermoML archive (https://trc.nist.gov/ThermoML).

ThermoML is an XML-based IUPAC Standard for storage and exchange of thermophysical and thermochemical property data; ThermoML was developed initially within IUPAC Project 2002-055-3-024[5] and later extended under the IUPAC project 2007-039-1-024[6]. The namespace, ThermoML, has been reserved by IUPAC.[7,8] The framework of ThermoML has been described in detail[9–13]; the resultant ThermoML XML schema definition (XSD) contains all supported elements and is available in the data.nist.gov entry.[14] In brief, each ThermoML file contains a single Citation entry containing metadata corresponding to the published article along with all compounds (with associated sample metadata) with experimental data supported by the XSD. The ThermoML archive contains property data for organics systems from the initial publisher announcement of journal cooperation with TRC through the 2019 calendar year as present in SOURCE on 9/30/2020 for five major journals:[15]

- Journal of Chemical Engineering and Data (ISSN: 1520-5134)[16]
- The Journal of Chemical Thermodynamics (ISSN: 1096-3626)[17]
- Fluid Phase Equilibria (ISSN: 0378-3812)[18]
- Thermochimica Acta (ISSN: 0040-6031)[19]
- International Journal of Thermophysics (ISSN: 1572-9567)[20]

FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles have been established to provide guidelines to improve the infrastructure of scientific data such that machines are able to find and use the data with minimal human intervention.[21] FAIR data facilitates new research efforts to build on prior efforts more efficiently and effectively. The original ThermoML Archive was designed to mirror a typical journal issue layout at the time. The journal issue design did not conform with FAIR principles. Because all metadata would flow through the journal issue paradigm, the ThermoML data were difficult to find and access. Simple searches by property, compound or author were not possible using the website; to retrieve a ThermoML file, the user would need to navigate through journal issues provided by html pull-down menus for each journal. Several years ago, all ThermoML files for each journal were provided as five separate compressed files (.tgz) for bulk download to make the data more interoperable and reusable. However, there was no static, citable version of the data that would facilitate reproducible research. The present work describes how the ThermoML Archive has been completely reworked to make the resource more FAIR for humans and machines.

Methods

All TRC data for organics systems are stored in the SOURCE database.[3] SOURCE is an SQL database populated using in-house software tools and consumed for use in external software; SOURCE is an internal database that is not FAIR for external users. The need for ThermoML became clear around the same time that the cooperation between TRC and major data journals was established. In the following, we describe the methods used to guide the flow of thermodynamics data through the TRC and out to the improved ThermoML archive that is more in line with FAIR data principles for external users.

Flagging System and NIST TRC-Journal Cooperation

TRC develops internal tools to facilitate the flow of information from primary sources into SOURCE and then out of SOURCE for external use in compiled software, such as Thermo-Data Engine (TDE), or web applications, such as Web Thermo Tables[22] and the ThermoML Archive. The data content of SOURCE grows daily and is potentially dynamic due to error identification and corrections. Errors may occur at any step in the process. Often, errors are clear in the context of data, collected widely from many sources, and models provided by TDE. TRC has developed a system to internally flag data points that may be erroneous. In general, problematic data are revealed in multiple ways (please see Ref 1 for details); the four most common criteria applied by TRC experts (who have extensive experimental backgrounds) when flagging data are as follows:

- Data anomalies revealed based on the property type and the dependence on state variables
- Inconsistencies between various sources including high-level models
- Inconsistencies between different thermodynamically related properties
- Experimental deficiencies that potentially cause significant data errors

Challenges with erroneous published data naturally led to the idea of journal cooperation, the motivation being a mutual benefit of improved data reporting.[15] TRC experts work with the editors of the five cooperating journals; after the review process, but before the paper is accepted for publication, the data are captured and evaluated using TRC tools. Most errors are caught and corrected. Occasionally, submissions are rejected by the editors as a result of TRC reports. As a result of the NIST-TRC/journal cooperation, the error rate of cooperating journals estimated using the evaluation process described above is significantly lower than that of other journals. For the 2015-2019 publications available in SOURCE, the average error rates are 0.5 % and 4.1 %, respectively.

Electronic submission of data by the authors was a further benefit that was imagined but never fully realized.[15] Early attempts to achieve electronic data submission were not successful due to the challenges of the data-capture problem with respect to software tools available at the time. We continue to pursue a successful electronic data submission system with hopes that widespread adoption would facilitate future updates of the ThermoML Archive.

