

ICS 01.040.03; 03.100.01

CCS A 20



团 体 标 准

T/CSPSTC XXXX—2024

基于知识图谱的突发事件应对标准数字化 技术规程

Digital technical code for emergency response standards based on knowledge graph

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2024 - XX - XX 发布

2024 - XX - XX 实施

中国科技产业化促进会 发布
中国标准出版社 出版

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 总体要求	2
4.1 通用性	2
4.2 基础性	2
4.3 规范性	2
5 基本流程	2
6 突发事件应对标准知识建模	2
6.1 标准结构要素知识建模	3
6.2 标准核心技术内容知识建模	3
6.3 知识建模活动输出	4
6.4 质量评估	4
7 突发事件应对标准知识获取	4
7.1 知识获取内容和要求	4
7.2 文本内容知识获取	4
7.3 表格内容知识获取	5
7.4 图内容知识获取	5
7.5 数学公式知识获取	5
7.6 质量评估	5
8 突发事件应对标准知识融合	6
8.1 实体对齐	6
8.2 知识一致性校验	6
9 突发事件应对标准知识存储	6
9.1 执行存储操作	6
9.2 存储管理	6
10 突发事件应对标准知识演化	6
10.1 知识更新	6
10.2 知识图谱版本管理	6
11 数字化成果验收	6
11.1 验收方式	7
11.2 验收内容	7
11.3 验收指标	7
11.4 验收结论	7
附录 A（资料性） 标准结构要素实体类别的属性和关系	8

附录 B（资料性） 突发事件应对领域本体模型 9

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由中国标准化研究院提出。

本文件由中国科技产业化促进会归口。

本文件起草单位：中国标准化研究院、中国矿业大学（北京）、北京市科学技术研究院、北京邮电大学、清华大学、应急管理部大数据中心、湖北省标准化与质量研究院、福建省特种设备检验研究院、北京辰安科技股份有限公司

本文件主要起草人：

基于知识图谱的突发事件应对标准数字化技术规程

1 范围

本文件规定了采用知识图谱技术进行突发事件应对标准数字化处理的总体要求，以及基本流程、知识建模、知识获取、知识融合、知识存储、知识演化和成果验收等要求。

本文件适用于采用知识图谱技术对突发事件应对标准进行数字化加工和处理。其他文件可参考应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 37967 基于XML的国家标准结构化置标框架

3 术语和定义

下列术语和定义适用于本文件。

3.1

标准数字化 standard digitalization

利用信息技术对标准进行结构化和语义化处理，转换为计算机可理解和可处理的形式，使其能够通过数字设备进行读取、传输和使用的过程。

3.2

突发事件应对 emergency response

为了预防和减少突发事件的发生，控制、减轻和消除突发事件引起的严重社会危害，规范突发事件应对活动，保护人民生命财产安全，维护国家安全、公共安全、环境安全和社会秩序所采取的一系列活动。

3.3

知识建模 knowledge modeling

构建知识图谱的本体及其形式化表达的活动。

注：知识建模活动可包括实体类型定义、关系定义及属性定义。

[来源：GB/T 42131—2022，3.19]

3.4

知识获取 knowledge acquisition

从不同来源和结构的输入数据中提取知识的活动。

注：知识获取的数据源通常按数据组织结构的维度可分为结构化数据、半结构化数据、非结构化数据（如纯文本、音频和视频数据等）。

[来源：GB/T 42131—2022，3.20]

3.5

知识融合 knowledge fusion

整合和集成知识单元（集），并形成拥有全局统一知识标识的知识图谱的活动。

[来源：GB/T 42131—2022，3.21]

3.6

知识存储 knowledge storage

设计存储架构，并利用软硬件等基础设施对知识进行存储、查询、维护和管理的活动。

注：常见的知识存储方式分为：基于关系数据库的存储方式、基于图数据库的存储方式、基于资源描述框架（RDF）数据库的存储方式等。

[来源：GB/T 42131—2022，3.22]

