Core Scientific Dataset Model: A lightweight and portable model and file format for multidimensional scientific data

Deepansh J. Srivastava¹, Thomas Vosegaard², Dominique Massiot³, Philip J. Grandinetti¹*

¹ Department of Chemistry, Ohio State University, 100 West 18th Avenue, Columbus, OH 43210, United States of America, ² Laboratory for Biomolecular NMR Spectroscopy, Department of Molecular and Structural Biology, University of Aarhus, DK-8000 Aarhus C, Denmark, ³ CEMHTI UPR3079 CNRS, Univ. Orléans, F-45071 Orléans, France

* grandinetti.1@osu.edu

Abstract

The Core Scientific Dataset (CSD) model with JavaScript Object Notation (JSON) serialization is presented as a lightweight, portable, and versatile standard for intra- and interdisciplinary scientific data exchange. This model supports datasets with a \( p \)-component dependent variable, \( \{U_0, \ldots, U_q, \ldots, U_{p-1}\} \), discretely sampled at \( M \) unique points in a \( d \)-dimensional independent variable \( \{X_0, \ldots, X_k, \ldots, X_{d-1}\} \) space. Moreover, this sampling is over an orthogonal grid, regular or rectilinear, where the principal coordinate axes of the grid are the independent variables. It can also hold correlated datasets assuming the different physical quantities (dependent variables) are sampled on the same orthogonal grid of independent variables. The model encapsulates the dependent variables’ sampled data values and the minimum metadata needed to accurately represent this data in an appropriate coordinate system of independent variables. The CSD model can serve as a re-usable building block in the development of more sophisticated portable scientific dataset file standards.

1 Introduction

A frustrating and common problem faced by scientists in many disciplines is the lack of a portable scientific dataset format and universal standards for exchanging and archiving multidimensional datasets—both experimental and computational. Scientific datasets are too often saved in vendor-specific file-formats using proprietary software, making archiving and data-exchange problematic even within a discipline, let alone across disciplines. A majority of scientists rely on vendor-specific proprietary software to interact with their datasets. These scientists are at a constant risk that the original dataset files could become unreadable if a future version of the software stops supporting older file formats or the vendor stops supporting the software, or even worse, goes out of business.

As a result of such risks and incompatibilities, many scientists resort to using comma-separated values (CSV) files for dataset exchange and archival. Such an approach, however, is not...
resourceful, especially in the case of multi-dimensional datasets. Furthermore, such approaches often leave out essential metadata about experimental or computational procedures. Other scientists resort to specialized library packages to import datasets from the vendor-specific file formats into their favorite programming languages such as Matlab, Python, R, Java, or use the third-party software for dataset imports. This is only a temporary fix since it just delays the original problem as the dataset files are translated to yet another third-party software or user-specific file-format—and again, often with metadata loss.

With increasing pressure from the funding agencies and scientific journals to archive and share primary and processed data, there is a growing sense of urgency for a stable, resourceful and future-proof file-format for the exchange of scientific datasets. Here we take the first step in addressing this problem by proposing a **Core Scientific Dataset (CSD) Model** that can encode a wide variety of multi-dimensional and correlated datasets. The objective of the CSD model is to encapsulate the data values and the minimum metadata needed to accurately represent the data in an appropriate coordinate system. We envision the CSD model as a re-usable building block in a hierarchical description of more sophisticated portable scientific dataset file standards.

## 2 Overview of CSD model

The CSD model supports a dataset of a continuous physical quantity (dependent variable) discretely sampled on a multi-dimensional grid with vertexes associated with one or more independent quantities (dimensions), e.g., a density as a function of temperature, a current as a function of voltage and time, an ionization energy as a function of element symbol, etc.

Similarly, the CSD model supports a dataset with a multi-component dependent variable. For example, a color image with a red, green, and blue (RGB) light intensity components as a function of two independent spatial dimensions, or the six components of the symmetric second-rank diffusion tensor MRI dataset as a function of three independent spatial dimensions. In the CSD model, a **dataset** is defined as an $p$-component dependent variable, $\{U_0, \ldots, U_q, \ldots, U_{p-1}\}$, discretely sampled at $M$ unique points in a $d$-dimensional $(X_0, \ldots, X_k, \ldots, X_{d-1})$ space. Moreover, this sampling is over an orthogonal grid, *regular* or *rectilinear*, where the principal coordinate axes of the grid are the dimensions. A regular grid is an orthogonal grid where the spacing between vertex coordinates along each dimension is uniform. If the spacing along any one of the dimensions is not uniform, the grid is rectilinear.

The CSD model can also hold multiple datasets when different physical quantities (dependent variables) are sampled on the same multi-dimensional (independent variables) grid. We refer to this case as **correlated datasets**. One such example would be the simultaneous sampling of current and voltage as a function of time. Another example would be datasets for air temperature, pressure, wind velocity, and solar-flux, all simultaneously sampled on a two-dimensional grid associated with the same region of latitude and longitude coordinates.

We adopt the JavaScript Object Notation (JSON) as the file-serialization format [1] for the CSD model because it is human-readable, if properly organized, as well as easily integrable with any number of programming languages and field related application-software.

### 2.1 UML class diagram

The schema for the CSD model, in the form of a UML class diagram [2], is shown in Fig 1. In such diagrams, each class is represented with a box that contains two compartments. The top compartment contains the name of the class, and the bottom compartment contains the attributes of the class. A composition is depicted as a binary association decorated with a filled black diamond. Inheritance is shown as a line with a hollow triangle as an arrowhead.
Each line in the bottom compartment of a box describes a single attribute of the class in the form:

```
name : type [multiplicity] = default [properties]
```

In this line `name` is the name of an attribute in the class, `type` defines the kind of object that may be placed in the attribute, `multiplicity` indicates how many objects are assigned
Table 1. The description of the attributes from the Dimension class in version 1.0 of the CSD model.

| Attribute        | Description |
|------------------|-------------|
| type             | Required attribute for all Dimension objects. Holds a String object with one of the allowed DimObjectSubtype enumeration literals—linear, monotonic or labeled. |
| labels           | Required attribute for LabeledDimension objects. Holds an ordered and unique array of String objects containing UTF-8 allowed characters. Invalid for LinearDimension and MonotonicDimension objects. |
| coordinates      | Required attribute for MonotonicDimension objects. Holds an ordered and unique array of strictly increasing or decreasing ScalarQuantity objects along the dimension. The dimensionality of ScalarQuantity objects must be consistent with each other and other dimension attributes. Invalid for LinearDimension and LabeledDimension objects. |
| count            | Required attribute for LinearDimension objects. Holds an Integer object specifying the number of coordinates, N_x, along the dimension. Invalid for MonotonicDimension and LabeledDimension objects. |
| increment        | Required attribute for LinearDimension objects. Holds a ScalarQuantity object specifying the increment, Δx, along the dimension. Invalid for MonotonicDimension and LabeledDimension objects. |
| coordinates_offset | Optional attribute for LinearDimension objects. Holds a ScalarQuantity object specifying the coordinates offset, b_x, used in Eq (3) to calculate the coordinates along the dimension. The default value is a physical quantity with a numerical value of zero. Invalid for MonotonicDimension and LabeledDimension objects. |
| origin_offset    | Optional attribute for LinearDimension and MonotonicDimension objects. Holds a ScalarQuantity object specifying the origin offset, α_x, along the dimension. The default value is a physical quantity with a numerical value of zero. Invalid for LabeledDimension objects. |
| complex_fft      | Optional attribute for LinearDimension objects. Holds a Boolean specifying how the coordinate, B_x, along the dimension are calculated from Eq (3). When false, the value of Z_k = 0 otherwise, Z_k = T_j/2 where T_1 = N_0 and N_0 = 1 for even and odd values of N_0, respectively. Invalid for MonotonicDimension and LabeledDimension objects. |
| period           | Optional attribute for MonotonicDimension and LinearDimension objects. Holds a ScalarQuantity object specifying the period of the dimension. The default value is a physical quantity with an infinite numerical value, that is, the absence of this key indicates that the dimension is non-periodic. When present it indicates that all dependent variables are periodic along the dimension. A ScalarQuantity object with a numerical value of zero is invalid for this attribute. Invalid for LabeledDimension objects. |
| quantity_name    | Optional attribute for MonotonicDimension and LinearDimension objects. Holds a String object containing the quantity name associated with the dimension. This value may resolve ambiguities which may otherwise be inherent. For example, with only a unit of "J/(mol·K)", one cannot distinguish between the thermodynamic quantities 'molar entropy' and 'molar heat capacity.' Similarly, the units "1/s", "Bq", and "Hz" all have the dimensionality of inverse time, but generally "Bq" would be an acceptable unit for the quantity of radioactivity and "Hz" for frequency. If unspecified the valid quantity name is left at the end-user's discretion. A list of CSDM-accepted physical quantity names and their corresponding dimensionalties can be found in the supporting information. Invalid for LabeledDimension objects. |
| label            | Optional attribute for all Dimension objects. Holds a String object of UTF-8 allowed characters containing the label for the Dimension object. The default value is an empty string. |
| description      | Optional attribute for all Dimension objects. Holds a String object of UTF-8 allowed characters describing an instance of the Dimension object. The default value is an empty string. |
| reciprocal       | Optional attribute for MonotonicDimension and LinearDimension objects. Holds a ReciprocalDimension object. See Fig 1 for the list of attributes in this object. These attributes follow the same definitions as described in this Table with the only difference being that these attributes describe the reciprocal dimension. Invalid for LabeledDimension objects. |
| application      | Optional attribute for all Dimension objects. Holds a generic dictionary object. See section 2.5 for expected behavior. |

There are three subtypes of this class—LinearDimension, MonotonicDimension, and LabeledDimension. See Fig 1 for the list of valid attributes for a given subtype. If an attribute is optional, its value should only be serialized to the file if it is not the default value. As a recommendation, when deserializing a JSON file the numerical value associated with the physical quantities should be converted to a 32-bit or higher floating-point number.

https://doi.org/10.1371/journal.pone.0225953.t001

to the attribute. The multiplicity can be a single number, e.g., "[1]", indicating that one object must be assigned to the attribute. Alternatively, the multiplicity can be given as a lower and upper bound for how many objects can be assigned to the attribute, e.g., "[0..1]" indicates that the assignment of a single object to an attribute is optional. An asterisk indicates an unlimited number of objects. For example, an attribute with a multiplicity of "[1..*]" must have no less than one object and an unlimited upper bound of objects that can be assigned to it. The default is the object assigned when an optional attribute is unspecified. The properties value at the end of the line gives additional information on the attribute. In
Table 2. The description of the attributes from the `DependentVariable` class in version 1.0 of the CSD model.

| attribute            | description                                                                 |
|----------------------|------------------------------------------------------------------------------|
| type                 | Required attribute for all `DependentVariable` objects. Holds a `String` object with one of the two allowed `DVObjectSubtype` enumeration literals—`internal` or `external`. |
| components           | Required attribute for `InternalDependentVariable` objects. Holds an ordered array of `p` components. When the value of `encoding` attribute is `none` each component, $U_{0,...,p}^n$ is an ordered array of numerical values. When the value of the `encoding` attribute is `base64`, each component is a Base64 string. Invalid for `ExternalDependentVariable` objects. |
| components_url       | Required attribute for `ExternalDependentVariable` objects. Holds a `String` object containing the Uniform Resource Locator (URL) of a local or a remote file where the ordered array of numerical values \{$U_{0,...,p}^n$\} are stored as binary data. The CSD model utilizes the `https` and `file` schemes for locating the files. For local data files, the URL is specified relative to the `.csdfe` file and is located either in the folder containing the `.csdfe` file or in a subfolder of the folder containing the `.csdfe` file. The corresponding syntax follows `file://relative/path/to/the/file`. Invalid for `InternalDependentVariable` objects. |
| quantity_type        | Required attribute for all `DependentVariable` objects. Holds a `String` object with any of the allowed `QuantityType` enumeration literals from Table 3. The value specifies the number, $p$, and interpretation of the `DependentVariable` components. |
| numeric_type         | Required attribute for all `DependentVariable` objects. Holds a `String` object with one of the allowed `NumericType` enumeration literals from Table 4. This value represents the numeric type and the number of bits associated with each numerical value in Eq (5) when the component data is stored in an external file or when Base64 encoded into a string. When numerical values are expressed as JSON numbers, this value specifies the numerical precision needed for import. |
| unit                 | Optional attribute for all `DependentVariable` objects. Holds a `String` object representing the unit associated with the data values in Eq (5). The default value is `""`, i.e., the data values are dimensionless. |
| quantity_name        | Optional attribute for all `DependentVariable` objects. Holds a `String` object containing the `quantity name` associated with the physical quantity. See the description for the `quantity_name attribute in Table 1 for further details. |
| encoding             | Optional attribute for `InternalDependentVariable` objects. Holds a `String` object with one of the allowed `EncodingType` enumeration literals in Table 5. This value specifies the encoding method used to store the data values in the `components` attribute. The default value is `none`. Invalid for `ExternalDependentVariable` objects. |
| component_labels     | Optional attribute for all `DependentVariable` objects. Holds an ordered array of `String` objects where the $q^{th}$ `String` is the label associated with the $q^{th}$ component. The default value is an ordered set of empty strings. |
| name                 | Optional attribute for all `DependentVariable` objects. Holds a `String` object of UTF-8 allowed characters containing the name associated with the dependent variable. Naming is good practice as it improves the human readability of the serialized file when multiple dependent variables might be present. The default value is an empty string. |
| description          | Optional attribute for all `DependentVariable` objects. Holds a `String` object of UTF-8 allowed characters describing an instance of the `DependentVariable`. The default value is an empty string. |
| sparse_sampling      | Optional attribute for all `DependentVariable` objects. Holds a `SparseSampling` object, which contains the attributes `dimension_indexes` and `sparse_grid_vertexes`. The attribute `dimension_indexes` holds an array of integers indicating which dimensions in the ordered array of dimensions are sparsely sampled and form the sparse grid. The attribute `sparse_grid_vertexes` holds an array of integers defining the ordered set of sampled vertexes on the sparse grid. See section 2.4.1 for further details. |
| application          | Optional attribute for all `DependentVariable` objects. Holds a generic dictionary object. See section 2.5 for expected behavior. |

If an attribute is optional, its value may only be serialized to the file if it is not the default value.

[https://doi.org/10.1371/journal.pone.0225953.t002](https://doi.org/10.1371/journal.pone.0225953.t002)

**Fig 1** This is used to indicate whether a set of objects assigned to an attribute is ordered and/or unique.

For object attribute names we adopt the “snake case” convention with all lower case characters and “camel case” for class or type names. Attribute value types used in the model are given in Table 6 along with the corresponding JSON value type used for serialization of the model. Of particular importance in the CSD model is the `ScalarQuantity` type, which is composed of a numerical value and any valid SI unit symbol or any number of accepted non-SI unit symbols. It is serialized in the JSON file as a string containing a numerical value followed by the unit symbol, for example, “3.4 m” (SI) or “2.3 bar” (non-SI). The CSD model follows the International System of Units guideline [3] for defining the physical quantities. In software usage, one must adhere to stricter conventions for unit and physical constant symbols to avoid ambiguities and symbol collisions. All unit symbols are case sensitive. For derived
Table 3. QuantityType enumeration literals allowed in version 1.0 of the CSD model.

| literals            | description                                                                                     |
|---------------------|-------------------------------------------------------------------------------------------------|
| scalar              | This value represents a \( p = 1 \) single-component dependent variable where the \( i \)-\( th \) data value is interpreted as a scalar value, \( S_i = U_{i,i} \). |
| vector\(_m\)        | The value represents a \( p = n \) component dependent variable where the \( i \)-\( th \) data value is interpreted as a vector, \( V_i = [U_{i,1}, U_{i,2}, \ldots, U_{i,n}] \). |
| symmetric\(_m\)_\(_n\) | The value represents a \( p = mn \) component dependent variable where the \( i \)-\( th \) data value is interpreted as a \( m \times n \) matrix, with \( m \) rows and \( n \) columns. The \( p \) components of the matrix are in column-major order.  

\[
\mathbf{M} = \begin{bmatrix}
U_{i,i} & U_{i,1} & \cdots & U_{i,(n-1)i} \\
U_{i,1} & U_{i,1+1} & \cdots & U_{i,1+(n-1)} \\
\vdots & \vdots & \ddots & \vdots \\
U_{i,n-1} & U_{i,n-1+1} & \cdots & U_{i,n-1+(n-1)}
\end{bmatrix}
\]

Here, the entry at the \( i \)-\( th \) row and the \( j \)-\( th \) column is \( U_{i,j} \). |
| symmetric\(_n\)_\(_n\) | The value represents a \( p = n^2 \) component dependent variable. This is a special case of matrix data value where \( n = m \) and the matrix is symmetric about the leading diagonal. In this case, only the upper half of the matrix is specified. The \( n \times n \) symmetric matrix, \( \mathbf{M}^{(n)} \), of the \( i \)-\( th \) data value is interpreted as,  

\[
\mathbf{M}^{(n)} = \begin{bmatrix}
U_{i,i} & U_{i,1} & \cdots & U_{i,(n-1)i} \\
U_{i,1} & U_{i,1+1} & \cdots & U_{i,1+(n-1)} \\
\vdots & \vdots & \ddots & \vdots \\
U_{i,n-1} & U_{i,n-1+1} & \cdots & U_{i,n-1+(n-1)}
\end{bmatrix}
\]

The literals are String objects and correspond to the value of the quantity\_type attribute of the DependentVariable object. In the description, the index \( i \) refers to the \( i \)-\( th \) data value from the ordered array, \( U_{i,j} \) in Eq (5). |
| pixel\(_n\)          | The value represents a \( p = n \) component dependent variable where the \( i \)-\( th \) data value is interpreted as a pixel, \( [U_{1,0}, U_{1,1}, \ldots, U_{1,n}] \), with the \( n \) components corresponding to pixel component intensities. Note this quantity type, as do all quantity types, is restricted to components that share the same physical dimensionality, i.e., can be added or subtracted, making it, for example, appropriate for holding RGB or CMYK components but not HSV components. |

The literals are String objects and correspond to the value of the quantity\_type attribute of the DependentVariable object. In the description, the index \( i \) refers to the \( i \)-\( th \) data value from the ordered array, \( U_{i,j} \) in Eq (5). |

https://doi.org/10.1371/journal.pone.0225953.t003

Table 4. NumericType enumeration literals allowed in version 1.0 of the CSD model.

| literals   | description                                      |
|------------|--------------------------------------------------|
| uint8      | 8-bit unsigned integer                           |
| uint16     | 16-bit unsigned integer                          |
| uint32     | 32-bit unsigned integer                          |
| uint64     | 64-bit unsigned integer                          |
| int8       | 8-bit signed integer                             |
| int16      | 16-bit signed integer                            |
| int32      | 32-bit signed integer                            |
| int64      | 64-bit signed integer                            |
| float32    | 32-bit floating-point number                     |
| float64    | 64-bit floating-point number                     |
| complex64  | two 32-bit floating-point numbers                |
| complex128 | two 64-bit floating-point numbers                |

The literal is a String object corresponding to the value of the numeric\_type attribute of the DependentVariable object. |

https://doi.org/10.1371/journal.pone.0225953.t004
unit symbols, the multiplication and division of the units are represented by the asterisk symbol, “”, and the solidus symbol, “/”, respectively. For example, a unit of speed is “m/s”.

Note that derived unit symbols in the CSD model require explicit use of the multiplication symbol instead of multiplication implied with spacing between symbols, e.g., use “N•m” instead of “N m”. Similarly, avoid the use of compound symbols, e.g., use “kW•h” instead of “kWh”. The caret symbol, “^” is used for raising unit symbols to a power—a unit of force is “kg•m^2/s^2”, and a unit of concentration is “g/cm^3”. Operator precedence can be specified using parentheses, e.g., “J/(mol*K)”. Also, note that while both °C and °F are valid units, they are not proper thermodynamic temperature units and are discouraged due to their ambiguity. Further details on the SI system and how units are used in the CSD model are given in the supporting information.

2.2 CSDM object

At the root level of the CSD model is the CSDM object. The CSDM object includes a required version attribute whose value is a string representing the version number of the CSD model, here assigned a string value of “1.0”.

The optional timestamp attribute indicates when the CSDM file was last serialized and holds a combined date and time string.
representation of the Coordinated Universal Time (UTC) formatted according to the ISO-8601 standard. The optional `geographic_coordinate` attribute indicates where the CSDM file was last serialized and holds a `GeographicCoordinate` object, inside which are three attributes: the required `latitude` and `longitude`, and the optional `altitude`. Positive latitude values indicate latitudes north of the equator, while negative values indicate latitudes south of the equator. Longitude values are relative to the zero meridian, with positive values extending east of the meridian and negative values extending west of the meridian. Positive altitudes indicate above sea level while negative values indicate below sea level. All three are `ScalarQuantity` types. The optional boolean `read_only` attribute is set to true for archived datasets—informing applications that the dataset should not be modified or overwritten. The optional `tags` attribute holds a set of UTF-8 allowed string values describing keywords associated with the dataset. The `description` attribute appears in nearly every CSD model object and holds a UTF-8 allowed string describing the instance of the model object. The `application` attribute also appears in nearly every CSD model object and is a generic object that can be used for storing application-specific metadata within the CSD model. Further details on the expected behavior of `application` attributes are given in section 2.5.