Overall Speci cations

In the following sections, we describe how the information described above can be accessed in the updated ThermoML archive. In this section we describe a minimal set of specifications of the ThermoML Archive to clarify implementation and use. The specifications are listed as follows:

- The ThermoML.xsd specifies the structure of information for each ThermoML entry.
- The ThermoML archive provides JSON and XML files; each JSON file contains all information necessary to generate the corresponding valid ThermoML XML file.
- All JSON and XML files are collected in a versioned, compressed file that may be downloaded from `https://data.nist.gov/od/id/mds2-2422`.
- The versioned, compressed file downloaded from data.nist.gov is the source of information for the ThermoML Archive web application and RESTful API.
- The RESTful API provides programmatic access to metadata.
- The web application exposes the information in a way that facilitates data discovery and reuse.

The above specifications guide and restrain any future developments to retain long-lived functionality regardless of how the web application is implemented; web technologies evolve rapidly. The objective of the web application is to expose the information contained within the versioned collection of files, thus allowing users to use the resource productively and reproducibly. The previous version of the ThermoML archive did not satisfy these specifications.

Version 2020-09-30 of ThermoML Archive

ThermoML entries for each journal correspond to the time between the initial point of cooperation through the end of the calendar year 2019. There are around 2.7M data points in the archive, with over 2M data points collected from J. Chem. Eng. Data. and J. Chem. Thermodyn. combined (Table 1). All data points are extracted to the ThermoML archive regardless of error flag state. Data points are assigned expanded uncertainties for 0.95 level of confidence. The uncertainties of the values include both propagated uncertainties of the variables and internal TRC estimates based on the method used and data consistency.

The ThermoML archive was extracted directly from SOURCE on September 30, 2020 as javascript object notation (JSON) files serialized with a structure closely related to the ThermoML XSD (Listings 1 and 2). These files are provided as some data consumers may prefer JSON for their applications. The JSON files are also used to generate all files, including the ThermoML files, used in the three major components of the archive: the web application, the RESTful API, and the data.nist.gov entry that provides bulk downloads. The web application was developed using the Vue reactive javascript framework[23] to provide dynamic browsing and searching. Searching capabilities of the web application are enabled by a RESTful API provided by the Cordra open source software used to manage digital objects; the RESTful API can be accessed programmatically. The RESTful API improves our alignment to FAIR Principle A1 and A1.1, where (meta)data are retrievable via a standardized communication protocol that is open and universally implementable. The Digital Object Interface Protocol (DOIP) that is supported by Cordra may be exposed by the ThermoML API in the future. The data.nist.gov entry uses robust NIST infrastructure to provide a reusable fixed version of all the data in the archive. All three use shared information with the deliberate intent to enhance human interaction with the archive and development of automated workflows for data consumption. We provide additional details for each component in the following sections.

Web application for browsing and searching

The dynamic web application is accessed at the index of the site and provides two additional routes `/ThermoML/Browse` and `/ThermoML/Search`. A small number of json files (located in the `/ThermoML/jsons` directory) are used to populate the browsing menus based on journal issues and properties located at the `/ThermoML/Browse` route. A search box that passes the input to the REST API is located at the `/ThermoML/Search` route. The property browsing follows the same property groups and property names as defined in the ThermoML XSD. Within the summary tab of the index route, the accumulated property counts can be browsed by journal (Figure 2); within the `/ThermoML/Browse` route, the same browsing component provides the ability to extract DOIs containing data for a given property. For example, browsing can be used to fetch 51 DOIs that contain 266 pure compound heat capacities at saturation pressure (Figure 3).

The web application also serves static html pages of metadata for each journal article; these files were introduced with embedded `https://schema.org/Dataset` JSON-LD (linked data) such that ThermoML Archive entries are indexed and findable via major commercial search engines (FAIR Principle F4), such as Google Dataset Search tool. The embedded JSON-LD makes significant use of the variableMeasured and measurementTechnique properties (Listing 3). A recent study indicated that only 20.90 % of datasets report variableMeasured and fewer than 1 % of datasets report measurementTechnique.[24] Thus, the ThermoML JSON-LD implementation is forward-looking. The JSON-LD can be explored by viewing the source of the html page. The JSON files extracted from SOURCE are served alongside the html and ThermoML files; thus, each journal article has three corresponding files: a JSON, a ThermoML, and an HTML file. These files are organized in web application filesystem by DOI where the three journal DOI prefixes (10.1007, 10.1016, 10.1021) provide the three root directories holding three files for each suffix corresponding to the extensions (.json, .xml, and .html), see Listing 4.