3.7

知识演化 knowledge evolution

随本体模型、数据资源等变化产生的新知识对原有知识的补充、更新或重组的活动。

[来源：GB/T 42131—2022，3.25]

4 总体要求

4.1 各类型突发事件应对标准数字化处理应以本文件为基础，并结合自身特征进行细化和完善。

4.2 数字化处理应保留标准文本结构的完整性和内容的规范性，同时满足知识的细粒度表示。

5 基本流程

基于知识图谱的突发事件应对标准数字化的基本流程见图1，主要包括以下步骤。

- 知识建模：突发事件应对标准知识建模包括结构要素和核心技术内容知识建模。对于突发事件应对标准结构要素，应基于标准文本逻辑结构进行知识建模；对于突发事件应对标准核心技术内容，应通过梳理突发事件应对领域知识，结合突发事件应对标准和实战应用需求进行知识建模。
- 知识获取：先对突发事件应对标准进行数据预处理，之后针对突发事件应对标准内容组成和表示方式，分别采用相应的知识抽取方法实现突发事件应对标准的知识抽取。
- 知识融合：对于已获取的突发事件应对标准知识，通过实体对齐和知识一致性检验对知识进行整合和校验。
- 知识存储：将经过知识获取和知识融合后的突发事件应对标准知识存入数据存储系统。
- 知识演化：针对版本修订和新发布的突发事件应对标准，进行突发事件应对标准知识更新，并对知识图谱的版本进行管理。

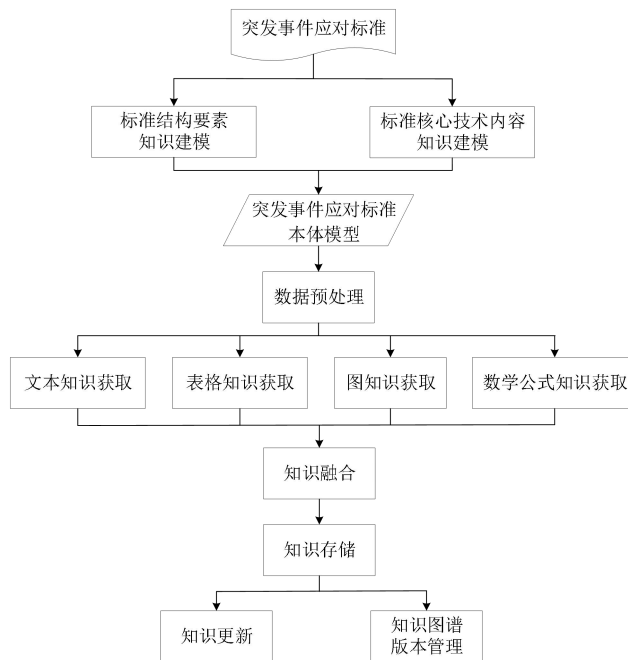


图1 基于知识图谱的突发事件应对标准数字化基本流程

6 突发事件应对标准知识建模

6.1 标准结构要素知识建模

6.1.1 突发事件应对标准结构要素包括：

- a) 封面（必备）：文件名称、文件的层次或类别、文件代号、文件编号、国际标准分类（ICS）号、中国标准文献分类（CCS）号、发布日期、实施日期、发布机构等；
- b) 目次（可选）：前言、引言、章编号和标题、条编号和标题，附录编号、“（规范性）”/“（资料性）”和标题、附录条编号和标题，参考文献、索引等；
- c) 前言（必备）：起草依据、与其他文件的关系、提出单位、归口单位、起草单位和起草人等；
- d) 引言（可选）：编制原因、编制目的等；
- e) 规范性引用文件（必备）：引导语和文件清单；
- f) 术语和定义（必备/可选）：术语条目；
- g) 符号和缩略语（可选）：符号和缩略语条目；
- h) 分类和编码/系统构成（可选）；
- i) 总体原则和/或总体要求（可选）；
- j) 核心技术要素（必备）：章、条、段等；
- k) 其他技术要素（可选）；
- l) 参考文献（可选）；
- m) 索引（可选）。

