

材料科学数据共享工程技术标准

材料科学数据元目录

Data Element Dictionary in Materials Science

（草案大纲）

（本稿完成日期：2009年10月）

2009-10 发布

材料科学数据共享标准规范课题组

材料科学数据元目录

1、引言

材料科学数据共享网是“物理上合理分布，逻辑上高度统一”的资源共享服务系统，其中涉及到金属材料、无机非金属材料、有机高分子材料、复合材料等多个数据资源共享节点的主体数据库的构建与共享。数据元则是共享资源在共享网层面上实现数据的一致性表达、交换和共享的基础。

制定本标准目的在于确立面向材料科学数据共享活动的数据元标准化的基本原则与方法，指导材料科学数据共享网各建设单位规范数据元的定义、命名、著录、管理等各个环节，形成标准化、规范化的材料领域数据元资源库，供本领域科学数据共享的建设与实施使用。

数据元的标准化对于材料领域科学数据共享建设与运行过程中数据的交换、共享具有重要的意义。数据元标识、描述与表达的规范化和通用性，保证了相同概念的数据元在语义上无歧义理解，将提高数据在组织、生产、交换和共享的效率和质量，并推动和促进材料科学数据的集成与共享。

本标准遵循《科学数据共享工程技术标准-数据元标准化的基本原则与方法》(SDS/T 2132—2004)，在《科学数据共享工程技术标准-公共数据元目录》(SDS/T 2131—2004)的基础上，参照相关标准和规范，结合材料领域各种类型科学数据库以及数据交换对数据元标准化的需要，提出了材料科学数据共享工程中涉及到的数据元的内容及其表达。

2 范围

本标准规定了材料科学数据共享工程中涉及到的数据元的内容及其表达，适用于材料科学数据共享工程中的数据采集、加工、处理、交换和共享过程，用于指导材料领域科学数据共享工程各建设单位研究和制定数据元目录使用。

3 规范性引用文件

下列文件中的条款通过本标准的引用而成为本部分的的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准。然而，鼓励根据本标准达成协议的各方，研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 15237.1—2000 术语工作 词汇 第1部分：理论与应用（eqv ISO 1087: 1-2000）