The `dependent_variables` and `dimensions` attributes each hold a set of `DependentVariable` and `Dimension` objects, respectively. The ordered and unique set of `Dimension` objects, indexed from \( k = 0 \) to \( d - 1 \), define the \( d \)-dimensional coordinate grid where discrete samples of the dependent variables are taken.

### 2.3 Dimension object

The mapping of grid vertexes along the \( k \)th dimension to an ordered set of coordinates, \( X_k \), are defined by one of three `Dimension` subtypes: `LabeledDimension`, `MonotonicDimension`, and `LinearDimension`. Fig 1 gives the required and optional attributes along with their default values for the three subtypes. Descriptions of the attributes for all three subtypes are also given in Table 1, and examples of various instances are given in section 3.

#### 2.3.1 LabeledDimension object

An ordered set, \( A_k \), of \( N_k \) character string labels in the `labels` attribute of a `LabeledDimension` object are mapped to the grid vertexes along the \( k \)th dimension, becoming the ordered set of coordinates along the dimension, as given by

\[
X_k = A_k.
\]

This is a purely qualitative dimension, with no physical significance given to the spacing between grid vertexes along the dimension.

#### 2.3.2 MonotonicDimension object

An ordered set, \( A_k \), of \( N_k \) strictly ascending or descending coordinates in the `coordinates` attribute of a `MonotonicDimension` object are similarly mapped to the grid vertexes along the \( k \)th dimension and become the ordered set of coordinates along the dimension, also given by Eq (1).

For the `MonotonicDimension` and `LinearDimension` objects, the CSD model allows the mapping of grid vertexes along a dimension to an ordered set of absolute coordinates, \( X_k^{\text{abs}} \), using the `origin_offset` attribute according to

\[
X_k^{\text{abs}} = X_k + o_k \mathbf{1},
\]

where \( o_k \) is the value of the `origin_offset` attribute. Note, the `ScalarQuantity` objects in \( X_k \) and \( o_k \) must all share the same unit dimensionality.
2.3.3 LinearDimension object. The ordered set of $N_k$ uniformly spaced coordinates along the $k^{th}$ LinearDimension object are given by

$$X_k = \Delta x_k \ (J_k - Z_k) + b_k 1,$$

where $\Delta x_k$ and $b_k$ are the ScalarQuantity objects in the increment and coordinates_offset attributes, respectively, and $J_k$ is an ordered set of coordinate indexes along the $k^{th}$ dimension,

$$J_k = [0, 1, 2 \ldots j_k, \ldots N_k - 1].$$

Here, $N_k$ is the Integer object in the count attribute. As before, the absolute coordinates along the $k^{th}$ dimension are given by Eq (2). Again, the ScalarQuantity objects $\Delta x_k$, $b_k$, and $o_k$ must all share the same unit dimensionality.

The $Z_k$ variable in Eq (3) is an integer with a value of $Z_k = 0$ when the LinearDimension attribute complex_fft is false. The complex_fft is set to true when a complex fast Fourier transform (FFT) has been applied to the dataset along the $k^{th}$ dimension, and then the value of $Z_k$ becomes $T_k/2$, where $T_k = N_k$ and $N_k - 1$ for even and odd values of $N_k$, respectively. There are two reasons for the inclusion of the attribute complex_fft and the different values of $Z_k$. First, it provides the metadata needed for determining whether a forward (false) or reverse (true) complex FFT should be performed on the dataset. Second, a value of $Z_k = T_k/2$ in Eq (3) when complex_fft is true associates $b_k$ with the zero “frequency” after a complex FFT. This definition makes $b_k$ independent of count and the increment in the Reciprocal dimension, i.e., the dimension before the complex FFT.

2.3.4 ReciprocalDimension object. An optional attribute named reciprocal can be present in both the LinearDimension and MonotonicDimension objects. This attribute holds a ReciprocalDimension object which contains metadata about the coordinate that is reciprocal to the $X_k$ coordinate. This metadata is useful for datasets which are frequently transformed into the reciprocal dimension, such as NMR, FTIR and x-ray datasets.

2.4 DependentVariable object

The DependentVariable object can be one of two subtypes: InternalDependentVariable and ExternalDependentVariable, depending on whether the serialized components are stored internally with the rest of the serialized metadata or externally at a location specified by a uniform resource locator (URL) [4], respectively. Descriptions of all DependentVariable attributes are given in Table 2, as well as through examples given in section 3. See Fig 1 for the required and optional attributes along with their default values.

A DependentVariable object holds an ordered set of $p$ components indexed from $q = 0$ to $p - 1$,

$$\{U_0, \ldots, U_q, \ldots, U_{p-1}\}.$$  

Each component, $U_q$, contains an ordered array of $M$ physical quantity values indexed from $i = 0$ to $M - 1$. These values represent samples on the coordinates grid and are ordered to follow a column-major order relative to the ordered set of dimensions. If $U_q$ contains a sample at every vertex of the $d$-dimensional grid, then

$$M = \prod_{k=0}^{d-1} N_k.$$
and the mapping of the values in \( U_q \) to the grid vertexes follows a simple reshaping of \( U_q \) to a \( N_0 \times N_1 \times \ldots \times N_{d-1} \) matrix where \( d \) is the number of \textbf{Dimension} objects. In this case, the location or memory offset of the \( i^{th} \) value in a component array maps to a grid vertex with coordinate indexes, \((j_0, j_1, \ldots, j_{d-1})\), given by

\[ j_k = \frac{i}{\prod_{t=0}^{k-1} N_t} \mod N_k, \quad k = 0, \ldots, d - 1. \]  

Conversely, the memory offset of the \( i^{th} \) value in a component array is obtained from the ordered array of coordinate indexes \((j_0, j_1, \ldots, j_{d-1})\), according to

\[ i = \sum_{k=0}^{d-1} \left( \prod_{t=0}^{k-1} N_t \right) j_k. \]  

It is also helpful to recall that the value of the empty product, \( \prod_{n} a_n \) where \( m > n \) is 1.

Taken together, the \( i^{th} \) values from each of the \( p \) components form a quantity specified by one of the \texttt{quantity_type} attribute values given in Table 3.

\textbf{InternalDependentVariable}. The \texttt{components} attribute in an \textbf{InternalDependentVariable} object holds an \textit{ordered array} of \( p \) components, and each component, \( U_q \), is an \textit{ordered array} of \( M \) numerical values associated with the \( q^{th} \) component. When the value of the \texttt{encoding} attribute is \texttt{none} or unspecified, a JSON serialization of this object gives a human-readable list of numerical values. This approach, however, is not resourceful compared to the serialization of raw binary data. As JSON files are strictly text-based it is not possible to serialize raw binary data inside a JSON file. A commonly used approach to reduce JSON file sizes in such situations is to encode raw binary data into plain text using a binary-to-text encoding scheme. The CSD model allows this approach with the raw binary data for each component encoded into a Base64 string when the \texttt{encoding} attribute is set to \texttt{base64}. In this case, JSON serialization of the \texttt{components} attribute in an \textbf{InternalDependentVariable} object holds an \textit{ordered array} of \( p \) Base64 strings where the \( q^{th} \) string represents the array \( U_q \). Out of the various binary-to-text encoding schemes, we chose Base64 encoding because of its widespread use and easy access to decoders across most object-oriented programming languages. Base64 provides an efficiency of \( \sim 75\% \) compared to the serialization of raw binary data. When encoding and decoding raw binary data with Base64 we assume a ‘little-endian’ byte order for multi-byte numeric types such as 32-bit and 64-bit integers or floats. Typically, data saved on Intel x86 platforms use the little-endian as the native format. Also, binary floating-point standard IEEE 754 is assumed for float and complex numeric types.

\textbf{ExternalDependentVariable}. The \texttt{components_url} attribute is only \textit{valid} when the value of the corresponding \texttt{type} attribute is \texttt{external}. Its value is a \texttt{String} object containing the address of a local or remote file where the ordered array of numerical values \( \{U_0, \ldots, U_p \} \) are stored as binary data. In this case we also assume little-endian byte order and the binary floating-point standard IEEE 754 for float and complex numeric types. The CSD model utilizes the \texttt{https} and \texttt{file} schemes of the Uniform Resource Locator (URL) for locating the files. For local data files, the URL is specified relative to the \texttt{.csdfe} file (see section 2.6) and is located either in the folder containing the \texttt{.csdfe} file or in a subfolder of the folder containing the \texttt{.csdfe} file. The corresponding syntax follows, \texttt{file:./relative/path/to/the/file}.

2.4.1 \texttt{SparseSampling object}. Eqs (6), (7) and (8) are no longer valid when the \textbf{DependentVariable} components are sparsely sampled on the \( d \) dimensional grid. In this case, additional metadata is required to determine the grid vertex, \((j_0, j_1, \ldots, j_{d-1})\), where the \( i^{th} \)
sampled component value belongs. If the component is sparsely sampled along all \(d\) dimensions, then the additional metadata can be an ordered set of \(M\) grid vertexes. We must, however, consider the general mixed case of \(s\) fully sampled dimensions and \(d - s\) sparsely sampled dimensions. In this case, we adopt an approach where the component values are organized into a set of fully sampled \(s\)-dimensional cross-sections taken at vertexes of a sub-grid formed from the sparsely sampled dimensions, which we will call the sparse grid. In adopting this approach, we require that the component values along the fully sampled dimensions are packed together into the array in column-major order relative to the ordered set of fully sampled dimensions, i.e., excluding the sparsely sampled dimensions.

The SparseSampling object provides this metadata in its two attributes `dimension_indexes` and `sparse_grid_vertexes`. The `dimension_indexes` attribute holds an ordered and unique set of integers indicating along which dimensions the DependentVariable is sparsely sampled. These dimensions form the sparse grid. The `sparse_grid_vertexes` holds an ordered set of vertexes on the sparse grid. Each sparse grid vertex is an ordered array of \(d - s\) indexes. To make the serialization more resourceful, we flatten the ordered set of arrays intended for the `sparse_grid_vertexes` attribute into an ordered array of integers, for example,

\[
[[1, 0], [3, 4], [5, 7], [8, 11], \ldots] \rightarrow [1, 0, 3, 4, 5, 7, 8, 11, \ldots],
\]

in a case of two sparsely sampled dimensions. The set of arrays (on the left) can be easily reconstructed from the array of integers (on the right) given the number of indexes specified in the `dimension_indexes` attribute. Additional storage reduction can be had by encoding the `sparse_grid_vertexes` array as a Base64 character string of specified unsigned_integer_type and little-endian byte ordering. The `encoding` attribute in the SparseSampling object would indicate this option with a value of `base64`.

### 2.5 Generic application objects—Beyond the CSD model

As stated earlier, the objective of the CSD model is to encapsulate the data values and the minimum metadata needed to accurately represent the data in an appropriate coordinate system, that is, the minimum metadata for defining the current state of the dataset. Thus, the goal of the CSD model is to always remain relevant as the state of the dataset changes. In our refinement of the CSD model, we identified any metadata attribute as extraneous if it could become irrelevant as the state of the dataset changes. Metadata attributes extraneous to the CSD model could generally be classified as belonging in one of four broad and somewhat overlapping categories: acquisition, process, analysis, and presentation. The design of models organizing these extraneous metadata attributes tends to be scientific domain specific, although some commonalities exist. The CSD model allows the inclusion of metadata models describing these other categories using generic application objects. An application can place its own attribute type, e.g., a dictionary object with application-specific metadata attributes inside each generic application object using a reverse domain name notation string as the attribute key, for example, “com.example.myApp”. The use of a reverse-DNS key provides a simple mechanism for reducing name-space collisions. Overall, we believe generic application objects give the CSD model enough flexibility to become the native file format of many applications.

This approach, however, creates a dilemma when CSDM files are saved and opened by different applications. Specifically, what does an application from company B (e.g., “com.B.process”) do with generic application objects placed in a CSDM file by an application?
from company A (e.g., “com.A.acquire”)? On the one hand, company B could retain the company A specific metadata (as found) in the generic application object using the “com.A. acquire” key as well as serialize its own metadata using the “com.B.process” key. If the company B application made any modification to the dataset, however, it runs the risk that parts of the company A application-specific metadata are now irrelevant or logically inconsistent with the newly saved dataset—potentially causing company A’s application to crash when it tries to open this newly saved dataset. On the other hand, company B could decide to discard the company A specific metadata, in which case the company A application can safely open the dataset saved by company B, but will have lost all of its previously saved metadata. Finding a consistent solution to this dilemma is critically important as one can easily envision a workflow where a dataset passes through many applications as it progresses from the raw dataset to the final “product.” During such a workflow there is often an expectation of an audit trail, which most likely could be determined from application metadata saved by each application used during the workflow.

One approach that could solve this dilemma is to allow the CSDM file to contain a time-ordered array of CSDM objects. In other words, company B would simply append a second CSDM object with only company B metadata to the array that already contains the CSDM object created by company A. No application metadata would be lost, and the metadata in each CSDM object would be relevant and logically consistent with its respective datasets. In this approach, the CSDM “array” file would grow as each application completes its task in the overall workflow.

It is our opinion, however, that it is better to delegate such a task of managing a time-ordered array of CSDM objects to the operating system. In this approach, we envision the workflow associated with a particular dataset to result in a folder containing a series of CSDM files, each a snapshot from the workflow as it progresses from the raw dataset to the final “product.” When each application is finished with its workflow task a CSDM file is saved with the \texttt{read\_only} flag set to true, so that any future work on the dataset would be performed on a copy of the CSDM dataset, leaving the “read-only” file with application metadata intact. Typically, the \texttt{read\_only} flag would be set to true immediately after the acquisition of raw data, after processing is complete, or after analysis of a dataset. Delegating the task of managing a time-ordered set of CSDM objects to the operating system also makes the workflow status involving individual CSDM files more transparent to the end-user. In adopting this solution we propose the general rule that \textit{while application attributes should be visible to any application opening a CSDM file, only the reverse-DNS owners have permission to use their respective keys to place an attribute in an application object.}

An application could implement an additional layer of protection from application metadata loss by saving CSDM compliant files with its own application-specific file extension. Other applications could still open the CSDM compliant file but would be discouraged from saving with another application’s file extension.

2.6 JSON file-serialization

A JSON file is ordinarily a UTF-8 encoded text file which is built on two structures: a collection of unordered key-value pairs and an ordered list of values. The “\texttt{key}”: \texttt{value} pair is separated by a \texttt{colon} symbol, with the \texttt{key} to the left and the \texttt{value} to the right of the colon. Different key-value pairs are separated using \texttt{commas}. The JSON keys are always wrapped in double quotation marks, as in “\texttt{key}”, and the value type can either be (a) a string, (b) a number, (c) a JSON object, (d) an array, (e) a boolean or (d) null. A string is a composition of
JSON allowed characters \[1\] wrapped in double quotation marks. A number can be integer or float. A JSON object is an unordered set of key-value pairs which begins with a left curly brace, \{, and ends in a right curly brace, \}. An array is an ordered collection of JSON values that begins with a left square bracket, \[, and ends in a right square bracket, \]. A boolean is true or false. In the JSON serialization of the CSD model, the JSON “key” corresponds to the attribute name of the various CSD model objects while the JSON value and CSD model attribute value follow the relationship listed in Table 6.

Efforts have been made in the design of the CSD model to keep the keys intuitive and self-explanatory to all scientists and engineers. To further enhance the human-readability aspect of the files, we recommend, as a general rule, that no key be present in the file unless its value differs from the default value. With this in mind, the CSD model defines all boolean values as false when unspecified. In other words, the only boolean keys that need to appear in the file are those set to true.

The serialization file names are designated with two possible extensions: .csdf and .csdfe, the acronyms for Core Scientific Dataset Format and Core Scientific Dataset Format External. When all data values are stored within the file, i.e., there are no instances of an ExternalDependentVariable object in the serialization, then the .csdf file extension is allowed, otherwise, the serialization file name must use the extension .csdfe. This difference in extensions is intended to alert the end-user to a possible risk of failure if the external data file is inaccessible when deserializing a file with the .csdfe file extensions.

### 3 \(dD\{p_0, p_1, \ldots\}\) example datasets

In this section we examine the CSD model in a number of illustrative examples. We use a shorthand notation of \(dD(p)\) to indicate that a dataset has a \(p\)-component dependent variable defined on a \(d\)-dimensional coordinate grid. In the case of correlated datasets the number of components in each dependent variable is given as a list within the curly braces, i.e., \(dD\{p_0, p_1, p_2, \ldots\}\).

Efforts have been made to include examples across disciplines, although given our expertise in magnetic resonance spectroscopy, we include multiple examples from this field. It is worth noting, however, that magnetic resonance datasets prove to be excellent test cases for the CSD model as they are diverse and often multi-dimensional in nature. We have converted a variety of datasets from various fields to the CSD model format. To accomplish this, we utilize several Python packages \[5, 6\] to import the original field-specific scientific datasets as Numpy \[7\] array(s) and export the latter in the CSD model format using the csdmpy package for Python, described in the appendix.

#### 3.1 1D{1} examples

In this section, we examine the JSON serialization for illustrative cases of 1D{1} datasets. These are the simplest cases, with one dimension, \(d = 1\), and one single-component dependent variable, \(p = 1\). The supplementary material gives further 1D{1} examples from FITR, UV-vis, and EPR spectroscopies.

**GMSL.csdf.** An example of a JSON serialized CSD model holding a 1D{1} dataset is shown in Listing 1. This dataset is a measurement of the global mean sea level \[8\] (GMSL) based on the satellite altimeter data from 1993-2009.

**Listing 1. CSD model depiction of the global mean sea level dataset.** A JSON serialized CSD model describing the global mean sea level dataset. The listing was created by the authors using data from reference \[8\].
Fig 2. CSD model depiction of the global mean sea level dataset. A line plot, derived from Listing 1, depicting the global mean sea level as a function of time. The figure was created by the authors using data from reference [8].

https://doi.org/10.1371/journal.pone.0225953.g002
At the root level is the csdm key, an acronym for the core scientific dataset model. The value of this key is a JSON object which is a serialization of the CSD model’s CSDM object and includes six keys—version, timestamp, tags, description, dimensions, and dependent_variables. The value of the dimensions key is an array (lines 7-14) with a single JSON object defined in-between lines 8 and 13. This object is a JSON serialization of the CSD model’s Dimension object. In this example, it represents a LinearDimension object, as indicated by the value of linear in the type key, and with a coordinate count of 1608 as defined by the value of the key count. Furthermore, it is a temporal dimension with ScalarQuantity values of 0.0833 yr for increment and 1880.0417 yr for coordinates_offset. The coordinates at vertexes along this temporal dimension are obtained from Eq (3).

The value of the dependent_variables key is an array (lines 15-31) with a single JSON object describing the global mean sea level. This object is a JSON serialization of an InternalDependentVariable object, with data values stored within the object, as indicated by the value of internal for the type key. The data values are serialized as JSON numbers as seen in-between lines 23-28 of Listing 1. Ellipses indicate where superfluous lines were omitted from the listing. The value of float32 for the numeric_type key indicates that the array of JSON numbers should be converted into a numerical array of data values with 32-bit floating-point precision on import. The value of mm for the unit key is the unit associated with the data values. The value of the component_labels is an array with a single entry holding the label associated with the component values. The value of scalar for the quantity_type key indicates that the component of the dependent variable is interpreted as scalar.

A plot of the dataset is shown in Fig 2. Note that meta-data on how a dataset is presented in a plot or otherwise is not included in the CSD model. While such presentation metadata is outside the scope of the core model, it can be included in an application dictionary.

Another simple example of a 1D dataset, acquired by the authors for this work, is shown in Listing 2. This example corresponds to a $^{13}$C free induction decay signal from a nuclear magnetic resonance spectroscopy of ethanol.

The value of the dimensions key is an array (lines 12-26) with a single JSON serialized LinearDimension object (lines 13-25) representing a temporal dimension with 4096 coordinate positions sampled every 0.1 ms starting at −0.3 ms. The coordinate values along the dimension are evaluated using Eq (3). This LinearDimension object also contains an optional JSON serialized ReciprocalDimension object (lines 19-24) as the value of the reciprocal key. In this example, it provides the metadata needed for describing the reciprocal time or the frequency dimension, i.e., after a Fourier transform.

The value of the dependent_variables key is an array (lines 27-41) with a single JSON serialized InternalDependentVariable object (lines 28-40) describing the signal response. While the keys and values in this object are similar to the corresponding object from the previous example, a key difference is that the value of the numeric_type key denotes a complex64 numeric type. Complex numbers are stored as an ordered array of alternating real and imaginary data values, starting with the real value. In this example, the first and the last complex numbers of the signal in Fig 3 are (−8899.406 − 1276.773) and (−193.923 − 67.065), respectively. Note that the length of the ordered data array is 2M for complex numeric types, where M is the total number of sampled data points. Fig 3 shows a line plot of the time domain NMR decay signal.