6.1.2 应遵循 GB/T 1.1 中对于各结构要素内容和逻辑结构的规定，进行突发事件应对标准结构要素知识建模。

6.1.3 突发事件应对标准结构要素知识建模的过程包括但不限于如下内容：

- a) 确定标准结构要素知识范畴内的关键术语：通过整理标准结构要素中所包含的共性要素，形成标准结构要素知识范畴内的关键术语；
- b) 定义实体类别：基于突发事件应对标准，结合标准结构要素知识范畴内的关键术语，定义突发事件应对标准结构要素的实体类别；
- c) 定义实体类别的属性和关系：通过分析知识的内在含义和关联关系确定实体类别的属性和关系。标准结构要素实体类别的属性和关系可参考附录 A；
- d) 确定并创建本体模型：通过整合突发事件应对标准各结构要素中的实体类别层次体系、属性和关系，创建突发事件应对标准结构要素本体模型。

6.2 标准核心技术内容知识建模

6.2.1 突发事件应对标准的核心技术内容主要包括突发事件应对标准核心技术要素部分的内容，包括但不限于以下知识：突发事件分类、等级判定、组织架构、应急流程、应急处置措施、原则和要求等。

6.2.2 突发事件应对标准核心技术内容知识建模应基于突发事件应对领域知识，结合突发事件应对标准内容和实战应用需求，进行知识体系梳理和本体模型构建。

6.2.3 突发事件应对标准核心技术内容知识建模的过程如下：

- a) 整合突发事件应对标准应用需求：梳理用户意图，对突发事件应对标准应用需求进行识别和整合；
- b) 梳理突发事件应对领域知识：针对突发事件应对全流程（预防与应急准备、监测与预警、应急处置与救援、恢复与重建），对突发事件应对领域知识进行梳理；
- c) 确定现有可复用本体模型：查找与突发事件应对领域相关的现有本体模型，并对其适用性、覆盖范围、可拓展性和可信度进行评估，确定可复用本体模型；
- d) 整理突发事件应对领域关键术语：梳理突发事件应对领域的关键术语，抽象出突发事件应对领域共性知识；
- e) 构建实体类别层级体系：基于现有可复用本体模型，通过扩展和继承的方式，对突发事件应对领域中所包含的实体类别进行层次梳理，形成突发事件应对领域实体类别层次体系；
- f) 定义实体类别的属性与关系：对实体类别层次关系之外的关系进行梳理和定义，并依据语义和是否与其他知识产生关联定义实体类别的属性；
- g) 定义应用需求相关的规则、公理等：面向突发事件应对标准应用需求，建立相关的规则和公理，辅助突发事件应对标准知识推理；

- h) 确定并创建本体模型及图式：通过整合实体类别层次体系、属性和关系、规则和公理等，建立突发事件应对标准核心技术内容本体模型。突发事件应对领域本体模型可参考附录 B。

6.3 知识建模活动输出

- 6.3.1 根据突发事件应对标准结构要素知识建模流程，输出突发事件应对标准结构要素本体模型。
- 6.3.2 根据突发事件应对标准核心技术内容知识建模流程，输出突发事件应对标准核心技术内容本体模型。
- 6.3.3 基于突发事件应对标准结构要素与核心技术内容之间的逻辑关联，包括但不限于来源于、引用等，建立突发事件应对标准结构要素本体模型与核心技术内容本体模型中实体之间的关联关系，输出突发事件应对标准本体模型。

6.4 质量评估

- 6.4.1 突发事件应对标准本体模型应满足合理性、可用性、完整性、可拓展性、兼容性、可复用性和简洁性。
- 6.4.2 应采用专家评估或其他方式对突发事件应对标准本体模型的质量进行评估。
- 6.4.3 应依据突发事件应对标准本体模型质量评估意见实施质量改进，对突发事件应对标准本体模型进行完善和优化。