RESTful API

The RESTful API is provided by Cordra, open source software for the management of digital objects.[25] The designed purpose of the API is to provide searching capability on ThermoML metadata without providing the actual data points. Thus, the ThermoML JSON files are modified to contain no data points before they are posted to Cordra; the data points are replaced with a summary of the data point counts. For programmatic use of the archive, the user hits against their local copy of the files downloaded from data.nist.gov. This approach reduces server loads and increases the rate of data access.

Cordra uses the Lucene query syntax for searching. In addition to more routine searching on simple strings, powerful searches can be constructed using the structure of the JSON files. For example, a simple search on all fields, such as abstract, title, etc., can be carried out using the string "octanol AND water" that finds 119 ThermoML entries. A more precise search string, which returns 102 ThermoML entries, can be constructed to find entries with a compound of the common names within the digital object,

/Compound/_/sCommonName/_:water AND /Compound/_/sCommonName/_:octanol,

where the underscores (_) sweep through the lists of compounds with lists of common names. Users can use the XSD or browse (search) the web app for example JSON files to develop well-structured searches.

The amount of information returned from the API can be reduced dramatically by appending the ids parameter (&ids) to the query. For example,

/Compound/_/sCommonName/_:octanol&ids,

returns a list of 285 ids. The ids are handles that combine a TRC prefix with the DOI. Currently, ThermoML handles are not registered with a handle server. They may be registered in the near future; thus, users should be aware that the TRC prefix is subject to change. Using Cordra lifecycle hooks, the behavior of the requestContext parameter has been introduced to allow the response to be reduced to contain a subset of elements. For example, calls to the API from the /ThermoML/Search route use the added URL parameter for the request,

requestContext={"Citation":true,"data_summary":true,"Compound":true},

to reduce the response information to contain only the citation, compounds, and data summary for display in the web application. We request that API users invoke searches with paging (pageNum and pageSize url parameters), when possible, in order to reduce server loads.

Bulk downloads via data.nist.gov entry DOI: 10.18434/mds2-2422

All the corresponding JSON and ThermoML files for each journal entry are provided as a downloadable compressed file (.tgz) from data.nist.gov.[14] Extracting the .tgz file creates new directories for each DOI prefix (Listing 4). Any future updates would provide a new .tgz containing changes; extracting the .tgz would clobber files in the same directory tree with any new versions. Automated tools and workflows using ThermoML Archive data should use this bulk data to reduce server loads by leveraging a local copy of the files.

Conclusions

ThermoML is a valuable source of thermophysical and thermochemical data. The ThermoML archive has been completely reworked using the FAIR Data Principles as an aspirational guide for continuous improvement. Citation and material property metadata has been leveraged to make the data more findable and accessible using both JSON-LD that is indexed by major commercial search engines such as Google Datasets Search and using a new RESTful API provided by the Cordra digital object management software. The scope of the archive has been expanded to include JSON files containing all information contained in the ThermoML files. The ability of programming languages to parse JSON into native data structures makes the archive more accessible and interoperable by lowering the barrier of entry for machine workflows. All JSON and ThermoML files in the archive are provided as a bulk .tgz download from data.nist.gov,[14] making data within the archive more findable, accessible, interoperable, and reusable as a resource on the local machine of the user.

```
<DataReport ... >
  <Version> ...</Version>
  <Citation>...</Citation>
  <Compound> ...</Compound>
  <Compound> ...</Compound>
  <PureOrMixtureData> ...</PureOrMixtureData>
  <PureOrMixtureData> ...</PureOrMixtureData>
  <PureOrMixtureData> ...</PureOrMixtureData>
  <ReactionData>... </ReactionData>
</DataReport>
```

Listing 1: The overall structure of a ThermoML file. The above file contains two compounds, three PureOrMixtureData sets and one ReactionData set. The files are written directly from JSON files (Listing 2) such that the information in the ThermoML is in sync with that in the JSON files. Both the JSON and ThermoML files are included in the ThermoML Archive.

```
{
    ReactionData: [...],
    tml_elements: [
        "Version",
        "Citation",
        "Compound",
        "PureOrMixtureData",
        "ReactionData"
    ],
    Compound: [...],
    Version: {...},
    THERMOML_MD5_CHECKSUM: "...",
    Citation: {...},
    PureOrMixtureData: [...]
}
```