**Listing 2.** CSD model depicting an 1-D NMR dataset. JSON serialized CSD model describing the $^{13}$C NMR Bloch decay time signal along with the relevant metadata of the reciprocal frequency dimension.
In Listing 3 is an illustration of a 1D[1] mass spectrum dataset [9] serialized with sparse sampling. Here the InternalDependentVariable object (lines 17-43) holds a SparseSampling object (lines 23-34) in the sparse_sampling key. Inside the SparseSampling object are the three keys dimensions_indexes, sparse_grid_vertexes, and unsigned_integer_type. The dimensions_indexes key holds an array of integers specifying the indexes of the dimensions along which the dependent variable is sparsely sampled. In this case, it is the zeroth dimension, i.e., the only dimension in the dataset. The sparse_grid_vertexes key holds an array of integers specifying the vertexes on the one-dimensional sparsely sampled grid. Again, in this example with only one dimension, the array of integers corresponds to the sampled sparse grid vertexes, i.e., the
coordinate indexes, \( j_0 \), along the zeroth dimension. The value uint8 for the unsigned_d_integer_type key is the numeric type used when importing the JSON serialized integer array from the sparse_grid_vertexes key.

### 3.2 2D\{1\} examples

**TEM.csdfe.** In Fig 4 is an intensity plot of a Transmission Electron Microscopy (TEM) dataset of a section of the early larval brain of *Drosophila melanogaster* used in the analysis of neuronal microcircuitry [10]. The CSDM JSON serialization for this 2D\{1\} dataset is given in Listing 4. This dataset has two dimensions, \( d = 2 \), and one single-component dependent variable, \( p = 1 \).

The value of the dimensions key is an array with two JSON serialized LinearDimension objects, defined in-between lines 8-12 and 13-17. Both these objects describe a linearly sampled spatial dimension with 512 points sampled every 4 nm. As before, Eq (3) gives the ordered list of the coordinates along the respective dimensions. The value of the dependent_variables key is an array containing a single JSON serialized InternalDependentVariable object (lines 20-26). Unlike the previous examples, the value of the components key is an array with a single element. This element is a Base64 encoded string, as indicated by the encoding key, and decodes to an array of binary data values which are interpreted as an array of numerical values with a uint8 numeric type. The array of numerical values is then mapped to the 512 × 512 coordinate grid according to Eqs (7) and (8).

**bubble.csdfe.** In Fig 5 and in Listing 5 we present a 2D\{1\} astronomy dataset of the bubble nebula acquired at 656 nm wavelength by the Hubble Heritage Project [11] team. In this example, the value of the dimensions key is an array with two JSON serialized LinearDimension objects defined in lines 8-15 and 16-23. Both these objects describe a linearly sampled angular dimension. The value of the dependent_variables key is an array with a single JSON serialized ExternalDependentVariable object, described in lines 26-32. In this example, the value of the type key is external, indicating that the data values are stored in an external file located at the Uniform Resource Locator (URL) address given by the components_url key. In this case, the address corresponds to a local file, designated by the file scheme of the URL, relative to the location of the bubble.csdfe file. The external file holds an ordered array of 11592 × 11351 binary values, which are specified by the numeric_type key as 32-bit floating-point numbers.
Listing 3. CSD model depiction of a sparse mass spectrum. JSON serialized CSD model describing the mass spectrum of acetone.

```json
acetone_csdf
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-06-23T17:53:26Z",
    "description": "Mass spectrum of acetone",
    "dimensions": [
      {
        "type": "linear",
        "count": 51,
        "increment": "1",
        "coordinates_offset": "10",
        "label": "m/z",
        "quantity_name": "dimensionless"
      }
    ],
    "dependent_variables": [
      {
        "type": "internal",
        "name": "acetone",
        "numeric_type": "float32",
        "quantity_type": "scalar",
        "component_labels": ["relative abundance"],
        "sparse_sampling": {
          "dimension_indexes": [0],
          "sparse_grid_vertices": [
            27,
            28,
            ...
            ...
            48,
            49
          ],
          "unsigned_integer_type": "uint8"
        },
        "components": [
          [9,
           9,
           ...
           ...
           270,
           10]
        ]
      }
    ]
  }
}
```

Listing 4. CSD model depiction of a TEM image dataset. JSON serialized listing of a TEM dataset containing one single-component `InternalDependentVariable` object and two `LinearDimension` objects. The listing was created by the authors using data from reference [10].
Fig 4. CSD model depiction of a TEM image dataset. An intensity plot, derived from Listing 4, of a TEM dataset depicting the early larval brain of Drosophila melanogaster used in the analysis of neuronal microcircuitry. The figure was created by the authors using data from reference [10].

https://doi.org/10.1371/journal.pone.0225953.g004
Listing 5. CSD model depiction of an astronomy image dataset. JSON serialized listing of the astronomy dataset describing the bubble nebula observed at 656 nm wavelength. The listing was created by the authors using data from reference [11].

```
bubble.csdfe
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2016-02-26T16:41:00Z",
    "tags": ["Bubble Nebula","Bubble"],
    "description": "The dataset is a new observation of the Bubble Nebula acquired by The Hubble Heritage Team, in February 2016."
  },
  "dimensions": [
    {
      "type": "linear",
      "count": 11596,
      "increment": "-2.27930619e-05 m",
      "coordinates_offset": "350.311874957 °",
      "quantity_name": "plane angle",
      "label": "Right Ascension"
    },
    {
      "type": "linear",
      "count": 11351,
      "increment": "1.10055218e-05 m",
      "coordinates_offset": "61.12851495 °",
      "quantity_name": "plane angle",
      "label": "Declination"
    }
  ],
  "dependent_variables": [
    {
      "type": "external",
      "name": "Bubble Nebula, 656nm",
      "numeric_type": "float32",
      "components_url": "file:./Bubble_1.dat"
    }
  ]
}
```

``satRec.csdf``. A monotonic dimension is employed when measurements are not uniformly spaced or span several orders of magnitude along a dimension. An example of a 2D(1) dataset with a monotonic dimension, acquired by the authors, is given in Listing 6. Here the dataset comes from a $^{29}$Si NMR magnetization recovery measurement of a highly siliceous ZSM-12 zeolite sampled on a 2D rectilinear grid. Fig 6 depicts a stacked plot corresponding to the dataset from Listing 6.

The value of the dimensions key is an array with two JSON serialized Dimension objects (lines 18-30 and 31-38). The first is a LinearDimension object, labeled as $t_2$, describing a temporal dimension with 1024 points sampled at every 80 μs with a coordinates_offset of −41.04 ms. Additionally, this LinearDimension object contains a ReciprocalDimension object serialized as the value of the reciprocal key. The second is a MonotonicDimension object, labeled as $t_1$, with the coordinates associated with grid
vertices along the dimension explicitly given in the ordered set of values in the `coordinates` key. The value of the `dependent_variables` key is an array with a single JSON serialized `InternalDependentVariable` object (lines 41-47) describing the signal response. Here, the data values are encoded as an array with one Base64 string in the `components` key.

This listing also gives an example of the use of the `application` key in the `csdm` dictionary. Here an application owning the domain name `physyapps.com` has placed an attribute in the `application` dictionary using the reverse domain name key `com.physyapps.rmn`. Domain name owners are free to place any valid JSON object as the value of their respective reverse domain name attribute inside the `application` dictionary. In this case, the domain name owner has used the reverse domain name key `com.physyapps.rmn` to place a dictionary holding two keys, `focus` and `dimension_precedence`.

An `application` key can also be placed in any `Dimension`, `ReciprocalDimension`, `DependentVariable`, and `SparseSampling` object. Again, according to the rule in section 2.5, only the reverse domain name owner has permission to serialize a file using their respective reverse domain name as a key in the `application` attribute.

`iglu_1d.csdf`. Listing 7 is a 2D[1] example of an NMR signal shown in Fig 7 with sparse sampling along one dimension [12]. Here the `InternalDependentVariable` object (lines 30-44) holds a `SparseSampling` object (lines 37-43) in the `sparse_sampling` key. The `SparseSampling` object contains three keys `dimension_indexes`, `sparse_grid_vertexes`, and `unsigned_integer_type`. The `dimension_indexes` key holds an array of integers specifying the indexes of the dimensions along which the dependent variable is sparsely sampled, in this case, the $k=1$ dimension. The `sparse_grid_vertexes` key holds an array of integers specifying the vertexes on the one-dimensional sparsely sampled grid. Since there are two dimensions in the dataset the array of integers corresponds to the coordinate indexes, $j_1$, along the $k=1$ dimension. In this example, the dependent variable values are fully sampled along the $k=0$ dimension. The value of the `unsigned_integer_type` key holds the numeric type used in importing the integer array from `sparse_grid_vertexes`.

---

Fig 5. CSD model depiction of an astronomy image dataset. A log intensity plot, derived from Listing 5, of the bubble nebula [11] observed at 656 nm wavelength. The figure was created by the authors using data from reference [11].

https://doi.org/10.1371/journal.pone.0225953.g005
Listing 6. CSD model depiction of a 2-D NMR dataset. JSON serialized listing of $^{29}$Si NMR magnetization saturation relaxation dataset containing one single-component dependentVariable object and two Dimension objects.

```json
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2016-03-12T16:41:00Z",
    "tags": ["$^{29}$Si", "NMR", "nuclear magnetism relaxation", "zeolite ZSM-12"],
    "description": "A $^{29}$Si NMR magnetization saturation recovery measurement of highly siliceous zeolite ZSM-12."
  },
  "application": {
    "com.physyapps.rnm": {
      "focus": {
        "mem_offset": 166,
        "component_index": 0,
        "dependent_variable_index": 0
      },
      "dimension_precedence": [0,1]
    }
  },
  "dimensions": [
    {
      "type": "linear",
      "count": 1024,
      "increment": "0.08 ms",
      "coordinates_offset": "-41.04 ms",
      "label": "t2",
      "reciprocal": {
        "origin_offset": "79.578822262 MHz",
        "coordinates_offset": "-8.7660626 kHz",
        "quantity_name": "frequency",
        "label": "$^{29}$Si nuclear magnetic resonance"
      }
    },
    {
      "type": "monotonic",
      "coordinates": [
        "1 s", "5 s", "10 s", "20 s", "40 s", "80 s"
      ],
      "label": "t1",
      "quantity_name": "time"
    }
  ],
  "dependent_variables": [
    {
      "type": "internal",
      "quantity_type": "scalar",
      "numeric_type": "complex64",
      "encoding": "base64",
      "components": ["AEU2Q7h/...y0NGPyJD"]
    }
  ]
}
```
Listing 8 is a 2D example of an NMR signal shown in Fig 8 with sparse sampling along two dimensions [12]. As before, the sparse_sampling key holds a \texttt{SparseSampling} object with the dimension_indexes, \texttt{sparse_grid_vertexes}, and \texttt{unsigned_integer_type} attributes. The dimension_indexes key holds an array of two integers, \( k = 0 \) and \( 1 \), specifying the sparse sampling dimensions. The \texttt{sparse_grid_vertexes} key holds an array of integers defining the vertexes on the two-dimensional sparsely sampled grid. As described in section 2.4.1 this array is a flattened ordered set of arrays which can be reshaped into the ordered set of sparse grid vertexes, i.e.,

\[ [0, 0, 1, 0, \ldots, 972, 511, 1015, 511] \rightarrow [[0, 0], [1, 0], \ldots, [972, 511], [1015, 511]] \]

The \( i \)-th vertex in the ordered set of sparse grid vertexes specifies the sparse grid location of the \( i \)-th value in each component array of the dependent variable.

### 3.3 2D example

**RGB_image.csdf.** A simple example of a 2D dataset with multiple components is a color image [13], such as the one shown in Fig 9. This is a 2D dataset, with two \texttt{LinearDimension} objects and one three-component dependent variable, \( p = 3 \). The CSDM serialization is shown in Listing 9. The \texttt{dimensions} key holds an array with two JSON serialized \texttt{LinearDimension} objects (lines 8-13 and 14-19) with 1024 and 768 points, respectively, and a unit sampling interval. The \texttt{dependent_variable} key holds an array with a single JSON serialized \texttt{InternalDependentVariable} object (lines 22-34) containing an image dataset as indicated by the \texttt{pixel_3} value of the \texttt{quantity_type} key. The first part, \texttt{pixel}, indicates pixel data, and the last part, \texttt{3}, gives the number of pixel components. An array holding to the three components, i.e., the red, green, and blue color intensities with each encoded as a Base64 string, is the value of the \texttt{components} key. The Base64 decoded binary data values are then interpreted as an array of 8-bit unsigned integer (\texttt{uint8}), for each component, and
subsequently mapped onto a 1024 × 768 coordinate grid. The value of the component_labels key is an array of the labels ordered to match the order of the components.

Listing 7. CSD model depiction of a sparse NMR dataset with one sparse dimension. JSON serialized listing of $^{13}$C-$^{15}$N NMR HSQC dataset containing one single-component DependentVariable object and two Dimension objects. The listing was created by the authors using data from reference [12].

```json
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-07-07T23:41:50Z",
    "dimensions": [
      {
        "type": "linear",
        "count": 1024,
        "increment": "0.192 ms",
        "label": "1H t2",
        "quantity_name": "time",
        "reciprocal": {
          "origin_offset": "400.13 MHz",
          "quantity_name": "frequency",
          "coordinates_offset": "-3.32 Hz"
        }
      },
      {
        "type": "linear",
        "count": 512,
        "increment": "0.192 ms",
        "label": "1H t1",
        "quantity_name": "time",
        "reciprocal": {
          "origin_offset": "400.13 MHz",
          "quantity_name": "frequency",
          "coordinates_offset": "-3.32 Hz"
        }
      }
    ],
    "dependent_variables": [
      {
        "quantity_name": "dimensionless",
        "numeric_type": "complex64",
        "components_url": "cos.data",
        "quantity_type": "scalar",
        "type": "external",
        "name": "cos",
        "sparse_sampling": {
          "dimension_indexes": [1],
          "sparse_grid_vertices": [
            0, 1, ..., 470, 499
          ],
          "unsigned_integer_type": "uint16"
        }
      }
    ]
  }
}
```
3.4 3D[2] example

`wind_velocity.csdfe`. An example of a 3D[2] dataset, i.e., with three dimensions, $d = 3$, and one two-component dependent variable, $p = 2$, is the wind velocity prediction [14] dataset as a function of latitude, longitude and time, shown in Listing 10.

The value of the dimensions key is an array with three JSON serialized Dimension objects. The first two LinearDimension objects, labeled as longitude and latitudes respectively, describe two linear dimensions sampled at every 0.5˚ for 49 points starting at $-102.5˚$ longitudes and $13.5˚$ latitudes. Together, these two objects create a two-dimensional grid that spans the region around the Gulf of Mexico as depicted in Fig 10. The third dimension is a LabeledDimension object as indicated by the value of the type key. The corresponding labels array lists six date-time stamps entries.

The value of the dependent_variable key is an array with a single JSON serialized ExternalDependentVariable object (lines 34-43) containing a two-component vector dataset as identified by the quantity_type key-value. This value is vector_2 where the first part, vector, indicates vector data, and the last part, 2, gives the number of vector components. The two vector components are labeled as ugrd10m-eastward_wind and vgrd10m-northward_wind, in the array assigned to the component_labels key. The data values are located in an external file as a binary data whose address, relative to the `wind_velocity.csdfe` file, is the value of the components_url key. The binary data is interpreted as a 32-bit floating-point numerical array. Note, because the binary data does not support array indexing, unlike JSON serialization, the corresponding numerical array of data values is reshaped into a matrix which includes the number of components. In this case, the reshaped matrix is $49 \times 49 \times 6 \times 2$, where the last number is the number of components, $p = 2$, and the remaining three is the number of points from the Dimension objects. Table 3 contains a description of the number of components, $p$, for each quantity_type.
Listing 8. CSD model depiction of a sparse NMR dataset with two sparse dimensions. JSON serialized listing of $^1$H NMR TOCSY dataset containing two Dimension objects and one single-component DependentVariable object with sparsely sampled values in both dimensions. The listing was created by the authors using data from reference [12].

```json
{  
  "csdm": {  
    "version": "1.0",  
    "timestamp": "2019-07-14T15:13:09Z",  
    "dimensions": [  
      {  
        "quantity_name": "time",  
        "type": "linear",  
        "count": 1024,  
        "increment": "0.192 ms",  
        "label": "$^1$H t2",  
        "reciprocal": {  
          "origin_offset": "400.13 MHz",  
          "quantity_name": "frequency",  
          "coordinates_offset": "-3.32 Hz"  
        }  
      },  
      {  
        "quantity_name": "time",  
        "type": "linear",  
        "count": 512,  
        "increment": "0.192 ms",  
        "label": "$^1$H t1",  
        "reciprocal": {  
          "origin_offset": "400.13 MHz",  
          "quantity_name": "frequency",  
          "coordinates_offset": "-3.32 Hz"  
        }  
      }  
    ],  
    "dependent_variables": [  
      {  
        "type": "external",  
        "quantity_name": "dimensionless",  
        "numeric_type": "complex64",  
        "quantity_type": "scalar",  
        "components_url": "cos.data",  
        "sparse_sampling": {  
          "dimension_indexes": [0, 1],  
          "sparse_grid_vertices": [  
            0, 0, 1, 0, ..., 972, 511, 1015, 511  
          ],  
          "unsigned_integer_type": "uint16"  
        }  
      }  
    ]
  }
}
```

Fig 10 depicts a quiver plot of the wind velocity at three different date-time stamps. Underlaid these plots is a map of the Earth corresponding to the given range of latitudes and longitudes. These plots were generated using the Matplotlib library [16] for python in addition to the Matplotlib Basemap toolkit [15] for rendering maps.
3.5 3D\{6\} example

brain_MRI.csdf. A 3D\{6\} dataset has three dimensions, \(d = 3\), and one six-component dependent variable, \(p = 6\). An example of such a dataset is the second rank symmetric diffusion tensor MRI dataset \([17]\) of a brain given in Listing 11.

The value of the dimensions key is an array with three JSON serialized Dimension objects describing the three spatial dimensions, labeled as \(x\), \(y\), and \(z\) respectively. Here, all objects describe a linear dimension with the sampling resolution of 1 mm, and 148, 190 and 160 points along the respective dimension.

The value of the dependent_variables key is an array with a single JSON serialized InternalDependentVariable object (lines 30-51) describing a symmetric matrix dataset as indicated by the value of the quantity_type key. The value symmetric_matrix_3 emphasizes a six-component dataset as noted in Table 3. The six components, labeled as \(D_{xx}\), \(D_{xy}\), \(D_{xz}\), \(D_{yy}\), \(D_{yz}\), and \(D_{zz}\) respectively, are stored as Base64 strings as the value of the components key. Each Base64 decoded binary array is interpreted as 32-bit floating-points array and subsequently reshaped to a \(148 \times 190 \times 160\) matrix.

The symmetric matrix data from the brain_MRI.csdf file was partially processed as a second-rank symmetric diffusion tensor to determine the isotropic diffusion coefficients. The intensity plots in Fig 11 depicts the projection of the isotropic diffusion coefficients on to the three spatial dimensions.