7 突发事件应对标准知识获取

7.1 知识获取内容和要求

- 7.1.1 突发事件应对标准知识获取内容主要包括标准中文本内容知识获取、表格内容知识获取、图内容知识获取和数学公式知识获取。
- 7.1.2 应基于突发事件应对标准本体模型，结合突发事件应对标准结构要素和核心技术内容，对突发事件应对标准知识进行细粒度拆解，实现知识单元的输出。
- 7.1.3 根据突发事件应对标准内容组成和表示方式的不同，应采用相应知识获取方法进行知识单元的提取。

7.2 文本内容知识获取

- 7.2.1 突发事件应对标准中文本内容知识获取可划分为标准结构要素文本内容和核心技术文本内容知识获取。
- 7.2.2 突发事件应对标准结构要素具有统一的内容组成和表述形式，可按照 GB/T 37967 进行标准结构化处理，并对结构化文件解析，获取突发事件应对标准结构要素知识；也可采用基于规则的方法进行知识单元的提取，步骤如下：
- 规则定义：根据突发事件应对标准结构要素本体模型，定义一组规则来指定要抽取的知识模式；
 - 文本预处理：对突发事件应对标准进行预处理，例如：去除无关字符、分词等，将文本转化为适合规则匹配的形式；
 - 规则匹配：将定义好的规则应用于预处理后的突发事件应对标准文本，可采用正则表达式、模式匹配或其他匹配算法来进行规则匹配，实现知识单元的提取；
 - 三元组或多元组提取：根据突发事件应对标准结构要素本体模型，基于提取后的知识单元识别出实体、属性和关系，形成三元组或多元组。
- 7.2.3 突发事件应对标准核心技术内容由具有复杂特征的突发事件应对领域知识组成，宜采用基于规则的方法和基于深度学习的方法相结合进行知识获取。对于具体应急处置措施的要求、原则和注意事项等知识，可采用基于规则的方法进行知识抽取；对于突发事件应对标准中的突发事件分类、应急流程、应急处置措施等知识，可采用基于深度学习的方法进行知识获取。
- 7.2.4 采用基于深度学习的方法进行突发事件应对标准核心技术内容知识获取的流程包括：
- 文本预处理：对突发事件应对标准进行文本预处理，例如：去除无关字符、分词、语句成分补充等；

- b) 模型设计：根据突发事件应对标准内容，选择合适的知识抽取模型进行知识抽取；
- c) 数据标注：依据突发事件应对标准本体模型，设计标签，形成统一的数据标注规则，对突发事件应对标准进行数据标注；
- d) 模型训练和调优：基于已标注的突发事件应对标准文本，对深度学习模型进行训练，并使用各种优化算法和损失函数来优化模型的参数；
- e) 知识抽取：使用训练好的深度学习模型对突发事件应对标准进行知识抽取。

7.3 表格内容知识获取

7.3.1 突发事件应对标准中表格内容具有相对统一的格式规范，但涉及各类型突发事件应对知识，具有复杂特征，知识获取时需结合表格上下文信息。

7.3.2 根据表格内容知识的复杂性和格式的规范性，宜采用人工辅助和计算机自动抽取相结合的方式进行知识获取，包括以下步骤：

- a) 规则定义：基于不同表格类别，结合突发事件应对领域知识，根据表头设计不同规则建立表头之间的关联；
- b) 自动提取知识：依据表头之间的关联关系建立对应表格内容之间的关系，实现知识的自动抽取；
- c) 人工校对：基于突发事件应对标准中表格内容上下文，结合专家经验和领域知识，对自动抽取的知识进行检查和修正。

