ICS 01.140.20
CCS A 14

# 团　　体　　标　　准

T/GDES XX—2026

# 标准开源格式规范

Standard open source format specification

（征求意见稿）

（本草案完成时间：）

**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。**

2026－XX－XX 发布　　2026－XX－XX 实施

广东省节能减排标准化促进会　　发　布

# 目　　次

前言 ..... II
引言 ..... III
1　范围 ..... 4
2　规范性引用文件 ..... 4
3　术语和定义 ..... 4
4　标准包结构 ..... 5
5　标准概念文件 ..... 7

# 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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由广东财经大学提出。

本文件由广东省节能减排标准化促进会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

# 引　言

## 0.1 技术背景

2026年，大语言模型（Large Language Model，LLM）基于Transformer架构的自注意力机制，聚焦于整合文本与视觉的端到端架构，并融入强化学习，赋予类人的慢思考与推理能力，将百万级超长上下文能力进行开源普惠化。Agent能够连续自主编程60分钟以上，无需人工干预即可完成数据库设计、安卓模拟器开发等复杂工程任务，具备出色的长程规划与自我迭代能力，特别是OpenClaw从“传统聊天机器人”动口角色及到“数字员工”动手角色的突破，以及各地扶持政策的落地，Agent已到了迈向“规模化落地”的关键时期。

综上所述，基于大语言模型，沿着“LLM →AI ChatBot → Prompt → Context → Harness→ Loop”技术发展跃进路线，及“Token”、“API”、“MCP”、“CLI”、“Skill”词涌现，代表着AI工程化从“调教模型”到“设计系统”的演进路径。

在2024年搭建的标准开源社区（http://www.perfectquality.cn/）的基础上，围绕核心标准，定位于标准实际应用，基于现有大语言模型应用的技术环境，采用以下技术措施来建设标准智能体（Standard Agent）：

——标准规范的内容，基于现有大语言模型应用的技术环境，采用 markdown 格式来标识标准内容，即可实现机器可显示标准文档、机器可读取标准文档、机器可读取标准内容、机器可执行标准内容、机器可解释标准内容、机器可更新标准内容；

——标准应用的资源，主要包括知识库、开源代码和元数据格式发布，统一规范标准应用资源；

——标准参考的案例，主要包括数据集及数据库和典型案例，示范推广标准实际应用；

——标准更新的机制，依据大语言模型 markdown 格式的更新机制，借助 Agent 现有技术底座实现标准自主更新。

## 0.2 标准开源格式

SOSF（Standard Open Source Format，SOSF）是一种开放的、对人类和智能体都友好的格式，用于表示标准：围绕标准、数据和系统的元数据、上下文和精选洞察。

该规范的设计为由人类编写、由智能体生成、跨组织交换，并由两者共同使用。该格式刻意保持极简：一个包含带 YAML 头部元数据的 Markdown 文件的目录。没有模式注册表，没有中央权威机构，也不需要专用工具。如果你能用cat读取文件，你就能阅读SOSF；如果你能用 git clone 克隆仓库，你就能分发它。

本规范参考和基于Google cloud 2026年6月12日发布的开放知识格式OKF，是一套基于文件夹+Markdown的文件规范，通过定义一种通用标准的文档类格式，生成者（人类、智能体、导出管道）可以写入，指导使用者（智能体、用户界面、搜索索引、确定性代码）应如何阅读和遍历它，促进跨系统和组织的知识交换。目的是实现人类、AI Agent通用，可跨平台交换的面向标准的知识库规范，具体体现以下特性：

——可读性，人类可读，无需工具即可阅读；

——可解析，智能体可解析，无需定制 SDK；

——可对比，可在版本控制中对比差异；

——可移植，可跨工具、跨组织、跨时间移植。

## 0.3 标准开源应用

标准开源社区依据标准开源格式SOSF制作和发布markdown格式标准数据，实现标准数据开源。然后依据大语言模型，借助Agent现有技术底座，开发建设标准应用的AI智能体。

标准AI智能体的开发建设计划：零碳工厂智能体、零碳园区智能体、VCER碳信用资产智能体等SOSF兼顾人类可读+AI可自动解析，轻量化，适合Agent长期知识库落地。