Listing 2: The overall structure of ThermoML Archive JSON files extracted from SOURCE. The JSON displayed above is written to the main DataReport element of a ThermoML file. The JSON files contain minimal information not present or supported by the ThermoML XSD. The "THERMOML_MD5_CHECKSUM" is the MD5 hash of the ThermoML file that may be used to verify file downloads. Since JSON is not ordered by specification, the "tml_elements" lists are used to write the ordered XML file (Listing 1), recursively as each element contains a "tml_elements" list.

```
{
    @type:  Dataset ,
    name: "ThermoML Data for: ...",
    license:  https://www.nist.gov/open/license ,
    material: ..., creator: ..., description: ...,
    variableMeasured:  ..., measurementTechnique: ...,
    citation: {
        @type: [  CreativeWork ,  PublicationIssue  ],
        url: "http://dx.doi.org/...",
        identifier: {...},
        pageStart: ...,  pageEnd:  ...,
        isPartOf:  {
                        @type: [
                            "Periodical",
                            "PublicationVolume"
                        ],
                        name: ...,  volumeNumber: ..., issueNumber: ...,
                        yearPublished: ..., publisher: ..., issn: ...
                    }
    },
    distribution: {
        @type: "DataDownload",
        contentUrl: "https://trc.nist.gov/ThermoML/...xml",
        encodingFormat: "text/xml"
    }
}
```

Listing 3: The overall structure of the JSON-LD embedded in the detailed metadata html pages for each entry. The placeholders (...) are filled in with information from each entry.

```
trc.nist.gov/ThermoML/
trc.nist.gov/ThermoML/{DOI Prefix}/{DOI Suffix}.(xml|json|html)
trc.nist.gov/ThermoML-API/objects/{TRC Prefix}/{DOI Prefix}/{DOI Suffix}

ThermoML/
    10.1007/
        *.json
        *.xml
        *.html
    10.1016/
    10.1021/
    index.html
    css/
    js/
    jsons/
```

Listing 4: URL templates and file system organization for the ThermoML web application. Using a web browser to access trc.nist.gov/ThermoML will load the index of the web application written in the Vue reactive javascript framework; the web application provides dynamic browsing and searching capabilities. The site also serves static html, json, and thermoml files that may be accessed with the according DOI path. Searching is provided to the web application using a RESTful api that may be programmatically accessed.

Notes

The authors declare no competing financial interest. This article is a contribution of NIST, and is not subject to copyright in the United States for the authors. Trade names are provided only to specify the source of information and procedures adequately and do not imply endorsement by the National Institute of Standards and Technology. Similar products by other developers may be found to work as well or better.

Acknowledgements

TRC thanks Ken Kroenlein for his contributions to the development of the ThermoML Archive.

References

1. V. Diky, A. Bazyleva, E. Paulechka, J. W. Magee, V. Martinez, D. Riccardi, and K. Kroenlein, The Journal of Chemical Thermodynamics **133**, 208 (2019).

2. B. Wilthan, E. A. Pfeif, V. V. Diky, R. D. Chirico, U. R. Kattner, and K. Kroenlein, Calphad **56**, 126 (2017).

3. A. Kazakov, C. D. Muzny, K. Kroenlein, V. Diky, R. D. Chirico, J. W. Magee, I. M. Abdulagatov, and M. Frenkel, International Journal of Thermophysics **33**, 22 (2012).

4. V. Diky, R. D. Chirico, M. Frenkel, A. Bazyleva, J. W. Magee, E. Paulechka, A. Kazakov, E. W. Lemmon, C. D. Muzny, A. Y. Smolyanitsky, et al., *NIST Thermo-Data Engine, NIST Standard Reference Database 103b, version 10.4.2,* (2021), URL https://www.nist.gov/mml/acmd/trc/thermodata-engine/srd-nist-tde-103b.

5. *Project 2002-055-3-024: XML-based IUPAC standard for experimental and critically evaluated thermodynamic property data storage and capture* (2006), accessed: 2022-02-17, URL https://iupac.org/projects/project-details/?project_nr=2002-055-3-024.

6. *Project 2007-039-1-024: Extension of ThermoML - the IUPAC standard for thermodynamic data communications* (2011), accessed: 2022-02-17, URL https://iupac.org/projects/project-details/?project_nr=2007-039-1-024.

7. M. Frenkel, R. D. Chirico, V. Diky, Q. Dong, K. N. Marsh, J. H. Dymond, W. A. Wakeham, S. E. Stein, E. Königsberger, and A. R. H. Goodwin, Pure and Applied Chemistry **78**, 541 (2006).