Listing 9. CSD model depiction of a RGB image dataset. JSON serialized listing of an RGB image dataset containing two Dimension objects, and one DependentVariable object with three components corresponding to red, green and blue color intensities. The listing was created by the authors using the data \([13]\) available under (Creative common 0) CC0 license.
Fig 9. CSD model depiction of a RGB image dataset. An image plot, derived from Listing 9, of an RGB dataset depicting a raccoon face. Photo credit: Judy Weggelaar.

https://doi.org/10.1371/journal.pone.0225953.g009
Listing 10. CSD model depiction of a meteorology vector dataset. JSON serialized listing of the predicted wind velocities over and around the Gulf of Mexico. The model contains one two-component `DependentVariable` object and three `Dimension` objects. Listing was created by the authors using data from the national centers for environment information/national oceanic and atmospheric administration [14].

```json
wind_velocity.csdfe
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2018-12-12T10:00:00Z",
    "tags": ["wind velocity", "weather forecast"],
    "description": "Dataset from NOAA/NCEP Global Forecast System (GFS) Atmospheric Model. Latitudes and longitudes defined in geographic coordinate system.",
    "dimensions": [
      {
        "type": "linear",
        "label": "longitude",
        "count": 49,
        "coordinates_offset": "-102.5 °W",
        "increment": "0.5 °W"},
      {
        "type": "linear",
        "label": "latitude",
        "count": 49,
        "coordinates_offset": "13.5 °N",
        "increment": "0.5 °N"},
      {
        "type": "labeled",
        "description": "Forecast time for ForecastModelRunCollection. Data time stamps defined with ISO 8601 format.",
        "label": "UTC date-time stamp",
        "labels": [
          "2018-12-12T12:00:00Z",
          "2018-12-12T18:00:00Z",
          "2018-12-13T00:00:00Z",
          "2018-12-13T06:00:00Z",
          "2018-12-13T12:00:00Z",
          "2018-12-13T18:00:00Z"]
    }
  },
  "dependent_variables": [
    {"type": "external",
     "description": "Component labels are standard attribute names used by Dataset Attribute Structure (.das). 'ugrd10m' is 'eastward wind velocity at 10 m above ground level', and 'vgrd10m' is 'northward wind velocity at 10 m above ground level'.",
     "components_url": "file:/NCEP_Global.dat",
     "component_labels": ["ugrd10m-eastward_wind", "vgrd10m-northward_wind"],
     "quantity_type": "vector_2",
     "name": "Wind velocity dataset",
     "numeric_type": "float32",
     "quantity_name": "speed",
     "unit": "m/s"}
  ]
}
```
Fig 10. CSD model depiction of a meteorology vector dataset. A quiver plot, derived from Listing 10, of the wind velocities from the dataset in Listing 10 at three different date-time stamps. The underlaid map of the Earth corresponding to the latitudes and longitudes is rendered using the Matplotlib Basemap toolkit [15]. The figure was created by the authors using data from reference [14].

https://doi.org/10.1371/journal.pone.0225953.g010
Listing 11. CSD model depiction of an MRI tensor dataset. JSON serialized listing of the diffusion tensor MRI dataset [17] of the brain containing one six-component **Dependent-Variable** object and three **Dimension** objects. Listing was created by the authors using data from reference [17].

```json
brain_MRI.csdf
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2016-03-12T16:41:00Z",
    "tags": ["MRI", "image", "brain", "diffusion tensor"],
    "dimensions": [
      {
        "type": "linear",
        "label": "x",
        "quantity_name": "length",
        "count": 148,
        "increment": "1.0 mm"
      },
      {
        "type": "linear",
        "label": "y",
        "quantity_name": "length",
        "count": 190,
        "increment": "1.0 mm"
      },
      {
        "type": "linear",
        "label": "z",
        "quantity_name": "length",
        "count": 160,
        "increment": "1.0 mm"
      }
    ],
    "dependent_variables": [
      {
        "type": "internal",
        "quantity_type": "symmetric_matrix_3",
        "encoding": "base64",
        "numeric_type": "float32",
        "components": [
          "kIa/PisXkEAwp+4/PaSMQHsG...",
          "aCp1vVvYcA4kOe/QpCkv2Jq...",
          "AAAAAAAAAAAAAAAAAAAAAAAAAAFp/...",
          "4161wqjg2z0+I4DBzHZ2WbW0...",
          "pJ1UPy3pD0AaKo+i1F4PoyQ...",
          "kIa/v4XXkAwp+6/PaSNwDn4...
        ],
        "component_labels": [
          "Dxx",
          "Dxy",
          "Dxz",
          "Dyy",
          "Dyz",
          "Dzz"
        ]
      }
    ]
  }
}
```
3.6 2D{1,1,2,1,1} example

An example of a 2D{1,1,2,1,1} dataset using data from the US National Centers for Environment Information / National Oceanic and Atmospheric Administration [14] is given in Listing 12.

In this example, the value of the `dimensions` key is an array with two `LinearDimension` objects describing the two spatial dimensions, labeled as `longitude` and `latitude`, respectively. The value of the `dependent_variables` key is an array with five `ExternalDependentVariable` objects describing the surface temperature ($p_0 = 1$), the air temperature at 2 m above ground level ($p_1 = 1$), the two-component wind velocity vector at 10 m above surface ($p_2 = 2$), the relative humidity ($p_3 = 1$), and the air pressure at the sea level ($p_4 = 1$). Fig 12 depicts the intensity and quiver plots of four dependent variables.

3.7 0D{1,1} example

J_vs_s.csdf. The CSD model also allows the serialization of datasets without a coordinate grid. A 0D{1,1} dataset, for example, has no dimensions, $d = 0$, and two single-component dependent variable, $p_0 = 1$ and $p_1 = 1$. The listing for such a dataset [18] is given in Listing 13. In this example, the two “correlated” dependent variables are the $^2$J nuclear spin couplings, across a Si-O-Si linkage and the $s$-character product on the O and two Si along the Si-O bond across the Si-O-Si linkage [18]. The value of the `dependent_variables` key is an array with two JSON serialized `InternalDependentVariable` object (lines 7-16 and 17-27). The first object, named as Gaussian computed J-couplings, describes the $^2$J couplings. The data values are stored as a Base64 string in the `components` key. The Base64 decoded binary array is interpreted as a 32-bit floating-point numerical array following the value of the `numeric_type` key. The second object is named as the product of $s$-characters. Here, the data values are again stored as a Base64 string, which after decoding is interpreted as a 32-bit floating-point numerical array. A scatter plot revealing the correlation between the two dependent variables from the dataset in Listing 13 is presented in Fig 13.
Listing 12. CSD model depiction of a meteorology dataset with multiple dependent variables. JSON serialized listing of multiple dependent variables including scalar and vector on a two-dimensional grid. Listing was created by the authors using data from the US National Centers for Environment Information / National Oceanic and Atmospheric Administration [14].

```json
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2017-09-17T12:00:00Z",
    "read_only": true,
    "description": "Dataset from NOAA/NCEP Global Forecast System (GFS) Atmospheric Model."
  },
  "dimensions": [
    {
      "type": "linear",
      "count": 192,
      "increment": "0.5 degree",
      "coordinates_offset": "264.0 degree",
      "quantity_name": "plane angle",
      "label": "longitude"
    },
    {
      "type": "linear",
      "count": 89,
      "increment": "0.5 degree",
      "coordinates_offset": "53.0 degree",
      "quantity_name": "plane angle",
      "label": "latitude"
    }
  ],
  "dependent_variables": [
    {
      "type": "external",
      "name": "Surface air temperature",
      "unit": "K",
      "quantity_name": "temperature",
      "component_labels": ["tempsfc - surface air temperature"],
      "numeric_type": "float64",
      "quantity_type": "scalar",
      "components_url": "file://surface_temp.dat"
    },
    {
      "type": "external",
      "name": "Air temperature at 2 m",
      "unit": "K",
      "quantity_name": "temperature",
      "component_labels": ["tmp2m - air temperature at 2m"],
      "numeric_type": "float64",
      "quantity_type": "scalar",
      "components_url": "file://temp2m.dat"
    },
    {
      "type": "external",
      "name": "Wind velocity at 10 m",
      "unit": "m/s",
      "quantity_name": "speed",
      "component_labels": ["ugrd10m - eastward wind velocity at 10m", "vgrd10m - northward wind velocity at 10m"],
      "numeric_type": "float64",
      "quantity_type": "vector_2",
      "components_url": "file://wind_velocity.dat"
    },
    {
      "type": "external",
      "name": "Relative humidity",
      "unit": "\%",
      "quantity_name": "dimensionless",
      "component_labels": ["rh2m - relative humidity at 2m"],
      "numeric_type": "float64",
      "quantity_type": "scalar",
      "components_url": "file://relative_humidity.dat"
    }
  ]
}
```
Listing 13. CSD model depiction of a computational dataset. JSON serialized listing of quantum chemistry calculation of nuclear spin-spin coupling constant between $^{29}$Si nuclei across a Si-O-Si linkage in small cluster molecule. An example dataset with two Dependent-Variable objects and no Dimension objects. The listing was created by the authors using data from reference [18].

```json
J_vs_e_csd
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2016-08-12T16:41:00Z",
    "tags": ["magnetic resonance", "$^{29}$Si", "nuclear J coupling", "computational chemistry", "silicates"],
    "dependent_variables": [ {
      "type": "internal",
      "name": "Gaussian computed J-couplings ",
      "unit": "Hz",
      "component_labels": ["J-coupling"],
      "quantity_type": "scalar",
      "numeric_type": "float32",
      "encoding": "base64",
      "components": [ "Btjrv1/vtr/+ ... w2TJQYV80EE=" ]
    },
    {
      "type": "internal",
      "name": "product of s-characters",
      "unit": "%",
      "component_labels": ["s-character product"],
      "quantity_type": "scalar",
      "numeric_type": "float32",
      "encoding": "base64",
      "components": [ "xLIYP6N5Wj+6 ... +oDP3mOwz8=" ]
    } ]
  }
}
```

Fig 12. CSD model depiction of a meteorology dataset with multiple dependent-variables. The figure depicts (a) an intensity plot of the air temperature at 2 m above surface, (b) a quiver plot of the wind vectors at 10 m above surface, (c) an intensity plot of the relative humidity, and (d) an intensity of the air pressure at sea level corresponding to the last four dependent-variables from Listing 12. These plots are overlaid on the coastline map of the Earth corresponding to the latitude and longitudes. These coastline were rendered using the Matplotlib Basemap toolkit [15]. The plots were generated using the Matplotlib library [16] for python. The figure was created by the authors using data from reference [14].

https://doi.org/10.1371/journal.pone.0225953.g012
4 Conclusions

We have designed the Core Scientific Dataset (CSD) Model as a lightweight, portable, versatile, resourceful, and standalone data model that is capable of handling multi-dimensional and correlated datasets from various spectroscopies, diffraction, microscopy, and imaging techniques. A guiding principle in the design of this model was to encapsulate only the minimal metadata necessary to represent the correlated datasets sampled on a common orthogonal coordinate grid. The model also allows for sparse sampling on this grid. Throughout the model, we make use of the `ScalarQuantity` class, which is composed of a numerical value and any valid SI unit symbol or any number of accepted non-SI unit symbols. This approach enables tremendous flexibility in allowing the dataset model to be agnostic of the scientific domain. Historically, this may have been perceived as a potential barrier to software implementation of the CSDM, however, in recent years libraries capable of parsing units have become freely available for various computing environments such as Matlab, Mathematica, and python. The CSD model is independent of the hardware, operating system, application software, and file-serialization method used for data exchange. The model provides a mechanism for the inclusion of additional application-specific metadata without compromising its fundamental role as a data exchange and archiving standard. When serialized using JSON serialization the resulting file format is human readable and integrable with most object-oriented programming languages and software. The serialization of the CSD model has been adopted as an open dataset file format in NMR software development under our control, i.e., SIMPSON [19, 20], DMFIT [21], jsNMR [22], and RMN [23], which already have a large installed user base within the solid-state NMR scientific community. We envision the CSD model and its associated file format as playing an important role in community accessible databases and in greater data-trail integrity and compliance issues for many research laboratories.

Fig 13. CSD model depiction of a computational dataset. Two dependent variables [18] correlating $J_{\text{Si-O-Si}}$ couplings to the corresponding product of s-characters on the Si, O and Si atoms along the Si-O bond across the Si-O-Si linkage. The figure was created by the authors using data from reference [18].

https://doi.org/10.1371/journal.pone.0225953.g013
Appendix

Scaled variables

Coordinates along a dimension can also be converted into scaled quantities based on other attributes in the Dimension object or in application meta-data. For example, in nuclear magnetic resonance spectroscopy, the spectra are conventionally plotted as a function of a dimensionless frequency ratio. In CSD model, the origin_offset, $o_k$, is interpreted as the NMR spectrometer frequency and the coordinates_offset, $b_k$, as the reference frequency. Given the dimension coordinate, $X_k$, from Eq (3), the corresponding dimensionless-coordinate ratio follows,

$$X_{\text{ratio}} = \frac{X_k}{o_k - b_k}.$$  \hspace{1cm} (9)

csdmpy

The csdmpy module is the Python support for the core scientific dataset (CSD) model file-exchange format. The source code is available at https://github.com/DeepanshS/csdmpy and the corresponding documentation at https://csdmpy.readthedocs.io/en/stable, which includes links for downloading the CSDM compliant files used in this report.

The main objective of this python module is to facilitate the import and export of the CSD model serialized files for Python users. Moreover, the module utilizes Python libraries such as Numpy and therefore allowing the end-users to process or visualize the imported datasets with any third-party package(s) compatible with Numpy.

Supporting information

S1 File. Additional CSDM examples and review of units and constants as used in the CSD model.

(PDF)

Acknowledgments

This material is based upon work supported in part by the US National Science Foundation under Grant No. DIBBS OAC 1640899. T.V. acknowledges financial support from the European Commission (H2020 Future and Emerging Technologies Grant 731475 and Research Infrastructures Grant 731005). Early support for this work came from the Le Studium Loire Valley Institute for Advanced Studies, Orléans, France.

Author Contributions

Conceptualization: Deepansh J. Srivastava, Thomas Vosegaard, Philip J. Grandinetti.

Formal analysis: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

Funding acquisition: Dominique Massiot, Philip J. Grandinetti.

Investigation: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

Methodology: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.
Project administration: Philip J. Grandinetti.

Resources: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

Software: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

Supervision: Philip J. Grandinetti.

Validation: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

Visualization: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

Writing – original draft: Deepansh J. Srivastava, Philip J. Grandinetti.

Writing – review & editing: Deepansh J. Srivastava, Thomas Vosegaard, Dominique Massiot, Philip J. Grandinetti.

References

1. ECMA. Standard ECMA-404: The JSON Data Interchange Syntax; 2017. Available from: https://www.ecma-international.org/publications/standards/Ecma-404.htm.

2. Fowler M. UML Distilled, A Brief Guide to the Standard Object Modeling Language. Boston: Addison-Wesley; 2004.

3. Thompson A, Taylor BN. Guide for the use of the International System of Units (SI); 2008. Available from: https://nvlpubs.nist.gov/nistpubs/SP/nistspecialpublication811e2008.pdf.

4. Consortium WWW. Architecture of the World Wide Web, Volume One; 2004. Available from: http://www.w3.org/TR/webarch/.

5. PythonWare. Python Imaging Library (PIL):. Available from: http://www.pythonware.com/products/pil/.

6. Collaboration A, Robitaille TP, Tollerud EJ, Greenfield P, Droettboom M, Bray E, et al. Astropy: A community Python package for astronomy. aap. 2013; 558:A33.

7. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science & Engineering. 2011; 13:22–30. https://doi.org/10.1109/MCSE.2011.37

8. Church JA, White NJ. Sea-Level Rise from the Late 19th to the Early 21st Century. Surveys in Geophysics. 2011; 32:585–602. https://doi.org/10.1007/s10712-011-9119-1

9. Lancashire RJ. JCamp-DX; 2006. Available from: http://www.chem.uwimona.edu.jm/spectra/index.html.

10. Cardona A, Saalfeld S, Preibisch S, Schmid B, Cheng A, Pulokas J, et al. An Integrated Micro- and Macro architectural Analysis of the Drosophila Brain by Computer-Assisted Serial Section Electron Microscopy. PLoS Biology. 2010; 8:e1000502. https://doi.org/10.1371/journal.pbio.1000502 PMID: 20957184

11. The Hubble Heritage Project; 2016. Available from: https://archive.stsci.edu/prepds/heritage/bubble/introduction.html.

12. Balsgart NM, Vosegaard T. Fast Forward Maximum entropy reconstruction of sparsely sampled data. J Magn Reson. 2012; 223:164–169. https://doi.org/10.1016/j.jmr.2012.07.002 PMID: 22975245

13. Weggeala J.; Available from: https://pixnio.com/fauna-animals/raccoons/raccoon-procyon-lotor.

14. ERDDAP: Marine Domain Awareness (MDA) ERDDAP Server—JRC Italy.; Available from: http://mda.marine.ie/erddap/gridap/NCEP_Global_Best.html.

15. Whitaker J. Matplotlib Basemap Toolkit; 2011. Available from: https://matplotlib.org/basemap/.

16. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007; 9:90–95. https://doi.org/10.1109/MCSE.2007.55

17. Diffusion tensor MRI datasets; 2000. Available from: http://www.sci.utah.edu/~gk/DTI-data/.

18. Srivastava DJ, Florian P, Baltisserber JH, Grandinetti PJ. Correlating geminal JSi-O-Si couplings to structure in framework silicates. Phys Chem Chem Phys. 2018; 20:562–571. https://doi.org/10.1039/C7CP06486A
19. Bak M, Rasmussen JT, Nielsen NC. SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy. J Magn Reson. 2000; 147:296–330. https://doi.org/10.1006/jmre.2000.2179 PMID: 11097821

20. Tossner Z, Andersen R, Stevenson B, Edén M, Nielsen NC, Vosegaard T. Computer-intensive simulation of solid-state NMR experiments using SIMPSON. J Magn Reson. 2014; 246:79–93. https://doi.org/10.1016/j.jmr.2014.07.002 PMID: 25093693

21. Massiot D, Fayon F, Capron M, King I, Le Calvé S, Alonso B, et al. Modelling one- and two-dimensional solid-state NMR spectra. Magn Reson Chem. 2002; 40:70–76. https://doi.org/10.1002/mrc.984

22. Vosegaard T. jsNMR: an embedded platform-independent NMR spectrum viewer. Magn Reson Chem. 2015; 53:285–290. https://doi.org/10.1002/mrc.4195 PMID: 25641013

23. PhySy Ltd. RMN 2.0; 2019. Available from: https://www.physyapps.com/rmn.
Supplemental Material for
“Core Scientific Dataset Model: A light weight and portable model and file format for multi-dimensional scientific data”

Deepansh J. Srivastava¹, Thomas Vosegaard², Dominique Massiot³, and Philip J. Grandinetti¹

¹Department of Chemistry, Ohio State University, 100 West 18th Avenue, Columbus, OH 43210, USA,

²Laboratory for Biomolecular NMR Spectroscopy, Department of Molecular and Structural Biology, University of Aarhus, DK-8000 Aarhus C, Denmark,

³CNRS, UPR3079 CEMHTI, 1D Avenue de la Recherche Scientifique, 45071 Orléans Cedex 2, France

November 22, 2019

In section 1 of this supplement, we give additional serialization examples of the CSD model. In section 2 we give a brief overview of concepts associated with the international system of units[1], along with tables giving unit symbol conventions as well as symbol suggestions for various physical constants.

S1 Additional CSDM Examples

Listings 1, 2, 3, and 4 further illustrate the versatility of the CSD model for 1D{1} datasets from different scientific domains. Here, we give examples from the Fourier-transform infrared (FTIR) spectroscopy, Ultraviolet-visible (UV-vis) spectroscopy, Electron paramagnetic resonance (EPR) spectroscopy, and Gas Chromatography, respectively. These datasets were obtained from http://wwwchem.uwimona.edu.jm/spectra/specindex.html (Prof. Robert J. Lancashire) and converted from JCAMP to the CSD model.