7.4 图内容知识获取

7.4.1 突发事件应对标准中图包括流程图、线图和平面图等，图内容知识获取应满足知识细粒度的要求，同时保证图内容知识完整性。

7.4.2 突发事件应对标准中图内容知识获取的流程包括：

- a) 图格式转化：将突发事件应对标准中的图转化为清晰和可存储的数据形式；
- b) 图理解：采用计算机视觉技术和光学字符识别技术进行图理解，抽取出图中所包含的知识，并进行人工校对；
- c) 创建图标识符：根据抽取得到的图知识，结合突发事件应对标准中图的上下文信息，为图创建一个特定的标识符，用于描述图信息；
- d) 创建实体：创建一个实体用于表示图，可将图标识符作为实体的属性，并将抽取得到的图知识转化为三元组或多元组；
- e) 关联其他实体：通过定义适当关系，将抽取得到的图知识和图标识符与突发事件应对标准中其他实体建立关联。

7.5 数学公式知识获取

突发事件应对标准中数学公式的知识获取，包括以下步骤：

- a) 公式表示：确定突发事件应对标准中数学公式的表示方式，如 LaTeX、MathML 等，将突发事件应对标准中的数学公式进行格式转化；
- b) 创建标识符：依据数学公式在突发事件应对标准中所表示的含义，为数学公式创建一个唯一的标识符，以记录数学公式的基本信息；
- c) 创建实体：创建一个实体用于表示数学公式，可将数学公式标识符作为实体的属性；
- d) 关联其他实体：通过定义适当的关系，将数学公式标识符与突发事件应对标准中其他相关的实体建立关联。

7.6 质量评估

突发事件应对标准知识获取的质量评估工作，可从以下方面开展：

- a) 通过将获取知识单元与人工标注数据进行比较，采用精确率、召回率和 F1 测量值等指标评价突发事件应对标准知识获取质量；
- b) 依据突发事件应对标准本体模型中关系定义进行关系约束校验；
- c) 依据突发事件应对标准本体模型中属性定义进行属性约束校验。

8 突发事件应对标准知识融合

8.1 实体对齐

可通过识别出突发事件应对标准知识获取活动输出的知识单元中等效的实体进行知识融合,实现实体对齐,包括以下步骤:

- a) 通过实体归一将突发事件应对标准中相同实体的不同标识符映射到一个统一的标识符上,实现实体的唯一性和一致性;
- b) 根据突发事件应对标准上下文和语义信息,通过实体消歧解决可能存在多个相同名称或相似名称的实体的歧义性;
- c) 通过属性值对齐和融合将突发事件应对标准中相应的属性值进行对应、关联和整合,实现属性值的一致性和完整性。

8.2 知识一致性校验

通过检查和验证突发事件应对标准知识获取活动输出的知识是否存在矛盾或冲突,以确保知识的一致性和可靠性,知识一致性检验工作,包括但不限于:

- a) 实体一致性校验;
- b) 属性一致性校验;
- c) 关系一致性校验;
- d) 逻辑一致性检验。

9 突发事件应对标准知识存储

9.1 执行存储操作

突发事件应对标准知识存储工作,包括但不限于:

- a) 根据突发事件应对标准存储和应用需求选择数据存储系统;
- b) 将经过知识获取和知识融合的突发事件应对标准知识录入数据存储系统;
- c) 基于数据存储系统实现突发事件应对标准知识图谱的存储、可视化和查询等功能。

9.2 存储管理

突发事件应对标准知识存储管理工作,包括但不限于:

- a) 对突发事件应对标准知识存储过程进行记录;
- b) 建立数据安全机制,保障数据存储安全。

10 突发事件应对标准知识演化

10.1 知识更新

10.1.1 针对已修订的突发事件应对标准,应对突发事件应对标准知识图谱中原有知识进行变更和修正,确保知识的准确性和时效性。

10.1.2 针对新发布的突发事件应对标准,应通过知识新增的方式向知识图谱中添加相应的知识。

10.1.3 知识新增应按照规范的突发事件应对标准数字化流程进行,同时应严格控制与原有知识的关联性和冲突性。

10.2 知识图谱版本管理

突发事件应对标准知识图谱版本管理工作,包括但不限于:

- a) 对突发事件应对标准知识更新过程进行记录;
- b) 生成新知识的存储路径和知识演化日志;
- c) 对原版本突发事件应对知识图谱进行冻结或淘汰。

11 数字化成果验收

11.1 验收方式

应采用计算机自动检验和人机交互检验相结合的方式对基于知识图谱的突发事件应对标准数字化成果进行验收。

11.2 验收内容

11.2.1 应对最终形成的突发事件应对标准知识图谱进行验收,包括知识的准确性、完整性、规范性和可用性等。

11.2.2 应对突发事件应对标准数字化处理各环节的输出结果进行验收,包括输出结果的质量和完整性。

11.2.3 应对突发事件应对标准数字化处理的过程记录进行验收,包括数字化处理过程的规范性和完整性。

11.3 验收指标

抽检比率应达到总条目的5%以上(含5%),合格率应达到95%以上(含95%),合格率为抽检合格条目数占抽检条目总数的百分比。

11.4 验收结论

11.4.1 每个突发事件应对标准数字化成果质量检验达到 11.2 和 11.3 的要求,予以验收“通过”。验收未通过应视情况进行返工或修改后,重新进行验收。

11.4.2 验收完成后须经验收组成员签字。验收“通过”的结论,应经相关负责人审核、签字后方有效。

附录 A
(资料性)

标准结构要素实体类别的属性和关系

表A.1给出了标准结构要素实体类别的属性。

表 A.1 标准结构要素实体类别的属性

实体类别	属性
标准中文名称	标准英文名称
	国际标准ICS分类号
	发布时间
	实施时间
	状态
	中国标准文献分类 (CCS) 号
术语	术语英文
	术语定义
章标题	内容
条标题	内容

表A.2给出了标准结构要素实体类别间的关系。

表 A.2 标准结构要素实体类别间的关系

实体类别	关系	实体类别
标准中文名称	标准号	标准号
	提出于	提出单位
	归口于	归口单位
	起草于	起草单位
	起草人	起草人
	规定了	规定内容
	适用于	适用界限
	涉及术语	术语
	包含	章标题
	引用了/更新	标准中文名称
术语	定义	术语定义
章标题	包含	条标题
条标题	来源于	标准中文名称

附录 B
(资料性)
突发事件应对领域本体模型

突发事件应对领域本体模型见表B.1。

表 B.1 突发事件应对领域本体模型

类别	一级	二级	三级	四级		
抽象类	命题	方法				
		术语				
		分级依据				
		要求				
	关系	具有类关系	具有类关系			
			结构类关系			
			行动类关系			
			继承类关系			
			时序类关系			
	属性	承灾体属性	承灾体属性	受灾群体属性		
				财产属性		
				资源与环境属性		
		应急资源属性	应急资源属性	救援人员属性	角色	
					责任	
				应急物资属性	型号	
					编码	
		数量	数量	常量		
				单位		
				函数		
	一般属性	一般属性	目的			
			作用			
			条件			
	图	图	有向图			
			伪图			
			组合图			
			树状图			
			循环图			
图元素	图元素	节点				
		图弧				
		循环				
集合或类	集合或类	突发事件级别				
		突发事件分类				
		应急物资分类				
		救援主体级别				
		救援主体分类				
		应急响应级别				
物质类	对象	主体	个人			
			组织机构			
			体系			
		危险源				
		承灾体	承灾体	受灾群体		
				财产		
				资源与环境		
		物资(资源)		基本生活物资		

			应急装备及配套物资		
			工程材料与机械加工设备		
			信息		
		其他	孕灾环境		
			原则		
			内容		
			行动		
		过程	预防与应急准备	风险管理	
				机构与人员准备	
				物资准备	
			监测与预警	监测	
				预警	
			应急处置与救援	指挥研判	
				信息报送	
				信息发布	
	行动救援				
	事后恢复与重建		应急保障		
		调查评估			
		恢复重建			
	全生命周期	救助			
	文本	法律法规			
		政策文件			
		应急预案			
各级标准					
事故报告					

参 考 文 献

- [1] GB/T 38548.1—2020 内容资源数字化加工 第1部分：术语
 - [2] GB/T 42131—2022 人工智能 知识图谱技术框架
-