8. M. Frenkel, R. D. Chirico, V. Diky, P. L. Brown, J. H. Dymond, R. N. Goldberg, A. R. H. Goodwin, H. Heerklotz, E. Königsberger, J. E. Ladbury, et al., Pure and Applied Chemistry **83**, 1937 (2011).

9. M. Frenkel, R. D. Chirico, V. V. Diky, Q. Dong, S. Frenkel, P. R. Franchois, D. L. Embry,

T. L. Teague, K. N. Marsh, and R. C. Wilhoit, Journal of Chemical & Engineering Data **48**, 2 (2003).

10. R. D. Chirico, M. Frenkel, V. V. Diky, K. N. Marsh, and R. C. Wilhoit, Journal of Chemical & Engineering Data **48**, 1344 (2003).

11. M. Frenkel, R. D. Chirico, V. V. Diky, K. N. Marsh, J. H. Dymond, and W. A. Wakeham, Journal of Chemical & Engineering Data **49**, 381 (2004).

12. R. D. Chirico, M. Frenkel, V. Diky, R. N. Goldberg, H. Heerklotz, J. E. Ladbury, D. P. Remeta, J. H. Dymond, A. R. H. Goodwin, K. N. Marsh, et al., Journal of Chemical & Engineering Data **55**, 1564 (2009).

13. M. Frenkel, V. Diky, R. D. Chirico, R. N. Goldberg, H. Heerklotz, J. E. Ladbury, D. P. Remeta, J. H. Dymond, A. R. H. Goodwin, K. N. Marsh, et al., Journal of Chemical & Engineering Data **56**, 307 (2011).

14. D. Riccardi, A. Bazyleva, E. Paulechka, V. Diky, J. W. Magee, A. F. Kazakov, S. A. Townsend, and C. D. Muzny, *ThermoML/Data Archive* (2021), URL `https://data.nist.gov/od/id/mds2-2422`.

15. R. D. Chirico, M. Frenkel, J. W. Magee, V. Diky, C. D. Muzny, A. F. Kazakov, K. Kroenlein, I. Abdulagatov, G. R. Hardin, W. E. Acree, et al., Journal of Chemical & Engineering Data **58**, 2699 (2013).

16. K. N. Marsh, Journal of Chemical & Engineering Data **48**, 1 (2003).

17. Announcement, The Journal of Chemical Thermodynamics **36**, iv (2004).

18. Announcement, Fluid Phase Equilibria **226**, v (2004).

19. Announcement, Thermochimica Acta **421**, 241 (2004).

20. W. M. Haynes, International Journal of Thermophysics **26**, 307 (2005).

21. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., Scientific Data **3** (2016).

22. K. Kroenlein, C. D. Muzny, A. F. Kazakov, V. Diky, R. D. Chirico, J. W. Magee, I. Abdulagatov, and M. Frenkel, *NIST Standard Reference Database 202 Web Thermo Tables (WTT)* (2012), URL https://www.nist.gov/mml/acmd/trc/web-thermo-tables-wtt.

23. *Vue.js, the Progressive JavaScript Framework*, accessed: 2022-02-17, URL https://vuejs.org.