In Listing 5, we show a 2D{1} NMR dataset from a phase-adjusted spinning sideband (PASS) measurement[2] where the two LinearDimension objects describe a 2D coordinate grid of time and rotor phase. In Listing 6 we show an example of a 3D{1} NMR dataset for a phase incremented echo train acquisition (PIETA) measurement[3]. Additional details are given in the respective listing captions.
Listing 1: 1D(1) example dataset of the FTIR spectrum of caffeine with one LinearDimension object and one DependentVariable object. The dimension is sampled every 1.9305486 cm\(^{-1}\) for 1842 points starting at 449.41 cm\(^{-1}\). Here, the data values in the InternalDependentVariable object are stored as a Base64 encoded string of a 32-bit floating point array.

```
caffeine.csdf
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-07-01T21:03:42Z",
    "read_only": true,
    "tags": [
      "infrared spectrum",
      "caffeine"
    ],
    "dimensions": [
      {
        "type": "linear",
        "count": 1842,
        "increment": "1.9305486 cm\(^{-1}\)",
        "coordinates_offset": "449.41 cm\(^{-1}\)",
        "quantity_name": "wavenumber",
      }
    ],
    "dependent_variables": [
      {
        "type": "internal",
        "name": "Caffeine",
        "encoding": "base64",
        "numeric_type": "float32",
        "quantity_type": "scalar",
        "component_labels": ["Transmittance"],
        "components": ["/p7GQg...J5dhC"]
      }
    ]
  }
}
```

Figure S1: CSD model depiction of a FTIR dataset.
Listing 2: 1D(1) example dataset of the UV-vis spectrum of benzene vapor with one LinearDimension object and one DependentVariable object. The dimension is sampled every 0.01 nm for 4001 data points starting at 230 nm. The data values from the ExternalDependentVariable object are stored as 32 bit floating point binary array in an external file whose URL relative to the benzene.csdfe file is specified by the value of the components_url key.

```
benzene.csdfe
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-07-01T21:03:42Z",
    "read_only": true,
    "tags": [
      "uv-visible spectrum",
      "gas phase",
      "benzene"
    ],
    "dimensions": [
      {
        "type": "linear",
        "count": 4001,
        "increment": "0.01 nm",
        "coordinates_offset": "230.0 nm",
        "quantity_name": "wavelength"
      }
    ],
    "dependent_variables": [
      {
        "type": "external",
        "name": "Vapor of Benzene",
        "numeric_type": "float32",
        "quantity_type": "scalar",
        "component_labels": ["Absorbance"],
        "components_url": "file:./benzeneVap.dat"
      }
    ]
  }
}
```

Figure S2: CSD model depiction of a UV-Vis dataset.
Listing 3: 1D\{1\} example dataset of the EPR spectrum of vanadium in the mushroom *Amanita muscaria*[4] with one *InternalDependentVariable* and one *LinearDimension* object. The dimension is sampled for 298 points at every 4 G starting at 2750 G. The data values in the *InternalDependentVariable* object are stored as a Base64 encoded string of a 32-bit floating points array.

```json
AmanitaMuscaria.csdf
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-07-01T21:03:42Z",
    "read_only": true,
    "tags": [
      "epr spectrum",
      "Vanadyl",
      "Amanita muscaria"
    ],
    "dimensions": [
      {
        "type": "linear",
        "count": 298,
        "increment": "4.0 G",
        "coordinates_offset": "2750.0 G",
        "quantity_name": "magnetic flux density"
      }
    ],
    "dependent_variables": [
      {
        "type": "internal",
        "name": "Vanadyl in Amanita muscaria",
        "encoding": "base64",
        "numeric_type": "float32",
        "quantity_type": "scalar",
        "component_labels": ["Intensity Derivative"],
        "components": ["TDeJPZ...kMvg=="]
      }
    ]
  }
}
```

![Figure S3: CSD model depiction of an EPR dataset.](image-url)
Listing 4: 1D[1] example dataset of the gas chromatograph of cinnamon with one `InternalDependentVariable` and one `LinearDimension` object. The data values in the `InternalDependentVariable` object are stored as a Base64 encoded string of a 32-bit floating points array.

cinnamon.csdf
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-07-01T21:03:42Z",
    "read_only": true,
    "tags": [
      "gas chromatography",
      "cinnamon"
    ],
    "dimensions": [
      {
        "count": 6001,
        "increment": "0.0034 min",
        "quantity_name": "time"
      }
    ],
    "dependent_variables": [
      {
        "components": ["AEU9Rw...g7Rw=="],
        "encoding": "base64",
        "numeric_type": "float32",
        "quantity_type": "scalar",
        "name": "Headspace from cinnamon stick",
        "component_labels": ["FID response"]
      }
    ]
  }
}

Figure S4: CSD model depicting of a gas chromatography dataset.
Listing 5: A 2D{1} example dataset of the NMR phase-adjusted spinning sideband (PASS) measurement\(^2\) where the two LinearDimension objects describe a 2D coordinate grid of time and rotor phase. The first LinearDimension describes a linearly sampled time dimension with 128 points acquired at every 10 \(\mu s\). The second LinearDimension describes a linearly sampled rotor phase / 2 \(\pi\) dimension with 32 points sampled at every 0.03125 tr. The symbol “tr” represents a unit of turn. Since this dimension is periodic, a period of 1 tr is provided as the value of the period key. Here, the label in the corresponding ReciprocalDimension object is sideband order. The value of the dependent_variables key is an array with a single ExternalDependentVariable object describing the single-component complex signal response.

```json
pass.csdfe
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2010-11-22T22:26:00Z",
    "tags": ["13C NMR", "Phase Adjusted Spinning Sideband", "L-Histidine"],
    "description": "M.C. Davis, et al., J.Magn.Reson., 210, 51-58 (2011).",
    "dimensions": [
      {
        "type": "linear",
        "count": 128,
        "quantity_name": "time",
        "increment": "10 \(\mu s\)",
        "reciprocal": {
          "coordinates_offset": "1229 Hz",
          "origin_offset": "131.543 MHz"
        }
      },
      {
        "type": "linear",
        "count": 32,
        "quantity_name": "plane angle",
        "label": "rotor phase / 2 \(\pi\)",
        "increment": "0.03125 tr",
        "period": "1 tr",
        "reciprocal": {
          "label": "sideband order"
        }
      }
    ],
    "dependent_variables": [
      {
        "type": "external",
        "numeric_type": complex64",
        "quantity_type": "scalar",
        "components_url": "file:/pass.dat"
      }
    ]
  }
}
```
Listing 6: A 3D example dataset of the NMR phase incremented echo train acquisition (PIETA) measurement[3] with three LinearDimension objects. The 1st dimension, labeled as t2, is sampled every 15 µs for 256 points starting at 1.92 ms. The 2nd dimension, labeled as t1 (echo time), is sampled every 5 ms for 128 points. The 3rd dimension, labeled as pulse phase/2π, is periodic with period as 1 tr, and sampled every 0.0039065 turns for 256 points. Its ReciprocalDimension object is labeled as accumulated coherence order change. The dependent_variables holds an array with a single ExternalDependentVariable object with complex values stored in an external file as a binary array. The component_url gives the external file location on a remote https server.

```json
pieta.csdfe
{
  "csdm": {
    "version": "1.0",
    "timestamp": "2019-07-01T21:03:42Z",
    "dimensions": [
      {
        "type": "linear",
        "count": 256,
        "quantity_name": "time",
        "increment": "15 µs",
        "coordinates_offset": "-1.92 ms",
        "label": "t2",
        "reciprocal": {
          "origin_offset": "400.065795 MHz",
          "label": "1H frequency shift"
        }
      },
      {
        "type": "linear",
        "count": 128,
        "increment": "5.0 ms",
        "label": "t1 (echo time)"
      },
      {
        "type": "linear",
        "count": 256,
        "increment": "0.00390625 tr",
        "period": "1.0 tr",
        "label": "pulse phase / 2 $\pi$",
        "reciprocal": {
          "label": "accumulated coherence order change"
        }
      }
    ],
    "dependent_variables": [
      {
        "type": "external",
        "numeric_type": "complex64",
        "quantity_type": "scalar",
        "components_url": "https://osu.box.com/shared/static/o46giboeu8w0091a6q9jbezks7duymon.data"
      }
    ]
  }
}
```
S2 Overview of Physical Quantities, Units and Constants

S2.1 Dimensionality

In the International System of Units (SI) seven reference quantities are used to define seven dimensions whose symbols are given in Table S1. The dimensionality of any physical quantity, \( q \), can then be expressed in terms of these seven reference dimensions in the form of a dimensional product

\[
\text{dim } q = L^\alpha \cdot M^\beta \cdot T^\gamma \cdot I^\delta \cdot \Theta^\epsilon \cdot N^\zeta \cdot J^\eta,
\]

where the lower case greek symbols represent integers called the dimensional exponents. The dimensionality of any physical quantity can be represented as a point in the space of dimensional exponents \((\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta)\). Physical quantities with different meanings can have the same dimensionality. For example, the thermodynamic quantities entropy and heat capacity are different physical quantities having the same physical dimensions. Only physical quantities with the same dimensionality can be added. With the operation of multiplication, the physical dimensions form a mathematical group.

There also exists dimensionless quantities, such as plane angle which has dimensionality of \( L/L \), or solid angle with a dimensionality of \( L^2/L^2 \). Representing such dimensionalities requires us to split each dimension exponent into the numerator and denominator exponents. Thus, we redefine our set of dimensionalities to incorporate numerator and denominator exponents,

\[
\text{dim } q = \left[ L^{\alpha_+} \cdot M^{\beta_+} \cdot T^{\gamma_+} \cdot I^{\delta_+} \cdot \Theta^{\epsilon_+} \cdot N^{\zeta_+} \cdot J^{\eta_+} \right] \cdot \left[ L^{\alpha_-} \cdot M^{\beta_-} \cdot T^{\gamma_-} \cdot I^{\delta_-} \cdot \Theta^{\epsilon_-} \cdot N^{\zeta_-} \cdot J^{\eta_-} \right].
\]

(2)

A list of accepted physical quantity names in the CSD model and their corresponding dimensionality symbols are given in Table 8. When no accepted physical quantity name exists for a dimensionality the symbol obtained from Eq. (2) should be used as the quantity name in the CSD model.

S2.2 Unit

Inherent in the measurement of any physical quantity is a comparison to a previous measurement. What is most useful is the ratio of the new measurement to a previous measurement. For example, an ancient scientist might have used a cubit to measure the ratio of large tree trunk’s circumference to its diameter. Around 3000 B.C.E., a cubit was decreed to be the length of a forearm and hand. So, a scientist could make a string of 1 cubit in length using the distance from the back of her elbow to the tip of her middle finger, and then use the string to measure the ratio

\[
\frac{l_{\text{circumference}}}{l_{\text{diameter}}} \approx 3.14.
\]

(3)

While another scientist with longer arms might have cut a longer string to be a cubit, the procedure for finding the ratio of large tree trunk’s circumference to its diameter will be the same, and the result is independent of the absolute length of the string used, i.e., independent of the units of length used.

S2.2.1 SI Units

We represent a physical quantity, \( q \), using the notation

\[
q = \{q\} \cdot [q],
\]

(4)

where \( \{q\} \) is the numerical value and \([q]\) is the reference unit.

| Reference Quantity         | Dimension Symbol |
|----------------------------|------------------|
| length                     | L                |
| mass                       | M                |
| time                       | T                |
| electric current           | I                |
| thermodynamic temperature  | \( \Theta \)      |
| amount of substance        | N                |
| luminous intensity         | J                |

Table S1: Dimension symbols for the seven reference quantities.
Table S2: Coherent SI base units for the seven reference quantities.

| Base Dimension      | Coherent SI Base Unit          |
|---------------------|--------------------------------|
| Name                | Plural Name                    | Symbol |
| length              | meter                          | meters | m   |
| mass                | kilogram                       | kilograms | kg |
| time                | second                         | seconds | s   |
| electric current    | ampere                         | ampere | A   |
| thermodynamic temp. | kelvin                         | kelvin | K   |
| amount of substance | mole                           | moles   | mol |
| luminous intensity | candela                        | candelas | cd |

Table S3: Base root unit names and symbols for the seven reference quantities.

| Base Dimension      | Base Root Unit                  |
|---------------------|---------------------------------|
| Name                | Plural Name | Symbol |
| length              | meter       | meters | m   |
| mass                | gram        | grams  | g   |
| time                | second      | seconds | s   |
| electric current    | ampere      | ampere | A   |
| thermodynamic temp. | kelvin      | kelvin | K   |
| amount of substance | mole        | moles   | mol |
| luminous intensity | candela     | candelas | cd |

**Coherent SI Base Units:** The coherent SI base (reference) units form a set of seven units, described in Table S2 and given by the symbols

\[ [q]_{	ext{CBU}} \subseteq [Q]_{	ext{CBU}} = \{\text{m, kg, s, A, K, mol, cd}\}. \] (5)

**SI Base Root Units:** A minor complication is that the coherent base unit for mass in SI Units is defined as the kilogram and not the gram. For this reason, we define the set of seven base root units,

\[ [q]_{	ext{BRU}} \subseteq [Q]_{	ext{BRU}} = \{\text{m, g, s, A, K, mol, cd}\}. \] (6)

with names and symbols shown in Table S3.

**SI Base Units:** The set of Coherent SI Base Units only includes the seven SI base units, given in Table S2. The larger set of SI Base Units includes the Coherent SI Base Units as well as all decimal multiples of the root units created using the 20 prefix symbols given in Table S4 with the root unit names and symbols given in Table S3. These prefixed and unprefixed symbols form the set, \( [Q]_{\text{BU}} \), of 147 SI base units,

\[ [q]_{\text{BU}} \subseteq [Q]_{\text{BU}} = \{x_L^m, x_M^g, x_T^s, x_I^A, x_\Theta^K, x_N^\text{mol}, x_J^\text{cd}\}, \] (7)

where \( [Q]_{\text{CBU}} \subseteq [Q]_{\text{BU}} \), \( [Q]_{\text{BRU}} \subseteq [Q]_{\text{BU}} \), and \( x_i \) indicates that the root unit symbol may be modified with one of the SI prefixes given in Table S4.

**Coherent Derived SI Units:** Coherent derived SI units are an infinite set, \( [Q]_{\text{CDU}} \), defined as the products of powers of coherent SI base units,

\[ [q]_{\text{CDU}} \subseteq [Q]_{\text{CDU}} = \{m^\alpha \cdot g^\beta \cdot s^\gamma \cdot A^\delta \cdot K^\epsilon \cdot \text{mol}^\zeta \cdot \text{cd}^\eta\}, \] (8)

for all positive and negative integer values of the exponents. Here \( [Q]_{\text{CBU}} \subseteq [Q]_{\text{CDU}} \).

**Derived SI Root Units:** Derived SI root units are an infinite set, \( [Q]_{\text{DRU}} \), defined as the products of powers of SI base root units,

\[ [q]_{\text{DRU}} \subseteq [Q]_{\text{DRU}} = \{m^\alpha \cdot g^\beta \cdot s^\gamma \cdot A^\delta \cdot K^\epsilon \cdot \text{mol}^\zeta \cdot \text{cd}^\eta\}, \] (9)

for all positive and negative integer values of the exponents. Here \( [Q]_{\text{BRU}} \subseteq [Q]_{\text{DRU}} \).
corresponding coherent derived units are given in table set that, for convenience, have their own special names and symbols. The 22 special names and symbols and their incorporate the numerator and denominator exponents, into the numerator and denominator exponents. Thus, we redefine the infinite set of coherent derived SI units to used for the derived quantity of force, is given the special name “newton” and replaced with the symbol “N”.

| Special SI Units: |
|-------------------|
| Derived SI units are an infinite set, \([Q]_{\text{DU}}\), defined as the products of powers of SI base units, |
| \([q]_{\text{DU}} \in [Q]_{\text{DU}} = \{(xLm)\alpha \cdot (xMg)\beta \cdot (xTS)\gamma \cdot (xA)\delta \cdot (x\Theta K)\epsilon \cdot (xN\text{mol})\zeta \cdot (xJ\text{cd})\eta\}\), |
| for all positive and negative integer values of the exponents. Here \([Q]_{\text{CBU}} \subseteq [Q]_{\text{CDU}} \subseteq [Q]_{\text{DU}}\) and \([Q]_{\text{BRU}} \subseteq [Q]_{\text{DRU}} \subseteq [Q]_{\text{DU}}\). |

**Derived SI Dimensionless Units:** There also exists dimensionless units, such as the radian which has units of m/m, or the steradian with units of m^2/m^2. Representing such units requires us to split each dimension exponent into the numerator and denominator exponents. Thus, we redefine the infinite set of coherent derived SI units to incorporate the numerator and denominator exponents,

\[
[q]_{\text{CDU}} \in [Q]_{\text{CDU}} = \left\{ \left[ \frac{m^{n+}}{m^n} \right] \cdot \left[ \frac{\text{kg}^{\beta+}}{\text{kg}^{\beta-}} \right] \cdot \left[ \frac{s^{\gamma+}}{s^{\gamma-}} \right] \cdot \left[ \frac{A^{\delta+}}{A^{\delta-}} \right] \cdot \left[ \frac{\text{K}^{\eta+}}{\text{K}^{\eta-}} \right] \cdot \left[ \frac{\text{mol}^{\zeta+}}{\text{mol}^{\zeta-}} \right] \cdot \left[ \frac{\text{cd}^{\eta+}}{\text{cd}^{\eta-}} \right] \right\},
\]

for all positive integer values of the exponent numerator and denominators. Using this approach we represent dimensionless units with an exponent vector using the notation \((\alpha_+, -\alpha_-), (\beta_+, -\beta_-), (\gamma_+, -\gamma_-), (\delta_+, -\delta_-), (\epsilon_+, -\epsilon_-), (\zeta_+, -\zeta_-), (\eta_+, -\eta_-)\). Thus, the radian can be represented by the exponent vector \((1-1, 0-0, 0-0, 0-0, 0-0, 0-0, 0-0, 0-0)\) and the steradian by \((2-2, 0-0, 0-0, 0-0, 0-0, 0-0, 0-0, 0-0)\). Dimensionless quantities such as counts, are not derived SI units and are represented with the exponent vector \((0, 0, 0, 0, 0, 0, 0, 0)\).

Finally, we redefine the infinite set of derived SI units to incorporate numerator and denominator exponents,

\[
[q]_{\text{DU}} \in [Q]_{\text{DU}} = \left\{ \left[ \frac{x_L^m m^\alpha}{m^\alpha} \right] \cdot \left[ \frac{x_M^\beta}{m^\beta} \right] \cdot \left[ \frac{x_T^\gamma}{s^\gamma} \right] \cdot \left[ \frac{x_A^\delta}{A^\delta} \right] \cdot \left[ \frac{x_\Theta^\epsilon}{\text{K}^\epsilon} \right] \cdot \left[ \frac{x_N^\zeta}{\text{mol}^\zeta} \right] \cdot \left[ \frac{x_J^{\eta+}}{\text{cd}^{\eta-}} \right] \right\},
\]

for all positive integer values of the exponent numerator and denominators. The infinite set \([Q]_{\text{DU}}\), as described by Eq. (12), is a complete set of units for all quantities in the physical sciences and is allowed in the CSD model.

**Equivalent Units** While all units in the set \([Q]_{\text{DU}}\) have unique derived symbols and names, it should be noted that some are functionally equivalent. For example, a derived physical quantity such as speed in units of \([\text{m} \cdot \text{s}^{-1}]\) can be converted among any of the units below,

\[
[\text{m} \cdot \text{s}^{-1}] \equiv [\text{km} \cdot \text{ks}^{-1}] \equiv [\text{hm} \cdot \text{hs}^{-1}] \equiv [\text{dam} \cdot \text{das}^{-1}] \equiv [\text{dm} \cdot \text{ds}^{-1}] \equiv [\text{cm} \cdot \text{cs}^{-1}] \equiv [\text{mm} \cdot \text{ms}^{-1}],
\]

without modifying the numerical value of the quantity.

**Special SI Units:** In the SI system there is a set, \([Q]_{\text{SU}}\), of 22 coherent derived units contained within the \([Q]_{\text{CDU}}\) set that, for convenience, have their own special names and symbols. The 22 special names and symbols and their corresponding coherent derived units are given in table S5. For example, the coherent derived SI unit: \([\text{m} \cdot \text{kg} \cdot \text{s}^{-2}]\), used for the derived quantity of force, is given the special name “newton” and replaced with the symbol “N”.

| SI Prefix Name | x, SI Prefix Symbol | factor | SI Prefix Name | x, SI Prefix Symbol | factor |
|---------------|---------------------|--------|---------------|---------------------|--------|
| yotta         | Y                   | \(10^{24}\) | yocto         | y                    | \(10^{-24}\) |
| zetta         | Z                   | \(10^{21}\) | zepto          | z                    | \(10^{-21}\) |
| exa           | E                   | \(10^{18}\) | atto           | a                    | \(10^{-18}\) |
| peta          | P                   | \(10^{15}\) | femto          | f                    | \(10^{-15}\) |
| tera          | T                   | \(10^{12}\) | pico           | p                    | \(10^{-12}\) |
| giga          | G                   | \(10^9\) | nano           | n                    | \(10^{-9}\) |
| mega          | M                   | \(10^6\) | micro          | \(\mu\)              | \(10^{-6}\) |
| kilo          | k                   | \(10^3\) | milli          | m                    | \(10^{-3}\) |
| hecto         | h                   | \(10^2\) | centi          | c                    | \(10^{-2}\) |
| deca          | da                  | \(10^1\) | deci           | d                    | \(10^{-1}\) |

Table S4: SI prefixes used for the seven reference quantities.
Additional examples are given in Table S5: The 22 Coherent Derived SI Units with Special SI Names and Symbols.

### Special SI Units with prefixes:

The 22 coherent derived units with special SI units have the additional possibility of being modified by SI prefixes to create a set, $[Q]_{xSU}$, of 462 units. Some of the units in $[Q]_{xSU}$, however, may not have an equivalent in the set of derived SI units, $[Q]_{DU}$. That is, $[Q]_{DU} - [Q]_{xSU} \neq \emptyset$. For example, the Sievert is equivalent to the coherent derived unit m²·s⁻², and

$$[\text{Sv}] = [\text{m}^2 \cdot \text{s}^{-2}] \subseteq [Q]_{DU},$$

but the decisievert, [dSv], which is one-tenth of the coherent derived unit m²·s⁻², has no equivalent unit in the set $[Q]_{DU}$. That is,

$$[\text{dSv}] = 0.1[\text{m}^2 \cdot \text{s}^{-2}] \not\subseteq [Q]_{DU}. \quad (15)$$

On the other hand, the centiSievert, [cSv], which is one-hundredth of the coherent derived unit m²·s⁻², has an equivalent unit in the set $[Q]_{DU}$,

$$[\text{cSv}] = 0.01[\text{m}^2 \cdot \text{s}^{-2}] = [\text{dm}^2 \cdot \text{s}^{-2}] = [\text{m}^2 \cdot \text{das}^{-2}] \subseteq [Q]_{DU}. \quad (16)$$

### Derived Units employing Special SI Units:

With the introduction of special SI symbols there is the additional possibility of derived quantities with derived units employing a special SI unit. Generally, however, no more than one special SI symbol appears in such a derived unit symbol and the special SI symbol often appears as a linear term in the numerator. For example, the derived quantity dynamic viscosity can use the derived unit $[\text{Pa} \cdot \text{s}]$ as well as $[\text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-1}]$. The derived quantity of surface tension can use the derived unit $[\text{N} \cdot \text{m}^{-1}]$ as well as $[\text{kg} \cdot \text{s}^{-2}]$.

### S2.2.2 Non-SI Units:

There are several units outside the International System of Units that continue to be used in different science and engineering communities. While it is straightforward to include non-SI unit symbols in the same manner as special SI unit symbols, one has to be careful to avoid symbol and name collisions, particularly as some non-SI unit symbols employ SI prefixes.
### S2.2.3 Physical constants

It is also possible to use symbols for physical constants as unit symbols, although again one has to be careful to avoid symbol collisions. There are conflicts between unit symbols and commonly accepted symbols for physical constants, e.g., the planck constant and hour (h), or the newton gravitational constant and gauss (G). Table S7 gives a list of suggested symbols for physical constants (giving priority to unit symbols) which avoid collision with accepted unit symbols.