GB/T 18391.1—信息技术 数据元的规范与标准化 第1部分：数据元的规范与标准化框架（idt ISO/IEC 11179-1:1999）

GB/T 18391.2—信息技术 数据元的规范与标准化 第2部分：数据元的分类（idt ISO/IEC 11179-2: 2000）

GB/T 18391.3—2001 信息技术 数据元的规范与标准化 第3部分：数据元的基本属性（idt ISO/IEC 11179-3: 1994）

GB/T 18391.4—2001 信息技术 数据元的规范与标准化 第4部分：数据元定义的编写规则与指南（idt ISO/IEC 11179-4: 1995）

GB/T 18391.5—2001 信息技术 数据元的规范与标准化 第5部分：数据元的命名和标识原则（idt ISO/IEC 11179-5: 1995）

GB/T 18391.6—2001 信息技术 数据元的规范与标准化 第6部分：数据元的注册（ISO/IEC 11179-6: 1997）

SDS/T 2132—2004 科学数据共享数据元标准化的基本原则与方法

SDS/T 2131—2004 科学数据共享公用数据元目录

SDS/T 1003—2004 科学数据共享概念与术语_概念

SDS/T 1003.2—2004 科学数据共享概念与术语_术语

GB 2312 信息交换用自由文本编码字符集 基本集

GB / T 4880.2—2000 语种名称代码 第2部分：字母代码

GB/T 7408 数据元交换格式 信息交换日期和时间表示法（eqv ISO 8601）

GB/T 19253 信息技术 数据元值的格式表示法

4、术语和定义

4.1 数据元 data element

用一组属性描述其定义、标识、表示和允许值的数据单元，在一定语境下，通常用于构建一个语义正确、独立且无歧义的特定概念语义的信息单元。

4.2 数据元概念 data element concept

能以数据元的形式表示，且与任何特定的表示法无关的概念。

4.3 数据元目录 data element dictionary

列出并定义了所有相关数据元的一种信息资源。

4.4 数据元名称 data element name

用于标识数据元的主要手段，由一个或多个词构成的命名。

4.5 标识符 identifier

注册机构内与语言无关的唯一标识符。同数据标识符。给定相关环境的对象的无歧义的名称。

4.6 版本 version

注册机构内，一套数据元规范中的一个数据元发布的标识。

4.7 数据元注册系统 data element registry

由注册机构保存的用于描述数据元含义和表示形式的信息资源，包括注册标识符、定义、名称、值域、元数据和管理属性等。

4.8 实体 entity

任何具体或抽象的事物，包括事物间的联系。

4.9 对象 object

可以想象或感觉的世界的任一部分。

4.10 对象类 object class

现实世界中有清楚的边界和含义、其特性和其行为遵循同样的规则而能够加以标识的想法、抽象概念或事物的集合。

4.11 属性 attribute

某个对象或实体的特征。

4.12 属性值 attribute value

某种属性的一个实例表示。

4.13 定义 definition

表述人和事物的基本特性、或其类别的词或短语。

4.14 相关环境 context

对使用名称或产生名称的应用环境或应用规程的指明或描述。

4.15 特性 property

对象类的所有个体所共有的某种性质。

4.16 特性术语 property term

用于表述对象类特性的术语，它是数据元名称的一个部分。

4.17 限定术语 qualifier term

用于定义和区分数据库中某个名称的词或词组。

4.18 表示术语 representation term

数据元中用于描述数据元表示的术语，它也是数据元名称的一个部分。

4.19 表示 representation

值域、数据类型的组合，必要时也包括度量单位或字符集。

4.20 注册机构标识符 (RAI) registration authority identifier

赋予注册机构的标识符。

4.21 版本标识符 version identifier

赋予版本的一个标识符（用于提交或修改数据元注册）。

4.22 分类模式 classification scheme

依据对象所具有的共同特征，如：起源、组成、结构、应用和功能，而对其进行排列或划分的模式。

4.23 分类的组分 classified component

依据一个或多个分类模式可对其进行分类的一个数据元的任何组分。

注：组分包括对象类、特性、表示类、数据元概念、值域和数据元。

4.24 结构设置 structure set

在相关环境中放置对象的方法，用以展示与其他对象的关系，如各种实体关系模型、分类法 (tax-onomies) 和本体论法 (ontologies)。

4.25 关键字 keyword

用于数据元检索的一个或多个有意义的字、词。

4.26 分类法 taxonomy

依据类及子类间的固有关系进行分类的方法。

4.27 主题词表 thesaurus

按给定顺序排列参照词汇，其中显示和标识了词汇间的关系。

4.28 语义学 semantics

有关词义注释的语言学分支学科（韦氏词典）。

5、数据元的表达格式

本标准中，数据元是通过标记、名称、英文名称、缩略表达、定义、同义名称、表示、计量单位、语境、值域、版本这些属性来表达的，其中缩略表达、计量单位、同义名称和语境是可以根据需要选用的可选项。

- (1) 标记：在数据元目录中，为数据元分配的与语言无关的惟一标识。
- (2) 名称：数据元的中文名称，是标识数据元的主要手段。
- (3) 英文名称：数据元的英文全称。
- (4) 缩略表达：数据元的英文简称。(可选)

(5) 定义：关于数据元的含义和基本特性的说明，并使之区别于其他数据元。

(6) 同义名称：本数据元的其他许用名称。（可选）

(7) 表示：数据元值的数据类型及字符长度的组合表示格式。参见5。

(8) 计量单位：数值型数据元值的计量单位。（可选）

(9) 语境：产生或使用数据元的应用环境的说明。（可选）

(10) 值域：数据元取值范围或相关内容的说明。如所采用的代码标准，或通过枚举等方式给出所有可能的取值。

(11) 版本：数据元所在数据元目录的版本标识。

6、数据元值的表示方法

本标准中，数据元值基本采用GB/T 19253规定的的数据元值的数据类型与字符长度表示方法，根据需要做了少量修改和补充。包括数据类型表示法和字符长度表示法，还给出一些表示实例。

本标准数据元值的数据类型采用表 1 所示的方法表示。

表 1 数据类型表示法

表示符	数据类型	说 明
S	字符型	可以包含汉字(中、国……等)、字母字符(a-z,A-Z)、数字字符等。 (默认 GB 2312信息交换用汉字编码字符集基本集)
N	数值型	数字形式表示的数值。
T	日期时间型	采用 GB/T 7408 中规定的 YYYYMMDD 、 YYYYMMDDThhmmss 格式。

5.2 字符长度表示法

本标准数据元值的字符长度采用表 2 规定的方法表示。

表 2 字符长度表示法

分类	表示方法
固定长度	在数据类型表示符后直接给出字符长度的数目。
可变长度	1) 可变长度不超过定义的最大字符数。在数据类型表示符后加 “..” 后给出数据元最大字符数目。 2) 可变长度在定义的最小和最大字符数之间。在数据类型表示符后给出最小字符长度数后加 “..” 后，再给出最大字符数。
有若干字符行表示的长度	按固定长度或可变长度的规定给出每行的字符长度数后加 “X” 后，再给出最大行数。

5.3 应用示例

(1) S 字符型

S12 固定长度为 12 个字符（相当于 6 个汉字）长度的字符。

S..12 可变长度，最大为 12 个字符长度的字符。

S4..12 可变长度，最小为 4 个字符，最大为 12 个字符长度的字符。

S..40X3 3 行最大长度为 40 个字符长度的字符。

(2) N 数值型

N3 固定长度为 3 位数字。

N..3 最大长度为 3 位数字。

N9, 2 最大长度为 9 位的十进制小数格式（包括小数点），小数点后保留 2 位数字。

(3) T 日期时间型

T8 采用 YYYYMMDD 格式（8 位定长）表示年月日。

如 1998 年 1 月 8 日，应表示为 19980108。

T15 采用 YYYYMMDDThhmmss 格式（15 位定长）表示年月日时分秒。时分秒之前加大写

字母“T”。如 2004 年 3 月 25 日 9 时 8 分 5 秒为 20040325T090805。

7、数据元分组

对数据元进行分类有几个目的：分类可帮助用户从众多的数据元中找出某个单一的数据元；方便对数据元进行数据管理分析；通过继承使原本借助其他属性（如名称和定义）不能完整表述的语义内容得以表达。

本小节给出数据元分组方案和数据元标记的分配方式。

8、数据元目录

在数据的应用、管理和交换中，数据元目录的编写时很重要的手段和途径之一。本小节依照分组方式给出详细的数据元描述。

9、数据元的维护与管理

对于数据元的管理，采用动态维护的办法。注册机构、主管机构及提交机构的设定、作用和职责以及数据元的注册程序参考《GB/T 18391.6—2001 信息技术 数据元的规范与标准化 第6部分：数据元的注册》。