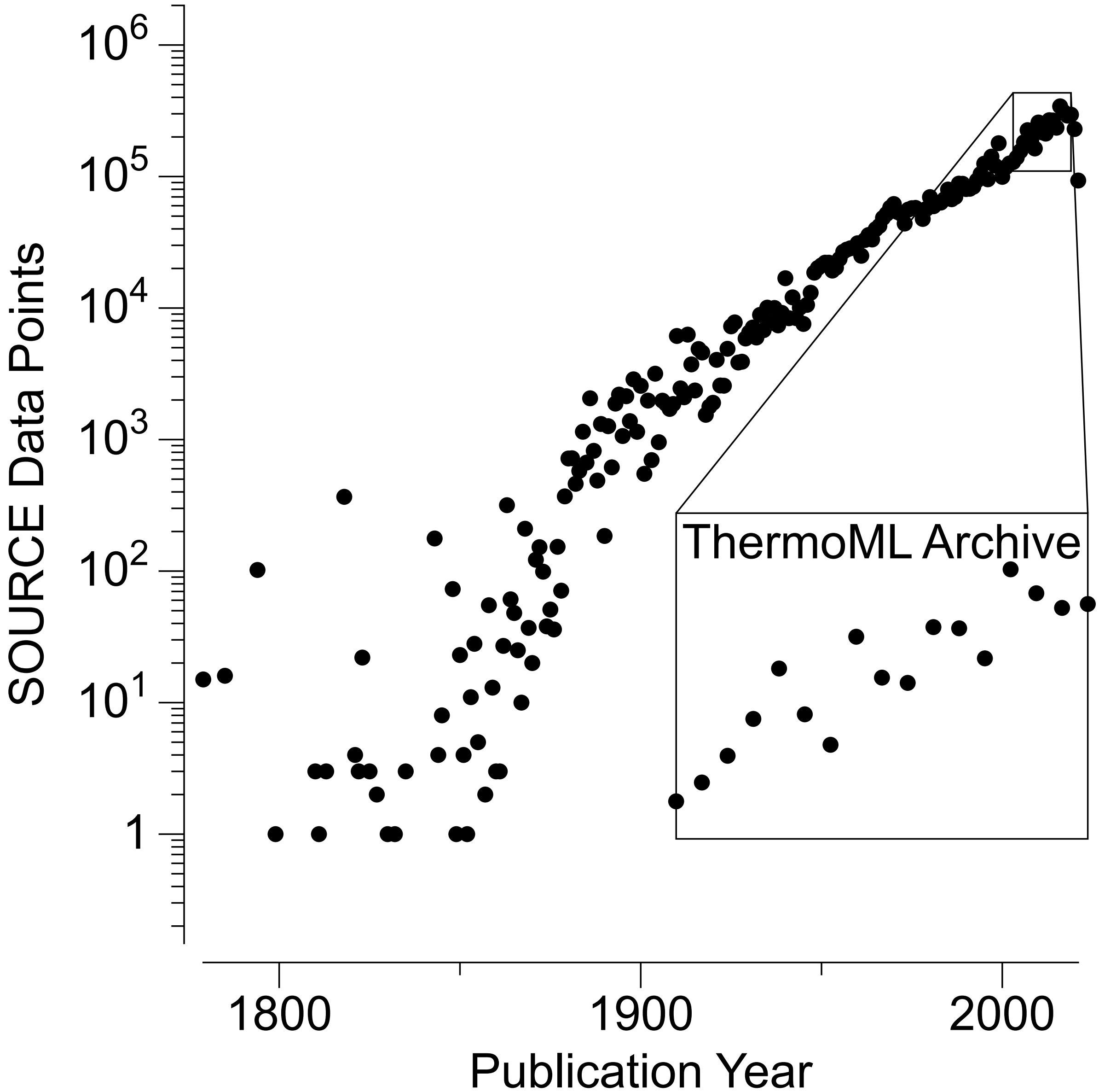
24. O. Benjelloun, S. Chen, and N. Noy, in *Lecture Notes in Computer Science* (Springer International Publishing, 2020), pp. 667–682.

25. *Cordra, Highly con gurable software for managing digital objects at scale*, accessed: 2022-02-17, URL https://cordra.org.

Figure 1: Number of experimental data points present in the SOURCE database accumulated by year of publication. TRC captures 300k-500k data points per year with most being associated with newly published articles in the cooperating journals. The statistics were collected from SOURCE around July 1, 2021. The inset corresponds, roughly, to the years covered by the ThermoML Archive for cooperating journals.

Figure 2: A rendered view of the ThermoML web application with the accumulated property point counts for all journals with the pure and binary properties displayed. The general ThermoML Archive descriptive component (shown) has General Info, Data Summary, and Searching Info tabs; the Data Summary tab provides a component for browse property groups and property names to discover data coverage.

Figure 3: A rendered view of the ThermoML web application for the /Browse route where the DOIs have been fetched for molar heat capacity at saturation pressure for the HeatCapacityAndDerivedProp group. There are 51 citations with 266 datasets. Links are provided from each DOI to the detailed metadata page.



10^6
10^5
10^4
10^3
10^2
10^1
1
SOURCE Data Points
1800
1900
2000
Publication Year
ThermoML Archive