Due of the symbol collision between the euler constant $e = 2.718281828459045...$ all string representations of numbers using scientific notation in the serialization of the CSD model should use an upper case “E”, e.g., 6.022140857E+23 instead of 6.022140857e+23. Since the euler constant is dimensionless the latter expression could mistakenly be interpreted as the valid equation: $6.022140857 \times e + 23 = 39.3698760600388...$

### Table S6: Quantities with Derived Symbols using Special SI Symbols.

| Derived Quantity                        | Derived Symbol | Coherent Derived SI Symbol |
|----------------------------------------|----------------|---------------------------|
| dynamic viscosity                      | Pa $\cdot$ s   | $m^{-1} \cdot kg \cdot s^{-1}$ |
| moment of force                        | N $\cdot$ m    | $m^2 \cdot kg \cdot s^{-2}$ |
| surface tension                        | N/m            | $kg \cdot s^{-2}$          |
| angular velocity                       | rad/s          | $m/(m \cdot s)$            |
| angular acceleration                   | rad/s$^2$      | $m/(m \cdot s^2)$          |
| heat flux density, irradiance          | W/m$^2$        | $kg \cdot s^{-3}$          |
| heat capacity, entropy                 | J/K            | $m^2 \cdot kg \cdot s^{-2} \cdot K^{-1}$ |
| specific heat capacity, specific entropy | J/(kg $\cdot$ K) | $m^2 \cdot s^{-2} \cdot K^{-1}$ |
| specific energy                        | J/kg           | $m^2 \cdot s^{-2}$          |
| thermal conductivity                   | W/(m $\cdot$ K) | $m \cdot kg \cdot s^{-3} \cdot K^{-1}$ |
| energy density                         | J/m$^3$        | $m^{-1} \cdot kg \cdot s^{-2}$ |
| electric field strength                | V/m            | $m \cdot kg \cdot s^{-3} \cdot A^{-1}$ |
| electric charge density                | C/m$^3$        | $m^{-3} \cdot s \cdot A$   |
| surface charge density                 | C/m$^2$        | $m^{-2} \cdot s \cdot A$   |
| electric flux density, electric displacement | C/m$^2$   | $m^{-2} \cdot s \cdot A$   |
| permittivity                           | F/m            | $m^{-3} \cdot kg^{-1} \cdot s^4 \cdot A^2$ |
| permeability                           | H/m            | $m \cdot kg \cdot s^{-2} \cdot A^{-2}$ |
| molar energy                           | J/mol          | $m^2 \cdot kg \cdot s^{-2} \cdot mol^{-1}$ |
| molar entropy, molar heat capacity    | J/(mol $\cdot$ K) | $m^2 \cdot kg \cdot s^{-2} \cdot K^{-1} \cdot mol^{-1}$ |
| exposure (x- and γ-rays)               | C/kg           | $kg^{-1} \cdot s \cdot A$   |
| absorbed dose rate                     | Gy/s           | $m^2 \cdot s^{-3}$          |
| radiant intensity                      | W/sr           | $m^4 \cdot m^{-2} \cdot kg \cdot s^{-3}$ |
| radiance                               | W/(m$^2 \cdot$ sr) | $m^2 \cdot m^{-2} \cdot kg \cdot s^{-3}$ |
| catalytic activity concentration      | kat/m$^3$      | $m^{-3} \cdot s^{-1} \cdot mol$ |

### References

[1] Ambler Thompson and Barry N. Taylor. Guide for the use of the International System of Units (SI). March 30, 2008 2008.

[2] W. T. Dixon. Spinning-sideband-free NMR-spectra. *J. Magn. Reson.*, 44:220–223, 1981.

[3] J. H. Baltisberger, B. J. Walder, E. G. Keeler, D. C. Kaseman, K. J. Sanders, and P. J. Grandinetti. Phase incremented echo train acquisition in NMR spectroscopy. *J. Chem. Phys.*, 136:211104–1–4, 2012.

[4] Rupert D. Gillard and Robert J. Lancashire. Electron spin resonance of vanadium in amanita muscaria. *Phytochemistry*, 23:179 – 180, 1984.

[5] O. N. Antzutkin, S. C. Shekar, and M. H. Levitt. Two-dimensional sideband separation in magic-angle spinning NMR. *J. Magn. Reson. A*, 115:7–19, 1995.
| Constant                                      | Suggested symbol | Value in Coherent Derived SI Unit                                      |
|-----------------------------------------------|------------------|-----------------------------------------------------------------------|
| avogadro constant                            | N_A              | 6.022140857E+23 (1/mol)                                              |
| bohr magneton                                 | µ_B              | 9.274009992054043E-24 m⁻²*A                                         |
| boltzmann constant                           | k_B              | 1.38064852E-23 m²*kg/(s²*K)                                          |
| characteristic impedance of vacuum           | Z_0              | 376.7303134617707 m⁻²*kg/(s⁻³*A⁻²)                                    |
| compton wavelength                           | λ_C              | 2.42631023609262E-12 m                                               |
| electric constant                            | ε_0              | 8.854187817620413E-12 s⁻⁴*A²/(m⁻³*Kg)                                |
| electron g factor                            | g_e              | -2.00231930436182 m²*A⁻¹/(m⁻²*Kg)                                    |
| electron magnetic moment                     | µ_e              | -9.28476462E-24 m²*A                                                |
| electron mass                                | m_e              | 9.10938356E-31 kg                                                    |
| elementary charge                            | q_e              | 1.6021766208E-19 s*A                                                 |
| euler constant                               | e                | 2.718281828459045 ...                                                |
| fine structure constant                      | α                | 0.00729735266206478 m⁻⁵*Kg*s⁻⁴*A⁻²/(m⁻⁵*Kg*s⁻⁴*A⁻²)                   |
| gas constant                                 | R                | 8.314459861448581 m⁻²*Kg/(s⁻²*Kmol)                                  |
| newton gravitational constant                | G_N              | 6.67408E-11 m³/(kg*s²)                                               |
| gravity acceleration                         | g_0              | 9.80665 m/s²                                                        |
| magnetic constant                            | µ_0              | 1.256637061435917E-06 m*Kg/(m²*A)                                    |
| magnetic flux quantum                        | Φ_0              | 2.067833831170082E-15 m²*Kg/(m²*A)                                   |
| muon g factor                                | g_µ              | -2.0023318418 m²*A⁻¹/(m²*A)                                          |
| muon magnetic moment                         | µ_µ              | -4.49044826E-26 m²*A                                                 |
| myon mass                                    | m_µ              | 1.883531594E-28 kg                                                   |
| neutron g factor                             | g_n              | -3.82608545 m²*A⁻¹/(m²*A)                                            |
| neutron magnetic moment                      | µ_n              | -9.662365E-27 m²*A                                                   |
| neutron mass                                 | m_n              | 1.674927471E-27 kg                                                   |
| nuclear magneton                             | µ_N              | 5.050783698211084E-27 m²*A                                          |
| pi                                           | π                | 3.141592653589793 m/m                                                |
| planck constant                              | h_P              | 6.62607004E-34 m²*Kg/s                                               |
| proton g factor                              | g_p              | 5.885694702 m²*A⁻¹/(m²*A)                                            |
| proton magnetic moment                        | µ_p              | 1.4106067873E-26 m²*A                                                |
| proton mass                                  | m_p              | 1.672621898E-27 kg                                                   |
| reduced planck constant                      | h                | 1.054571800139113E-34 m³*Kg/(m²*K)                                   |
| rydberg constant                             | R_∞              | 10973731.5705508 (1/m)                                               |
| speed of light                               | c_0              | 299792458 m/s                                                       |
| stefan-boltzmann constant                    | σ                | 5.670367E-08 m²*Kg/(m²*K⁻³)                                          |
| wien wavelength displacement constant        | b_λ              | 0.0028977729 m*K                                                     |

Table S7: Physical constants and suggested symbols for avoiding symbol collision with unit symbols.
[6] M.C. Davis, K. M. Shookman, J. D. Sillaman, and P. J. Grandinetti. TOP-PASS: A processing algorithm to reduce 2D PASS acquisition time. *J. Magn. Reson*, 210:51–58, 2011.
| Quantity                        | Dimensionality                  |
|--------------------------------|---------------------------------|
| absorbed dose                 | $L^2 \cdot M / (M \cdot T^2)$   |
| absorbed dose rate            | $L^2 / T^3$                     |
| acceleration                  | $L / T^2$                       |
| action                         | $L^2 \cdot M / T$               |
| amount                         | N                               |
| amount concentration          | $N / L^3$                       |
| amount of electricity          | $T \cdot I$                     |
| amount ratio                  | $N / N$                         |
| angular acceleration          | $L / (L \cdot T^2)$             |
| angular frequency             | $L / (L \cdot T)$               |
| angular momentum              | $L^2 \cdot M / T$               |
| angular speed                 | $L / (L \cdot T)$               |
| angular velocity              | $L / (L \cdot T)$               |
| area                           | $L^2$                           |
| area ratio                    | $L^2 / L^2$                     |
| capacitance                   | $T^4 \cdot I^2 / (L^2 \cdot M)$ |
| catalytic activity            | $N / T$                         |
| catalytic activity concentration | $N / (L^3 \cdot T)$            |
| catalytic activity content    | $N / (M \cdot T)$               |
| charge to amount ratio        | $T \cdot I / N$                 |
| charge to mass ratio          | $T \cdot I / M$                 |
| circulation                   | $L^2 / T$                       |
| compressibility               | $L \cdot T^2 / M$               |
| current                       | I                               |
| current density               | $I / L^2$                       |
| current ratio                 | $I / I$                         |
| density                       | $M / L^3$                       |
| diffusion coefficient         | $L^2 / T$                       |
| diffusion flux                | $N / (L^2 \cdot T)$             |
| dimensionless                 | 1                               |
| distance per volume           | $L / L^3$                       |
| dose equivalent               | $L^2 \cdot M / (M \cdot T^2)$   |
| dynamic viscosity             | $M / (L \cdot T)$               |
| elastic modulus               | $M / (L \cdot T^2)$             |
| electric charge               | $T \cdot I$                     |
| electric charge density       | $T \cdot I / L^3$               |
| electric conductance          | $T^3 \cdot I^2 / (L^2 \cdot M)$ |
| electric conductivity         | $T^3 \cdot I^2 / (L^3 \cdot M)$ |
| electric dipole moment        | $L \cdot T \cdot I$             |
| electric displacement         | $T \cdot I / L^2$               |
| electric field gradient       | $L^2 \cdot M / (L^2 \cdot T^3 \cdot I)$ |
| electric field strength       | $L \cdot M / (T^3 \cdot I)$     |
| electric flux                 | $L^3 \cdot M / (T \cdot I)$     |
| electric flux density         | $T \cdot I / L^2$               |
| electric polarizability       | $L^2 \cdot T^4 \cdot I^2 / (L^2 \cdot M)$ |
| electric potential difference | $L^2 \cdot M / (T^3 \cdot I)$   |
| electric quadrupole moment    | $L^2 \cdot T \cdot I$           |
| electric resistance           | $L^2 \cdot M / (T^3 \cdot I^2)$ |
| electric resistance per length| $L^2 \cdot M / (L \cdot T^3 \cdot I^2)$ |
| electric resistivity          | $L^3 \cdot M / (T^3 \cdot I^2)$ |
| electrical mobility           | $L^2 \cdot T^3 \cdot I / (L^2 \cdot M \cdot T)$ |
| electromotive force           | $L^2 \cdot M / (T^3 \cdot I)$   |
| energy                        | $L^2 \cdot M / T^2$             |
| energy density                | $M / (L \cdot T^2)$             |
| entropy                       | $L^2 \cdot M / (T^2 \cdot 8)$   |
| fine structure constant       | $L^5 \cdot M \cdot T^4 \cdot I^2 / (L^5 \cdot M \cdot T^4 \cdot I^2)$ |
| first hyperpolarizability     | $L^3 \cdot T^7 \cdot I^3 / (L^4 \cdot M^2)$ |
| Quantity                                      | Dimensionality                                      |
|----------------------------------------------|-----------------------------------------------------|
| fluidity                                     | L•T/M                                              |
| force                                        | L•M/T^2                                            |
| frequency                                   | 1/T                                                |
| frequency per electric field gradient        | L^2•T^3•I/(L^2•M•T)                                |
| frequency per electric field gradient squared| L^4•T^6•I^2/(L^4•M^2•T)                            |
| frequency per magnetic flux density          | T•I/M                                              |
| frequency ratio                             | T/T                                                |
| gas permeance                               | L•T^2•N/(L^2•M•T)                                 |
| gravitational constant                      | L^3/(M•T^2)                                       |
| gyromagnetic ratio                          | L•T^2•I/(L•M•T)                                    |
| heat capacity                               | L^2•M/(T^2•θ)                                      |
| heat flux density                           | L^2•M/(L^2•T^3)                                    |
| heat transfer coefficient                   | M/(T^3•θ)                                         |
| illuminance                                 | L^2•J/L^4                                         |
| inductance                                  | L^2•M/(T^2•I^2)                                    |
| inverse amount                              | 1/N                                                |
| inverse amount concentration inverse time    | L^3/(T•N)                                         |
| inverse area                                | 1/L^2                                             |
| inverse current                             | 1/I                                                |
| inverse length                              | 1/L                                                |
| inverse luminous intensity                  | 1/J                                                |
| inverse magnetic flux density                | T^2•I/M                                            |
| inverse mass                                | 1/M                                                |
| inverse temperature                         | 1/θ                                                |
| inverse time                                | 1/T                                                |
| inverse time squared                        | 1/T^2                                             |
| inverse volume                              | 1/L^3                                             |
| irradiance                                  | L^2•M/(L^2•T^3)                                    |
| kinematic viscosity                         | L^2/T                                             |
| length                                       | L                                                 |
| length ratio                                | L/L                                                |
| linear momentum                             | L•M/T                                             |
| luminance                                   | J/L^2                                             |
| luminous efficacy                           | T^3•J/(L^2•M)                                      |
| luminous energy                             | L^2•T•J/L^2                                       |
| luminous flux                               | L^2•J/L^2                                         |
| luminous flux density                       | L^2•J/L^4                                         |
| luminous intensity                          | J                                                  |
| luminous intensity ratio                    | J/J                                                |
| magnetic dipole moment                      | L^2•I                                             |
| magnetic dipole moment ratio                | L^2•I/(L^2•I)                                      |
| magnetic field gradient                     | M/(L•T^2•I)                                       |
| magnetic field strength                     | I/L                                               |
| magnetic flux                               | L^2•M/(T^2•I)                                      |
| magnetic flux density                       | M/(T^2•I)                                         |
| magnetizability                             | L^2•M•T^4•I^2/(M^2•T^2)                           |
| mass                                         | M                                                 |
| mass concentration                          | M/L^3                                             |
| mass flow rate                              | M/T                                               |
| mass flux                                    | M/(L^2•T)                                         |
| mass ratio                                  | M/M                                               |
| mass to charge ratio                        | M/(T•I)                                           |
| molality                                     | N/M                                               |
| molar conductivity                          | L^2•T^3•I^2/(L^2•M•N)                              |
| molar energy                                | L^2•M/(T^2•N)                                      |
| molar entropy                               | L^2•M/(T^2•θ•N)                                   |
| molar heat capacity                         | L^2•M/(T^2•θ•N)                                   |
| molar magnetic susceptibility               | L^3/N                                             |
| molar mass                                   | M/N                                               |
| Quantity                                      | Dimensionality                                      |
|----------------------------------------------|-----------------------------------------------------|
| moment of force                              | $L^2 \cdot M^2/(M \cdot T^2)$                      |
| moment of inertia                            | $L^2 \cdot M$                                       |
| permeability                                 | $L \cdot M/(T^2 \cdot I^2)$                         |
| permittivity                                 | $T^4 \cdot I^2/(L^3 \cdot M)$                       |
| plane angle                                  | $L/L$                                               |
| porosity                                     | $L^3/L^3$                                           |
| power                                        | $L^2 \cdot M/T^3$                                   |
| power per luminous flux                      | $L^3 \cdot M/(L \cdot T^3 \cdot J)$                 |
| pressure                                     | $M/(L \cdot T^2)$                                   |
| pressure gradient                            | $M/(L^2 \cdot T^2)$                                 |
| radiance                                     | $L^4 \cdot M/(L^4 \cdot T^3)$                       |
| radiant flux                                 | $L^2 \cdot M/T^3$                                   |
| radiant intensity                            | $L^4 \cdot M/(L^2 \cdot T^3)$                       |
| radiation exposure                           | $T \cdot I/M$                                       |
| radioactivity                                | $L^3 \cdot M/(L \cdot T)$                          |
| reduced action                               | $L^2/\sqrt{L}$                                     |
| refractive index                             | $L \cdot T/(L \cdot T)$                            |
| rock permeability                            | $L^2$                                               |
| second hyperpolarizability                   | $L^4 \cdot T^{10} \cdot I^4/(L^6 \cdot M^3)$       |
| second radiation constant                    | $L^3 \cdot M \cdot T^2 \cdot \theta/(L^2 \cdot M \cdot T^2)$ |
| solid angle                                  | $L^2$                                               |
| specific energy                              | $L^2/T^2$                                           |
| specific entropy                             | $L^2/(T^2 \cdot \theta)$                           |
| specific gravity                             | $L^3 \cdot M/(L^3 \cdot M)$                         |
| specific heat capacity                        | $L^2/(T^2 \cdot \theta)$                           |
| specific power                               | $L^2 \cdot M/(M \cdot T^3)$                         |
| specific surface area                         | $L^2/M$                                             |
| specific volume                              | $L^3/M$                                             |
| spectral power                               | $L^2 \cdot M/(L \cdot T^3)$                         |
| spectral radiance                             | $L^4 \cdot M/(L^5 \cdot T^3)$                       |
| spectral radiant energy                       | $L^2 \cdot M/(L \cdot T^2)$                         |
| spectral radiant flux density                 | $M/(L \cdot T^3)$                                   |
| spectral radiant intensity                    | $L^4 \cdot M/(L^3 \cdot T^3)$                       |
| speed                                        | $L/T$                                               |
| stefan-boltzmann constant                     | $L^2 \cdot M/(L^2 \cdot T^3 \cdot \theta^4)$       |
| stress                                       | $M/(L \cdot T^2)$                                   |
| stress-optic coefficient                      | $L \cdot T^2/M$                                     |
| surface area to volume ratio                  | $L^2/L^3$                                           |
| surface charge density                        | $T \cdot I/L^2$                                     |
| surface density                              | $M/L^2$                                             |
| surface energy                               | $L^2 \cdot M/(L^2 \cdot T^2)$                       |
| surface tension                              | $M/T^2$                                             |
| temperature                                  | $\theta$                                            |
| temperature gradient                         | $\theta/L$                                          |
| temperature ratio                            | $\theta/\theta$                                     |
| thermal conductance                          | $L^2 \cdot M/(T^3 \cdot \theta)$                   |
| thermal conductivity                         | $L \cdot M/(T^3 \cdot \theta)$                     |
| time                                         | $T$                                                 |
| time ratio                                   | $T/T$                                               |
| torque                                       | $L^3 \cdot M/(L \cdot T^2)$                         |
| velocity                                     | $L/T$                                               |
| voltage                                      | $L^2 \cdot M/(T^3 \cdot I)$                         |
| volume                                       | $L^3$                                               |
| volume per length                            | $L^3/L$                                             |
| volume power density                          | $L^2 \cdot M/(L^3 \cdot T^3)$                       |
| volume ratio                                  | $L^3/L^3$                                           |
| volumetric flow rate                          | $L^3/T$                                             |
| wavelength displacement constant              | $L \cdot \theta$                                    |
| wavenumber                                    | $1/L$                                               |
Table 9: Physical unit symbols accepted in the CSD model.