Figure 1


NIST
Thermodynamics Research Center / ThermoML Archive / Browse | Search
Cooperating Journals
Journal of Chemical and Engineering Data (JCED)
The Journal of Chemical Thermodynamics (JCT)
Fluid Phase Equilibria (FPE)
Thermochimica Acta (TCA)
International Journal of Thermophysics (IJT)
Please read Liability Statement below before using the information on this site.
Links and Downloads
For current ThermoML specifications, refer to the xsd file and the original IUPAC project.
ThermoML.xsd
IUPAC Project 2007-039-1-024
.tgz file containing all available versions of ThermoML and JSON files
Cordra
Lucene Query Syntax
Contact Us
email TRC Info @ nist.gov
General Info
Data Summary
Searching Info
Summary of ThermoML Archive data points through 2019 for all cooperating journals
The ThermoML Archive includes data through 2019 for all cooperating journals as present in TRC databases on the date 2020-09-30. The /ThermoML/Browse route, linked above in the navigation gray bar, provides browsing by journal issue or by property, collated for all journals, to the source of the data points. A summary of all data point counts collected for each journal, and then overall journals, may be browsed below by property group and name.
Journal | Data Sets | Data Points | Pure Data Sets | Pure Data Points | Binary Data Sets | Binary Data Points | Ternary Data Sets | Ternary Data Points | Reaction Data Sets | Reaction Data Points | Show Details
All Journals | 123727 | 2692934 | 45855 | 563741 | 58302 | 1385778 | 18324 | 740838 | 1246 | 2577 | Hide Details
pure
Group Name | Data Sets | Data Points | Group Properties
Criticals | 896 | 896 | Show Details
RefractionSurfaceTensionSoundSpeed | 8887 | 55505 | Show Details
PhaseTransition | 8532 | 9372 | Show Details
VaporPBoilingTAzeotropTandP | 5829 | 68383 | Show Details
TransportProp | 5151 | 73086 | Show Details
VolumetricProp | 12190 | 191799 | Show Details
HeatCapacityAndDerivedProp | 4370 | 164700 | Show Details
binary
Group Name | Data Sets | Data Points | Group Properties
VaporPBoilingTAzeotropTandP | 4763 | 136047 | Show Details
VolumetricProp | 8987 | 484338 | Show Details
TransportProp | 4695 | 185833 | Show Details
CompositionAtPhaseEquilibrium | 15509 | 172684 | Show Details
HeatCapacityAndDerivedProp | 559 | 49817 | Show Details
PhaseTransition | 3984 | 38347 | Show Details
RefractionSurfaceTensionSoundSpeed | 5634 | 204148 | Show Details
ExcessPartialApparentEnergyProp | 3443 | 56553 | Show Details
ActivityFugacityOsmoticProp | 10161 | 55106 | Show Details

Figure 2

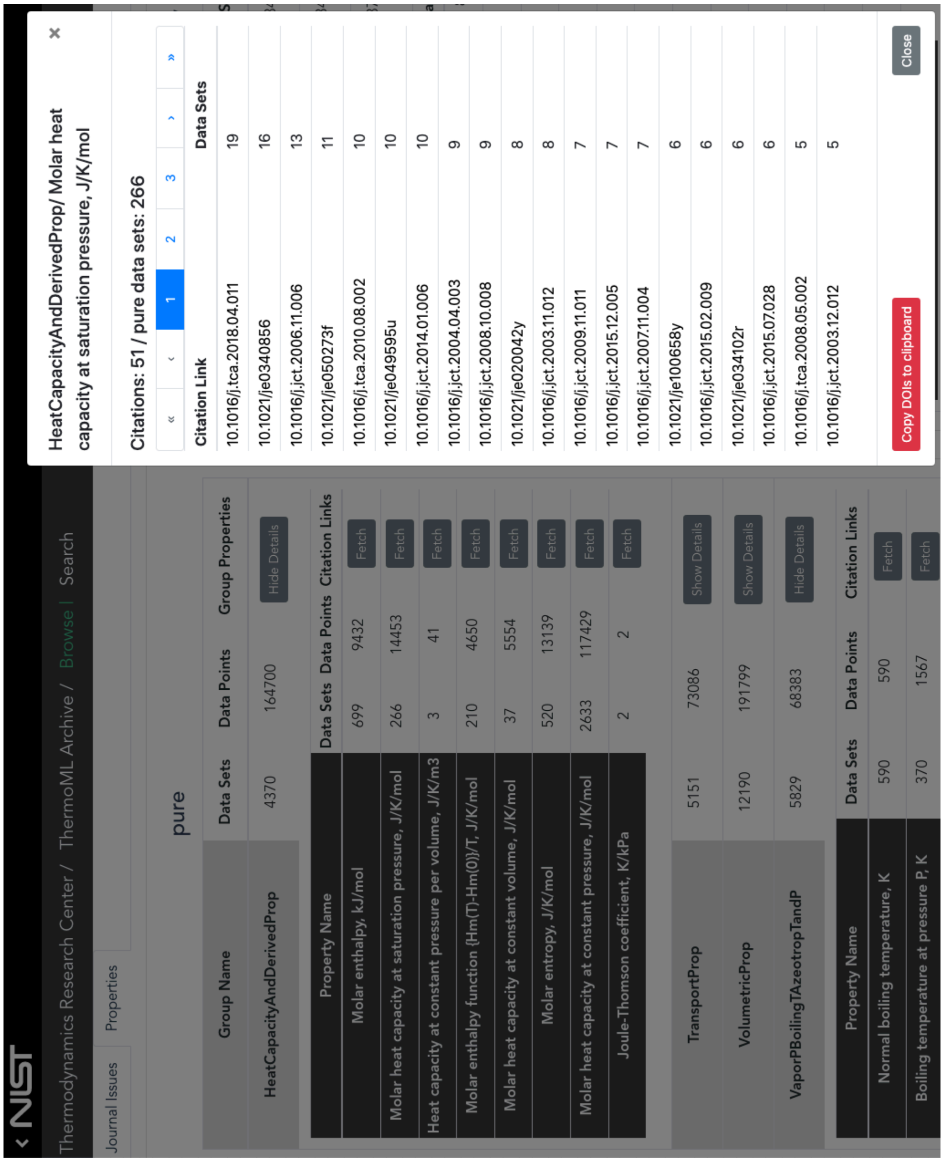

NIST
Thermodynamics Research Center / ThermoML Archive / Browse / Search
Journal Issues
Properties
pure
Group Name
Data Sets
Data Points
Group Properties
HeatCapacityAndDerivedProp
4370
164700
Hide Details
Property Name
Data Sets
Data Points
Citation Links
Molar enthalpy, kJ/mol
699
9432
Fetch
Molar heat capacity at saturation pressure, J/K/mol
266
14453
Fetch
Heat capacity at constant pressure per volume, J/K/m3
3
41
Fetch
Molar enthalpy function {Hm(T)-Hm(0)}/T, J/K/mol
210
4650
Fetch
Molar heat capacity at constant volume, J/K/mol
37
5554
Fetch
Molar entropy, J/K/mol
520
13139
Fetch
Molar heat capacity at constant pressure, J/K/mol
2633
117429
Fetch
Joule-Thomson coefficient, K/kPa
2
2
Fetch
TransportProp
5151
73086
Show Details
VolumetricProp
12190
191799
Show Details
VaporPBoilingTAzeotropTandP
5829
68383
Hide Details
Property Name
Data Sets
Data Points
Citation Links
Normal boiling temperature, K
590
590
Fetch
Boiling temperature at pressure P, K
370
1567
Fetch
HeatCapacityAndDerivedProp/ Molar heat capacity at saturation pressure, J/K/mol
Citations: 51 / pure data sets: 266
« ‹ 1 2 3 › »
Citation Link
Data Sets
10.1016/j.tca.2018.04.011
19
10.1021/je0340856
16
10.1016/j.jct.2006.11.006
13
10.1021/je050273f
11
10.1016/j.tca.2010.08.002
10
10.1021/je049595u
10
10.1016/j.jct.2014.01.006
10
10.1016/j.jct.2004.04.003
9
10.1016/j.jct.2008.10.008
9
10.1021/je020042y
8
10.1016/j.jct.2003.11.012
8
10.1016/j.jct.2009.11.011
7
10.1016/j.jct.2015.12.005
7
10.1016/j.jct.2007.11.004
7
10.1021/je100658y
6
10.1016/j.jct.2015.02.009
6
10.1021/je034102r
6
10.1016/j.jct.2015.07.028
6
10.1016/j.tca.2008.05.002
5
10.1016/j.jct.2003.12.012
5
Copy DOIs to clipboard
Close

Figure 3

Journal	**Total**	**Pure**	**Binary**	**Ternary**	**Reaction**
All Journals	123727	45855	58302	18324	1246
	2692934	563741	1385778	740838	2577
J. Chem. Eng. Data	57357	20604	27487	9141	125
	1285627	272263	632403	380519	442
J. Chem. Thermodyn.	36011	14043	16528	4603	837
	857345	176330	464525	214812	1678
Thermochim. Acta	8284	4109	3088	814	273
	144269	46674	60471	36678	446
Fluid Phase Equilib.	20531	6241	10635	3647	8
	364651	57466	203716	103461	8
	1544	858	564	119	3
Int. J. Thermophys.	41042	11008	24663	5368	3

Table 1: Counts for all data sets (first line) and data points (second line) for reaction, pure, binary, and ternary organics systems for each journal in the ThermoML Archive.