| Unit or constant name                  | Symbol | SI Prefix | Value in Coherent Derived SI Unit |
|----------------------------------------|--------|----------|-----------------------------------|
| acres                                  | ac     | No       | 4046.8564224 m^2                  |
| acre feet                              | ac*ft  | No       | 1233.48183754752 m^3              |
| alpha particle mass                    | m_a    | No       | 6.64665723E-27 kg                 |
| alpha particle mass energy             | m_a*c_0^2 | No       | 5.97192096393537E-10 kg           |
| amperes                                | A      | Yes      | 1 A                               |
| amperes hours                          | A*h    | No       | 3600 s*A                          |
| amperes hours per gram                 | A*h/g  | No       | 3600000 s*A/kg                    |
| amperes minutes                        | A*min  | No       | 60 s*A                           |
| amperes minutes per gram               | A*min/g| No       | 60000 s*A/kg                     |
| amperes per ampere                     | A/A    | Yes      | 1 A/A                            |
| amperes per meter                      | A/m    | Yes      | 1 A/m                            |
| amperes per square meter               | A/m^2  | Yes      | 1 A/m^2                          |
| amperes seconds                        | A*s    | No       | 1 s*A                            |
| amperes seconds per gram               | A*s/g  | No       | 1000 s*A/kg                      |
| amperes square meters                  | A*m^2  | Yes      | 1 m^2*A                          |
| astronomical units                    | ua     | No       | 149597870691 m                    |
| atmospheres                            | atm    | No       | 101325 kg/(m*s^2)                |
| atomic mass constant                   | m_u    | No       | 1.66053904E-27 kg                 |
| atomic mass units                      | u      | No       | 1.66053904E-27 kg                 |
| atomic units of 1st polarizability    | q_e^3*a_0^3/E_h^2 | No       | 3.206361325744867E-53 m^3*s^7*A^3/(m^4*kg^2) |
| atomic units of 2nd polarizability     | q_e^4*a_0^4/E_h^3 | No       | 6.235380075451029E-65 m^4*s^10*A^4/(m^6*kg^3) |
| atomic unit of charge density          | q_e/a_0 | No       | 10812023777.567 s*A/m^3           |
| atomic unit of current                 | q_e/E_h/h | No       | 0.00662361818438244 A             |
| atomic unit of electric dipole moment  | q_e*a_0 | No       | 8.47835550255471E-30 m*s*A        |
| atomic unit of electric field          | E_h/(q_e*a_0) | No       | 51422067094.8027 m*kg/(s^3*A)     |
| atomic unit of electric field gradient | E_h/(q_e*a_0^0.2) | No       | 9.717362362541966E+21 m^2*kg/(m^2*s^3*A) |
| atomic unit of electric field gradient | A_0 | No       | 9.717362362541966E+21 m^2*kg/(m^2*s^3*A) |
| atomic unit of electric polarization   | q_e^2*a_0^2/E_h | No       | 1.648777272087738E-41 m^2*s^4*A^2/(m^2*kg) |
| atomic unit of electric potential      | E_h/q_e | No       | 27.2113606243671 m^2*kg/(s^3*A)   |
| atomic units of electric quadrupole moment | q_e*a_0^2 | No       | 4.486551481998239E-40 m^2*s*A    |
| atomic unit of energy                  | E_h    | No       | 4.359744650780484E-18 m^2*kg/s^2  |
| atomic units of force                  | E_h/a_0 | No       | 0.23872360557656E-80 m*kg/s^2    |
| atomic unit of length                  | a_0    | No       | 5.291772105638424E-11 m           |
| atomic units of magnetic dipole moment | h*q_e/m_e | No       | 1.854801998418009E-23 m^2*A      |
| atomic units of magnetic flux density  | h/(q_e*a_0^0.2) | No       | 235651.755093854 kg/(s^2*A)      |
| atomic units of magnetizability        | q_e*a_0^2/m_e | No       | 4.925197684724207E-10 m^2*kg*s^4*A^2/(kg^2*s^2) |
| atomic units of momentum               | h/a_0  | No       | 1.992851802293756E-24 m^2*kg    |
| atomic units of permittivity           | q_e^2/(a_0^2/E_h) | No       | 1.112650560536321E-10 s^*4*A^2/(m^3*kg) |
| atomic units of time                   | h/E_h  | No       | 2.41088432605869E-17 s           |
| atomic units of velocity               | a_0*0^E_h/h | No       | 2187691.262715648 m/s           |
| avogadro constant                      | N_A    | No       | 6.022140857E+23 (1/mol)          |
| bars                                   | bar    | Yes      | 100000 kg/(m*s^2)                |
| barns                                  | b      | No       | 1E-28 m^2                       |
| becquerels                             | Bq     | Yes      | 1 (1/s)                         |
| bohr magnetons                         | µ_B    | No       | 9.27409992054043E-24 m^2*A      |
| boltzmann constant                     | k_B    | No       | 1.38064852E-23 m^2*kg/(s^2*K)   |
| Unit or constant name | Symbol | SI Prefix | Value in Coherent Derived SI Unit |
|-----------------------|--------|-----------|----------------------------------|
| brewsters             | B      | No        | 1E-12 m^2*s^2/kg                |
| british thermal units | Btu    | No        | 1055.05585257348 m^2*kg/s^2      |
| british thermal unit per hour | Btu/h | No | 0.2930710701593 m^2*kg/s^3 |
| british thermal unit per minute | Btu/min | No | 17.584264209558 m^2*kg/s^3 |
| british thermal unit per second | Btu/s | No | 1055.05585257348 m^2*kg/s^3 |
| Btus per hour per foot per rankine | Btu(h*ft*°R) | No | 1.730734666295077 m^2*kg/(s^3*K) |
| Btus per hour per rankine | Btu(h*°R) | No | 0.5275279262867396 m^2*kg/(s^3*K) |
| Btus per hour per square foot per rankine | Btu(h*ft^2*°R) | No | 5.678263340863113 kg/(s^3*K) |
| calories              | cal    | No        | 4.1868 m^2*kg/s^2                |
| calories per gram per kelvin | cal/(g*K) | No | 4186.8 m^2/(s^2*K) |
| calories per hour     | cal/h  | No        | 0.001163 m^2*kg/s^3              |
| calories per hour per kelvin | cal(h*K) | No | 0.001163 m^2*kg/(s^3*K) |
| calories per hour per meter per kelvin | cal(h*m*K) | No | 0.001163 m^2*kg/(s^3*K) |
| calories per hour per square meter per kelvin | cal(h*m^2*K) | No | 0.001163 kg/(s^3*K) |
| calories per minute   | cal/min | No        | 0.06978 m^2*kg/s^3               |
| calories per mole     | cal/mol | No        | 4.1868 m^2*kg/(s^2*mol)           |
| candelas              | cd     | Yes       | 1 cd                             |
| candelas per candela  | cd/cd  | Yes       | 1 cd/cd                          |
| candela steradians    | cd*sr  | Yes       | 1 m^2*cd/m^2                      |
| celsius               | °C     | No        | 1 K                              |
| celsius per meter     | °C/m   | No        | 1 K/m                            |
| centimeter squared gram | cm^2*g | No      | 1E-07 m^2*kg                      |
| centimeter squared kilogram | cm^2*kg | No | 0.0001 m^2*kg                      |
| centuries             | hyr    | No        | 3155760000 s                     |
| chains                | ch     | No        | 20.1168 m                        |
| characteristic impedance of vacuum | Z_0 | No | 376.73031346177707 m^2*kg/(s^3*A^2) |
| charge density        | A*h/L  | No        | 36000000 s*A/m^3                  |
| compton wavelengths   | λ_C    | No        | 2.42631023609262E-12 m            |
| conductance quantum   | G_0    | No        | 7.480891730820603E-05 s^3*A^2/(m^2*kg) |
| coulombs              | C      | Yes       | 1 s*A                            |
| coulomb meters        | C*m    | Yes       | 1 m*s*A                          |
| coulombs per cubic meter | C/m^3 | Yes | 1 s*A/m^3                        |
| coulombs per kilogram  | C/kg   | Yes       | 1 s*A/kg                         |
| coulombs per mole     | C/mol  | No        | 1 s*A/mol                        |
| coulombs per square meter | C/m^2 | Yes | 1 s*A/m^2                        |
| coulombs per volt meter | C/(V*m) | Yes | 1 s^4*A^2/(m^3*kg)               |
| cubic centimeters per hour | cm^3/h | No | 2.777777777777778E-10 m^3/s    |
| cubic centimeters per minute | cm^3/min | No | 1.666666666666667E-08 m^3/s    |
| cubic centimeters per mole | cm^3/mol | No | 1E-06 m^3/mol                    |
| cubic centimeters per second | cm^3/s | No | 1E-06 m^3/s                     |
| cubic centimeters per second foot | cm^3/ft | No | 0.028316846592 m^3              |
| cubic centimeters per second hour | cm^3/h | No | 7.86579072E-06 m^3/s            |
| cubic centimeters per second minute | cm^3/min | No | 0.0004719474432 m^3/s            |
| cubic centimeters per second second | cm^3/s | No | 0.028316846592 m^3/s            |
| cubic inches           | in     | No        | 1.6387064E-05 m^3                |
| cubic meters per cubic meter | m^3/m^3 | No | 1 m^3/m^3                        |
| cubic meters per hour  | m^3/h  | No        | 0.00027777777777778 m^3/s       |
| Unit or constant name | Symbol | SI Prefix | Value in Coherent Derived SI Unit |
|-----------------------|--------|-----------|-----------------------------------|
| cubic meters per kilogram | m³/kg | No | 1 m³/kg |
| cubic meters per minute | m³/min | No | 0.01666666666666667 m³/s |
| cubic meters per mole | m³/mol | No | 1 m³/mol |
| cubic meters per second | m³/s | No | 1 m³/s |
| cubic yards | yd³ | No | 1E-30 m³ |
| cubic ångströms | Å³ | No | 1E-30 m³ |
| cups | cup | No | 0.0002365882365 m³ |
| curies | Ci | Yes | 3700000000 (1/s) |
| daltons | Da | Yes | 1.66053904E-27 kg |
| darcys | Dc | No | 9.869233E-2 |
| days | d | No | 315576000 s |
| degrees | ° | No | 0.0174532925199433 m/m |
| drams | dr | No | 0.00017718451953125 kg |
| dynes | dyn | Yes | 1E-05 m·kg/s² |
| dynes per centimeter | dyn/cm | No | 0.001 kg/s² |
| electric constant | ε₀ | No | 8.854187817620413E-12 s⁴*A²/(m³*kg) |
| electron g factor | g_e | No | -2.00231930436182 m²*A/(m²*A) |
| electron magnetic moment | µ_e | No | -9.28476462E-24 m²*A |
| electron mass | m_e | No | 9.10938356E-31 kg |
| electronvolts | eV | Yes | 1.602176608E-19 m²·kg/s² |
| ergs | erg | Yes | 1E-07 m²·kg/s² |
| ergs per second | erg/s | No | 1E-07 m²·kg/s³ |
| euler constant | e | No | 2.718281828459045 |
| fahrenheit | °F | No | 0.5555555555555556 K |
| fahrenheit per foot | °F/ft | No | 1.82268830562848 K/m |
| farads | F | Yes | 1 s⁴*A²/(m²*kg) |
| farads per meter | F/m | Yes | 1 s⁴*A²/(m³*kg) |
| faraday constant | &F | No | 96485.332849877 s*A/mol |
| fathoms | ftm | No | 1.8288 m |
| fine structure constant | α | No | 0.007297352566206478 m⁵·kg·s⁻⁴·A²/(m⁵·kg·s⁻⁴·A²) |
| fluid ounces | floz | No | 2.95735295625E-05 m³ |
| feet | ft | No | 0.3048 m |
| feet per hour | ft/h | No | 8.466666666666667E-05 m/s |
| feet per minute | ft/min | No | 0.00508 m/s |
| feet per second | ft/s | No | 0.3048 m/s |
| feet per square minute | ft/min² | No | 8.466666666666667E-05 m²/s² |
| feet per square second | ft/s² | No | 0.3048 m²/s² |
| feet pound force | ft*lbf | No | 1.3558179483314 m²·kg·s⁻²/(kg·s²) |
| feet pound force per hour | ft*lbf/h | No | 0.000376616096758223 m²·kg·s⁻³ |
| feet pound force per minute | ft*lbf/min | No | 0.0225969658055234 m²·kg·s⁻³ |
| feet pound force per second | ft*lbf/s | No | 1.3558179483314 m²·kg·s⁻³ |
| furlongs | fur | No | 201.168 m |
| gallons | gal | No | 0.003785411784 m³ |
| gallons per hour | gal/h | No | 1.051503273333333E-06 m³/s |
| Unit or constant name                     | Symbol          | SI Prefix | Value in Coherent Derived SI Unit          |
|------------------------------------------|-----------------|-----------|-------------------------------------------|
| gallons per minute                      | gal/min         | No        | 6.30901964E-05 m^3/s                      |
| gallons per second                      | gal/s           | No        | 0.003785411784 m^3/s                      |
| gas constant                            | R               | No        | 8.314459861448581 m^2*kg/(s^2*K*mol)       |
| gas permeance unit                      | GPU             | No        | 0.33 m*s^2*mol/(m^2*kg*s)                 |
| gauss                                    | G               | Yes       | 0.0001 kg/(s^2*A)                         |
| gauss per centimeter                    | G/cm            | No        | 0.01 kg/(m*s^2*A)                         |
| gill                                     | gi              | No        | 0.00011829411825 m^3                      |
| grains                                   | gr              | No        | 6.479891E-05 kg                           |
| grams                                    | g               | Yes       | 0.001 kg                                  |
| gram meters per second                  | m*g/s           | Yes       | 0.001 m*kg/s                              |
| grams per cubic meter                   | g/m^3           | Yes       | 0.001 kg/m^3                              |
| grams per kilogram                      | g/kg            | Yes       | 0.001 kg/kg                               |
| grams per liter                         | g/L             | Yes       | 1 kg/m^3                                  |
| grams per microliter                    | g/µL            | Yes       | 1000000 kg/m^3                            |
| grams per milliliter                    | g/mL            | Yes       | 1000 kg/m^3                               |
| grams per mole                          | g/mol           | Yes       | 0.001 kg/mol                              |
| grams per second                        | g/s             | Yes       | 0.001 kg/s                                |
| grams per square meter                  | g/m^2           | Yes       | 0.001 kg/m^2                              |
| grams per square meter per second       | g/(m^2*s)       | Yes       | 0.001 kg/(m^2*s)                          |
| gram square meters per second           | g*m^2/s         | Yes       | 0.001 m^2*kg/s                            |
| gravitational constant                  | G_N             | No        | 6.67408E-11 m^3/(kg*s^2)                  |
| gravity acceleration                    | g_0             | No        | 9.80665 m/s^2                             |
| grays                                    | Gy              | Yes       | 1 m^2*kg/(kg*s^2)                         |
| grays per second                        | Gy/s            | Yes       | 1 m^2/s^3                                 |
| half teaspoons                          | halfsp          | No        | 2.46460796875E-06 m^3                     |
| hectares                                | ha              | No        | 10000 m^2                                 |
| henries                                 | H               | Yes       | 1 m^2*kg/(s^2*A^2)                        |
| henries per meter                       | H/m             | Yes       | 1 m*kg/(s^2*A^2)                          |
| hertz                                    | Hz              | Yes       | 1 (s/s)                                   |
| hertz per atomic unit of electric field gradient | Hz/Λ_0       | No        | 1.029085396028452E-22 m^2*s^3*A/(m^2*kg*s) |
| hertz per atomic unit of electric field gradient squared | Hz/Λ_0^2     | No        | 1.059017664088893E-44 m^4*s^6*A^2/(m^4*kg^2*s) |
| hertz per hertz                          | Hz/Hz           | Yes       | 1 s/s                                     |
| hertz per tesla                          | Hz/T            | Yes       | 1 s*A/kg                                  |
| horse power per ounce                    | hp/oz           | No        | 26303.78891073498 m^2*kg/(kg*s^3)          |
| horse power per pound                    | hp/lb           | No        | 1643.986806920936 m^2*kg/(kg*s^3)          |
| horsepower                               | hp              | No        | 745.699872 m^2*kg/s^3                      |
| hours                                    | h               | No        | 3600 s                                    |
| hundredweight                            | cwt             | No        | 45.359237 kg                              |
| hundredweightUK                          | cwtUK           | No        | 50.80234544 kg                            |
| imperial cups                            | cupUK           | No        | 0.000284130625 m^3                        |
| imperial fluid ounces                    | flocUK          | No        | 2.84130625E-05 m^3                        |
| imperial gallons                         | galUK           | No        | 0.00454609 m^3                            |
| imperial gallons per hour                | galUK/h         | No        | 1.262802777777778E-06 m^3/s               |
| imperial gallons per minute              | galUK/min       | No        | 7.57681666666667E-05 m^3/s                |
| imperial gallons per second              | galUK/s         | No        | 0.00454609 m^3                            |
| imperial gill                            | giUK            | No        | 0.0001420653125 m^3                       |
| imperial half teaspoons                  | halfspUK        | No        | 2.95964010416667E-06 m^3                   |
| imperial pints                           | ptUK            | No        | 0.00056826125 m^3                         |
| Unit or constant name                      | Symbol   | SI Prefix | Value in Coherent Derived SI Unit |
|------------------------------------------|----------|-----------|----------------------------------|
| imperial quarts                          | qtUK     | No        | 0.0011365225 m³                |
| imperial quarter teaspoons               | quartertspUK | No     | 1.479847005208333E-06 m³      |
| imperial tablespoons                     | tbspUK   | No        | 1.77581640625E-05 m³           |
| imperial teaspoons                       | tspUK    | No        | 5.9138802083334E-06 m³         |
| inches                                   | in       | No        | 0.0254 m                       |
| inch ounce force                         | in*ozf   | No        | 0.007061551814226041 m²/kg²/s² |
| inches per hour                          | in/h     | No        | 7.055555555555556E-06 m/s     |
| inches per minute                        | in/min   | No        | 0.00042333333333334 m/s       |
| inches per second                        | in/s     | No        | 0.0254 m/s                     |
| inches per square minute                 | in/min²  | No        | 7.055555555555557E-06 m²/s²   |
| inches per square second                 | in/s²    | No        | 0.0254 m²/s²                   |
| inch pound force                         | in*lbf   | No        | 0.1129848290276167 m²/kg²/s²  |
| inches pound force per hour              | in*lbf/h | No        | 3.138467472989352E-05 m²/kg²/s²|
| inches pound force per minute            | in*lbf/min| No      | 0.001883080483793611 m²/kg²/s²|
| inches pound force per second            | in*lbf/s | No        | 0.1129848290276167 m²/kg²/s²  |
| joules                                   | J        | Yes       | 1 m²/kg/s²                      |
| joules per cubic meter                   | J/m³     | Yes       | 1 kg/(m³/s²)                    |
| joules per gram                          | J/g      | Yes       | 1000 m²/s²                      |
| joules per gram kelvin                   | J/(g*K)  | Yes       | 1000 m²/(s²*K)                  |
| joules per kilogram                      | J/kg     | Yes       | 1 m²/kg/(s²*K)                  |
| joules per kilogram kelvin               | J/(kg*K) | Yes       | 1 m²/(s²*K)                     |
| joules per liter                         | J/L      | Yes       | 1000 kg/(m³/s²)                 |
| joules per mole                          | J/mol    | Yes       | 1 m²/kg/(s²*mol)                |
| joules per mole kelvin                   | J/(mol*K)| Yes       | 1 m²/kg/(s²*K*mol)              |
| joules per nanometer                     | J/nm     | Yes       | 1E+09 m²/kg/(m³/s²)             |
| joules per radian                        | J/rad    | Yes       | 1 m³/kg/(m³/s²)                 |
| joules per second                        | J/s      | Yes       | 1 m²/kg/s²                      |
| joules per square meter                  | J/m²     | Yes       | 1 m²/kg/(m²/s²)                 |
| joules per tesla                         | J/T      | Yes       | 1 m²*A                          |
| joules second                            | J/s      | Yes       | 1 m²/kg/s                       |
| katals                                   | kat      | Yes       | 1 mol/s                         |
| katals per cubic meter                   | kat/m³   | Yes       | 1 mol/(m³/s)                    |
| katals per kilogram                      | kat/kg   | Yes       | 1 mol/(kg*s)                    |
| katals per liter                         | kat/L    | Yes       | 1000 mol/(m³/s)                 |
| kelvin                                   | K        | Yes       | 1 K                             |
| kelvin per kelvin                        | K/K      | Yes       | 1 K/K                           |
| kelvin per meter                         | K/m      | No        | 1 K/m                           |
| kilocalories                             | kcal     | No        | 4186.8 m²/kg/s²                 |
| kilocalories per hour per kelvin         | kcal/(h*K)| No    | 1.163 m²/kg/(s³*K)              |
| kilocalories per hour per meter per kelvin| kcal/(h*m*K)| No | 1.163 m²/kg/(s³*K)             |
| kilocalories per hour per square meter per kelvin | kcal/(h*m²*K)| No | 1.163 kg/(s³*K) |
| kilocalories per mole                    | kcal/mol | No        | 4186.8 m²/kg/(s²*mol)           |
| kilograms force                          | kgf      | No        | 9.80665 m³/kg/s²                |
| kilogram force centimeters per radian    | kgf*cm/rad| No     | 0.0980665 m³/kg/(m³/s²)         |
| kilogram force meters per radian         | kgf*m/rad| No        | 9.80665 m³/kg/(m³/s²)           |
| kilohertz per atomic unit of electric field gradient | khz/Λ 0 | No | 1.029885839028452E-19 m²/s³/A/(m²/kg*s) |
| kilohertz per atomic unit of electric field gradient squared | khz/Λ_0^2 | No | 1.059017664088893E-41 m⁴/s⁶/A²/(m⁴/kg²*s²) |
| Unit or constant name                              | Symbol  | SI Prefix | Value in Coherent Derived SI Unit |
|--------------------------------------------------|---------|-----------|-----------------------------------|
| knots                                            | kn      | No        | 0.514444444444444 m/s            |
| knots per second                                 | kn/s    | No        | 0.514444444444444 m/s^2          |
| leagues                                          | lea     | No        | 4828.032 m                       |
| light years                                      | ly      | No        | 9460730472580800 m               |
| links                                            | li      | No        | 0.201168 m                       |
| liters                                           | L       | Yes       | 0.001 m^3                        |
| liters per 100 kilometers                        | L/(100 km) | No    | 1E-08 m^3/m                      |
| liter per mole per second                        | L/(mol*s) | No    | 0.001 m^3/(s*mol)                |
| lumens                                           | ln      | Yes       | 1 m^2*cd/m^2                     |
| lumens per square foot                           | ln/ft^2 | No        | 10.76391041670972 m^2*cd/m^4     |
| lumens per square meter                          | ln/m^2  | Yes       | 1 m^2*cd/m^4                     |
| lumens per square meter per steradian            | ln/(m^2*sr) | Yes | 1 cd/m^2                        |
| lumens per watt                                  | ln/W    | Yes       | 1 s^3*cd/(m^2*kg)                |
| lumen seconds                                    | ln*s    | Yes       | 1 m^2*s*cd/m^2                   |
| lux                                              | lx      | Yes       | 1 m^2*cd/m^4                     |
| magnetic constant                                | µ_0     | No        | 1.256637061435917E-06 m*kg/(s^2*A^2) |
| magnetic flux quantum                            | Φ_0     | No        | 2.067833831170082E-15 m^2*kg/(s^2*A) |
| maxwells                                         | Mx      | Yes       | 1E-08 m^2*kg/(s^2*A)             |
| megahertz per atomic unit of electric field gradient | MHz/Λ_0 | No    | 1.029865839028452E-16 m^2*s^3*A/(m^2*kg*s) |
| megahertz per atomic unit of electric field gradient squared | MHz/Λ_0^2 | No | 1.059017664088893E-38 m^4*s^6*A^2/(m^4*kg^2*s) |
| meters                                           | m       | Yes       | 1 m                              |
| meters per hour                                  | m/h     | Yes       | 0.000277777777777778 m/s         |
| meters per hour per second                       | m/(h*s) | Yes       | 0.000277777777777778 m^2/s       |
| meters per liter                                 | m/L     | Yes       | 1000 m/m^3                       |
| meters per meter                                 | m/m     | Yes       | 1 m/m                            |
| meters per minute                                | m/min   | Yes       | 0.0166666666666667 m/s           |
| meters per second                                | m/s     | Yes       | 1 m/s                            |
| meters per square second                         | m/s^2   | Yes       | 1 m/s^2                          |
| meter seconds per meter second                   | m*s/(m*s) | No   | 1 m*s/(m*s)                      |
| meters squared gram                              | m^2*g   | No        | 0.001 m^2*kg                     |
| meters squared kilogram                          | m^2*kg  | No        | 1 m^2*kg                         |
| microdarcys                                      | µDc     | No        | 9.869233E-19 m^2                |
| micrograms                                       | mcg     | No        | 1E-09 kg                         |
| miles                                            | mi      | No        | 1609.344 m                       |
| miles per gallon                                 | mi/gal  | No        | 425143.707430272 m/m^3           |
| miles per hour                                   | mi/h    | No        | 0.44704 m/s                      |
| miles per hour per second                        | mi/(h*s) | No   | 0.44704 m/s^2                    |
| miles per imperial gallon                        | mi/galUK | No  | 354006.1899346471 m/m^3          |
| miles per minute                                 | mi/min  | No        | 26.8224 m/s                      |
| miles per second                                 | mi/s    | No        | 1609.344 m/s                     |
| miles per square minute                          | mi/min^2 | No  | 0.44704 m/s^2                    |
| miles per square second                          | mi/s^2  | No        | 1609.344 m/s^2                   |
| millennia                                         | kyr     | No        | 31557600000 s                    |
| milliampere hours                                | mA*h    | No        | 3.6 s*A                          |
| milliampere hours per gram                       | mA*h/kg | No        | 3600 s*A/kg                      |
| milliampere minutes                              | mA*min  | No        | 0.06 s*A                         |
| milliampere minutes per gram                     | mA*min/g | No  | 60 s*A/kg                         |
| milliampere seconds                              | mA*s    | No        | 0.001 s*A                        |
| Unit or constant name                       | Symbol      | SI Prefix | Value in Coherent Derived SI Unit |
|--------------------------------------------|-------------|-----------|-----------------------------------|
| milliampere seconds per gram               | mA*s/g      | No        | 1 s*A/kg                          |
| millidarcys                                | mDc         | No        | 9.869233E-16 m^2                 |
| millimeters of Hg                          | mmHg        | No        | 133.322 kg/(m*s^2)               |
| million oil barrels                        | MMMbbl      | No        | 158987.295 m^3                   |
| minutes                                    | min         | No        | 60 s                              |
| moles                                      | mol         | Yes       | 1 mol                             |
| moles per cubic meter                      | mol/m^3     | Yes       | 1 mol/m^3                         |
| moles per kilogram                         | mol/kg      | Yes       | 1 mol/kg                          |
| moles per liter                            | mol/L       | Yes       | 1000 mol/m^3                      |
| moles per liter                            | M           | Yes       | 1000 mol/m^3                      |
| moles per microliter                       | mol/µL      | Yes       | 1E+09 mol/m^3                     |
| moles per milliliter                       | mol/ml      | Yes       | 10000000 mol/m^3                  |
| moles per minute                           | mol/min     | Yes       | 0.01666666666666667 mol/s         |
| moles per mole                             | mol/mol     | Yes       | 1 mol/mol                         |
| moles per second                           | mol/s       | Yes       | 1 mol/s                           |
| months                                     | month       | No        | 2629800 s                         |
| muon g factor                              | g_µ         | No        | -2.0023318418 m^2*A/(m^2*A)       |
| muon magnetic moment                       | µ_µ         | No        | -4.9044826E-26 m^2*A              |
| myon mass                                  | m_µ         | No        | 1.883513594E-28 kg                |
| nanodarcys                                 | nDc         | No        | 9.869233E-21 m^2                 |
| natural units of energy                    | m_e*c_0^2   | No        | 8.187105649650028E-14 m^2*kg/s^2 |
| natural units of length                    | X_C         | No        | 3.8615926764E-13 m                |
| natural units of momentum                  | m_e*c_0     | No        | 2.73092448831719E-22 m*kg/s      |
| natural units of time                      | h/(m_e*c_0^2)| No   | 1.28808866822313E-21 s            |
| newtons per meter                          | N/m         | Yes       | 1 kg/s^2                         |
| newtons per square meter                   | N/m^2       | Yes       | 1 kg/(m*s^2)                      |
| neutron g factor                           | g_n         | No        | -3.82608545 m^2*A/(m^2*A)         |
| neutron magnetic moment                    | µ_n         | No        | -9.662365E-27 m^2*A              |
| neutron mass                               | m_n         | No        | 1.674927417E-27 kg                |
| newton                                       | N           | Yes       | 1 m*kg/s^2                       |
| newton meters                              | N*m         | Yes       | 1 m^2*kg^2/(kg*s^2)              |
| newton meters per radian                   | N*m/rad     | Yes       | 1 m^3*kg/(m*s^2)                  |
| newtons per coulomb                        | N/C         | Yes       | 1 m*kg/(s^3*A)                    |
| newtons per square ampere                  | N/A^2       | Yes       | 1 m*kg/(s^2*A^2)                  |
| newton seconds                             | N*s         | Yes       | 1 m*kg/s                          |
| newton seconds per square meter            | N*s/m^2     | Yes       | 1 kg/(m*s)                        |
| nuclear magnetons                          | µ_N         | No        | 5.050783698211084E-27 m^2*A      |
| ohms                                       | Q           | Yes       | 1 m^2*kg/(s^3*A^2)                |
| ohms centimeter                            | Q/cm        | Yes       | 0.01 m^3*kg/(s^3*A^2)             |
| ohms meter                                 | Q/m         | Yes       | 1 m^3*kg/(s^3*A^2)                |
| ohms per feet                              | Q/ft        | Yes       | 3.280839895013123 m^2*kg/(m*s^3*A^2) |
| ohms per meter                             | Q/m         | Yes       | 1 m^2*kg/(m*s^3*A^2)              |
| oil barrels                                | bbl         | No        | 0.158987295 m^3                   |
| ounces                                     | oz          | No        | 0.028349523125 kg                 |
| ounces force                               | ozf         | No        | 0.2780138509537812 m*kg/s^2       |
| ounce force inches                         | ozf*in      | No        | 0.007061551814226041 m^2*kg^2/(kg*s^2) |
| parts per billion                          | ppb         | No        | 1E-09                             |
| parts per million                          | ppm         | No        | 1E-06                             |
| Unit or constant name                  | Symbol | SI Prefix | Value in Coherent Derived SI Unit |
|---------------------------------------|--------|----------|----------------------------------|
| parts per quadrillion                 | ppq    | No       | 1E-15                            |
| parts per ten thousand                | ‰      | No       | 0.0001                           |
| parts per thousand                    | ‰      | No       | 0.001                            |
| parts per trillion                    | ppt    | No       | 1E-12                            |
| pascals                               | Pa     | Yes      | 1 kg/(m*s^2)                     |
| pascals per meter                     | Pa/m   | Yes      | 1 kg/(m^2*s^2)                   |
| pascal seconds                        | Pa*s   | Yes      | 1 kg/(m*s)                       |
| percent                               | %      | No       | 0.01                             |
| photos                                | ph     | Yes      | 10000 m^2*cd/m^4                 |
| pi                                    | π      | No       | 3.141592653589793 m/m            |
| pints                                 | pt     | No       | 0.000473176473 m^3               |
| planck charge                         | q_P    | No       | 1.87554602272158E-18 s*A         |
| planck constant                       | h_P    | No       | 6.62607004E-34 m^2*kg/s          |
| planck length                         | l_P    | No       | 1.616228373080886E-35 m          |
| planck mass                           | m_P    | No       | 2.176470195634196E-08 kg         |
| planck temperature                    | T_P    | No       | 1.41680793748162E+32 K           |
| planck time                           | t_P    | No       | 5.391157549003072E-44 s          |
| poises                                | P      | Yes      | 0.1 kg/(m*s)                     |
| pounds                                | lb     | No       | 0.45359237 kg                    |
| pounds force                          | lbf    | No       | 4.482216152605 m*kg/s^2          |
| pound force feet                      | lbf*ft | No       | 1.3558179483314 m^2*kg^2/(kg*s^2) |
| pound force feet per radian           | lbf*ft/rad | No     | 1.3558179483314 m^2*kg^2/(m*s^2)  |
| pound force inches                    | lbf*in | No       | 0.1129848290276167 m^2*kg^2/(kg*s^2) |
| pound force inches per radian         | lbf*in/rad | No    | 0.1129848290276167 m^3*kg/(m*s^2) |
| pounds force per square feet          | lbf/ft^2 | No     | 47.880259 kg/(m*s^2)             |
| pounds force per square inch          | psi    | No       | 6894.75729 kg/(m*s^2)            |
| pounds force per square inch          | psi/in | No       | 6894.75729 kg/(m^2*s)            |
| pounds force per square inch per foot | psi/ft | No       | 22620.59478346456 kg/(m^2*s^2)   |
| proton g factor                       | g_p    | No       | 5.585694702 m^2*A/(m^2*A)        |
| proton magnetic moment                | µ_p    | No       | 1.406067873E-26 m^2*A            |
| proton mass                           | m_p    | No       | 1.672621898E-27 kg               |
| quantum of circulation                | h_P/(2*m_e) | No     | 1.054571800139113E-34 m^2/s      |
| quarts                                | qt     | No       | 3.0864198E-3 m^3                 |
| quarter teaspoons                     | quartersp | No     | 1.232330984375E-06 m^3           |
| radians                               | rad    | Yes      | 1 m/m                           |
| radians per second                    | rad/s  | Yes      | 1 m/(m*s)                       |
| radians per second per tesla          | rad/(s*T) | Yes   | 1 m*s^2*A/(m^2*A)               |
| radians per square second             | rad/s^2 | Yes    | 1 m/(m^2*s^2)                   |
| rankines                              | °R     | No       | 0.5555555555555555 K             |
| rankines per foot                     | °R/ft  | No       | 1.822688830562848 K/m            |
| reduced planck constant               | h      | No       | 1.054571800139113E-34 m^3*kg/(m*s) |
| rydbergs                              | Ry     | No       | 2.179872325390242E-18 m^2*kg/s^2 |
| rydberg constant                      | R_∞    | No       | 10973731.5705508 (1/m)           |
| seconds                               | s      | Yes      | 1 s                              |
| seconds per second                    | s/s    | Yes      | 1 s/s                           |
| second radiation constant             | h_P*c_0/k_B | No     | 0.0143877735382772 m^3*kg*s^2*K/(m^2*kg*s^2) |
| siemens                               | S      | Yes      | 1 s^3*A^2/(m^2*kg)               |
| Unit or constant name                                    | Symbol       | SI Prefix | Value in Coherent Derived SI Unit                                                                 |
|---------------------------------------------------------|--------------|-----------|--------------------------------------------------------------------------------------------------|
| siemens centimeter squared per mole                    | S*cm^2/mol   | Yes       | 0.0001 m^2*s^3*A^2/(m^2*kg*mol)                                                                  |
| siemens meter squared per mole                          | S*m^2/mol    | Yes       | 1 m^2*s^3*A^2/(m^2*kg*mol)                                                                      |
| siemens per centimeter                                 | S/cm         | Yes       | 100 s^3*A^2/(m^3*kg)                                                                            |
| siemens per meter                                      | S/m          | Yes       | 1 s^3*A^2/(m^3*kg)                                                                               |
| sieverts                                                | Sv           | Yes       | 1 m^2*kg/(kg*s^2)                                                                               |
| speed of light                                         | c_0          | No        | 299792458 m/s                                                                                  |
| ångström ångströms                                     | Å^2          | No        | 1E-20 m^2                                                                                       |
| square centimeters                                     | cm^2         | No        | 0.0001 m^2                                                                                      |
| square centimeters per second                          | cm^2/s       | No        | 0.0001 m^2/s                                                                                    |
| square chains                                           | ch^2         | No        | 404.6872515856 m^2                                                                             |
| square feet                                              | ft^2         | No        | 0.09290304 m^2                                                                                  |
| square inches                                           | in^2         | No        | 0.00064516 m^2                                                                                  |
| square kilometers                                       | km^2         | No        | 10000000 m^2                                                                                    |
| square meters                                           | m^2          | No        | 1 m^2                                                                                           |
| square meters per cubic meter                           | m^2/m^3      | No        | 1 m^2/m^3                                                                                       |
| square meters per liter                                 | m^2/L        | No        | 1000 m^2/m^3                                                                                    |
| square meters per second                                | m^2/s        | No        | 1 m^2/s                                                                                         |
| square meters per square meter                          | m^2/m^2      | No        | 1 m^2/m^2                                                                                       |
| square micrometers                                      | µm^2         | No        | 1E-12 m^2                                                                                       |
| square micrometers per second                           | µm^2/s       | No        | 1E-12 m^2/s                                                                                    |
| square miles                                             | mi^2         | No        | 2589988.110336 m^2                                                                              |
| square millimeters                                      | mm^2         | No        | 1E-06 m^2                                                                                       |
| square millimeters per second                           | mm^2/s       | No        | 1E-06 m^2/s                                                                                    |
| square nanometers                                       | nm^2         | No        | 1E-18 m^2                                                                                       |
| square rods                                              | rod^2        | No        | 25.2929532241 m^2                                                                               |
| square yards                                            | yd^2         | No        | 0.83612736 m^2                                                                                  |
| stefan-boltzmann constant                               | σ            | No        | 5.670367E-08 m^2*kg/(m^2*s^3*K^4)                                                                |
| steradians                                              | sr           | Yes       | 1 m^2/m^2                                                                                       |
| stilbs                                                  | sb           | Yes       | 100000 cd/m^2                                                                                  |
| stokes                                                  | St           | Yes       | 0.0001 m^2/s                                                                                    |
| stones                                                  | st           | No        | 6.35029318 kg                                                                                   |
| tablespoons                                              | tbsp         | No        | 1.478676478125E-05 m^3                                                                          |
| teaspoons                                               | tsp          | No        | 4.9289219375E-06 m^3                                                                            |
| tesla                                                   | T            | Yes       | 1 kg/(s^2*A)                                                                                    |
| tesla meter per ampere                                  | T*m/A        | Yes       | 1 m^2*kg/(s^2*A^2)                                                                              |
| tesla per centimeter                                    | T/cm         | No        | 100 kg/(m*s^2*A)                                                                                |
| tesla per meter                                         | T/m          | Yes       | 1 kg/(m*s^2*A)                                                                                  |
| thomson                                                 | Th           | No        | 1.036426957204542E-08 kg/(s*A)                                                                  |
| thousand oil barrels                                    | Mbbl         | No        | 158.987295 m^3                                                                                  |
| tons                                                    | ton          | No        | 907.18474 kg                                                                                    |
| tonnes                                                  | t            | No        | 1000 kg                                                                                         |
| tonsUK                                                  | tonUK        | No        | 1016.046908 kg                                                                                 |
| torrs                                                   | Torr         | No        | 133.3223684210526 kg/(m*s^2)                                                                    |
| townships                                               | twp          | No        | 93239571.972096 m^2                                                                             |
| turns                                                   | tr           | Yes       | 6.283185307179586 m/m                                                                          |
| volts                                                   | V            | Yes       | 1 m^2*kg/(s^3*A)                                                                                |
| volts meter                                             | V*m          | Yes       | 1 m^3*kg/(s*A)                                                                                  |
| volts per centimeter                                    | V/cm         | Yes       | 100 m*kg/(s^3*A)                                                                                 |
| Unit or constant name                          | Symbol     | SI Prefix | Value in Coherent Derived SI Unit                                      |
|---------------------------------------------|------------|-----------|------------------------------------------------------------------------|
| volts per meter                             | V/m        | Yes       | 1 m*kg/(s^3*A)                                                        |
| volts per square meter                      | V/m^2      | Yes       | 1 m^2*kg/(m^2*s^3*A)                                                  |
| von klitzing constant                       | h_P/(q_e^2)| No        | 25812.80745611641 m^2*kg/(s^3*A^2)                                   |
| watts                                       | W          | Yes       | 1 m^2*kg/s^3                                                         |
| watt hour                                   | W*h        | Yes       | 3600 m^2*kg/s^2                                                       |
| watts per cubic centimeter                  | W/cm^3     | Yes       | 100000 m^2*kg/(m^3*s^3)                                               |
| watts per cubic meter                       | W/m^3      | Yes       | 1 m^2*kg/(m^3*s^3)                                                    |
| watts per kilogram                          | W/K        | Yes       | 1 m^2*kg/(s^3*K)                                                      |
| watts per kilogram                          | W/kg       | Yes       | 1 m^2*kg/(kg*s^3)                                                     |
| watts per lumen                             | W/ lm      | Yes       | 1 m^3*kg/(m^3*s cd)                                                   |
| watts per meter kilvin                      | W/(m*K)    | Yes       | 1 m^2*kg/(s^3*K)                                                      |
| watts per nanometer                         | W/nm       | Yes       | 1E+09 m^2*kg/(m*s^3)                                                  |
| watts per square centimeter                 | W/cm^2     | Yes       | 100000 m^2*kg/(m^3*s^3)                                               |
| watts per square foot                       | W/ ft^2    | No        | 10.76391041670972 m^2*kg/(m^2*s^3)                                   |
| watts per square inch                       | W/ in^2    | No        | 0.89699253472581 m^2*kg/(m^2*s^3)                                    |
| watts per square meter                      | W/m^2      | Yes       | 1 m^2*kg/(m^2*s^3)                                                    |
| watts per square meter per kelvin           | W/(m^2*K)  | Yes       | 1 kg/(s^3*K)                                                          |
| watts per square meter per nanometer        | W/(m^2*nm)| Yes       | 1E+09 kg/(m*s^3)                                                      |
| watts per square meter per steradian        | W/(m^2*sr)| Yes       | 1 m^4*kg/(m^4*s^3)                                                    |
| watts per square meter steradian per nanometer | W/(m^2*sr*nm)| Yes       | 1E+09 m^4*kg/(m^5*s^3)                                               |
| watts per steradian                         | W/sr       | Yes       | 1 m^4*kg/(m^2*s^3)                                                    |
| watts per steradian per nanometer           | W/(sr*nm)  | Yes       | 1E+09 m^4*kg/(m^3*s^3)                                                |
| webers                                      | Wb         | Yes       | 1 m^2*kg/(s^2*A)                                                     |
| webers per ampere meter                     | Wb/(A*m)   | Yes       | 1 m*kg/(s^2*A^2)                                                     |
| weeks                                       | wk         | No        | 604800 s                                                              |
| wien wavelength displacement constant       | b_λ        | No        | 0.0028977729 m*K                                                      |
| yards                                       | yd         | No        | 0.9144 m                                                              |
| years                                       | yr         | No        | 31557600 s                                                            |
| ångströms                                   | Å           | No        | 1E-10 m                                                              |
| œrsteds                                     | Oe         | Yes       | 79.57747154594767 A/m                                                 |