# 标准开源格式规范

## 1 范围

本文件给出了标准开源格式的标准包结构和标准概念文件等内容。

本文件适用于基于大语言模型的AI智能体标准知识库建设，其他文档类的AI智能体知识库建设建设也可参照执行。

## 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 22373—2021 标准文献元数据

## 3 术语和定义

下列术语和定义适用于本文件。

3.1

**OKF 术语**

3.1.1

**知识包 knowledge Bundle**

一个自包含的、分层的知识文档集合。分发的单位。整个知识库为一个根目录（可Git仓库、zip压缩包），支持多层子目录。

3.1.2

**知识概念 knowledge Concept**

包内的单个知识单元，表示为一个 Markdown 文档。它可以描述有形资产（表、API）、抽象概念（指标、业务流程），或介于两者之间的任何事物。文件相对路径为概念全局唯一ID。

**示例**：tables/orders.md → ID： tables/orders3. 文件编码强制：UTF-8。

3.1.3

**头部元数据 YAML frontmatter**

Markdown 文件顶部由 `---` 分隔的 YAML 元数据块。

3.1.4

**正文 body**

文件中头部元数据之后的所有内容。

3.1.5

**链接 link**

从一个概念到另一个概念的标准 Markdown 链接，用于表达隐式父子层次结构之外的关系。

3.1.6

**来源 source**

概念衍生自的材料，可以是包内或包外的，记录在 `sources` 头部元数据字段中。

3.1.7

**来源追溯 provenance**

概念衍生自的来源集合。

3.1.8

**可信度信号 credibility signal**

每个来源的客观事实（`author`、`usage_count`、`last_modified`），用于推断信任度；为记录信号，而非结论。

3.1.9

**角色 actor**

标识谁或什么执行了操作的字符串，对智能体使用 `<producer>/<version>` 约定，对人类使用 `human:<id>`，对自动化流程使用 `process:<id>`。

3.1.10

**信任层级 trust tier**

从概念的 `verified` 字段得出的级别：未验证、机器确认或人工审核。

3.1.11

**认证计算 attested computation**

一种概念（`type: Attested Computation`），承载一种经认可的计算值的方式，以便使用者可以确认该值是通过运行它生成的。

3.1.12

**执行器 executor**

运行指令或代码，执行计算并返回回执。

3.1.13

**回执 receipt**

运行返回的证据，由 `executor.receipt` 定义；是运行时产物，不存储在包中。

3.1.14

**认证器 attester**

确定性（非 LLM）代码，检查回执并返回结论。

3.2

**SOSF 术语**

3.2.1

**标准包 standard bundle**

作为知识包的一种类型，一份标准可以转化为标准包，整个标准报为一个根目录（可Git仓库、zip压缩包），支持多层子目录。

3.2.2

**标准概念 standard concept**

作为知识概念的一种类型，一份标准可以作为一个标准概念，表示为一个 Markdown 文档。文件相对路径为概念全局唯一ID。

**示例：**tables/orders.md → ID： tables/orders3. 文件编码强制：UTF-8。

## 4 标准包结构

### 4.1 包结构

包是 Markdown 文件的目录树，如图1所示。目录结构与领域无关：生成者可以按照对所捕获知识有意义的任何方式组织概念。

```
path/to/bundle/
index.md              # 可选。用于渐进式披露的目录列表。
log.md                # 可选。按时间顺序的更新历史。
<concept>.md          # 包根目录的一个概念。
<subdirectory>/     # 子目录将概念组织成组。
  index.md
  <concept>.md
  <subdirectory>/
 ...
```

图1 OKF 知识包结构目录

包可以以下列方式分发：

——Git 仓库（推荐，因为它提供历史记录、归属和差异对比）；

——目录的 tarball 或 zip 归档；

——更大仓库中的子目录。

保留文件名以下文件名在层次结构的任何级别都有定义的含义，`index.md`目录列表文件和`log.md`更新历史文件不得用于概念文档。

## 4.2 标准包结构

标准包结构依据包结构，可参考如图2所示。

```
path/to/standard-bundle/
index.md              # 用于渐进式披露的目录列表文件（Agent 优先读取），可选。
log.md                # 按时间顺序的更新历史文件，可选。
<standard>.md         # 包根目录的一个概念文件，必选。
<subdirectory>/       # 子目录将概念组织成组，可选
  index.md            # 文件，目录索引（Agent 优先读取），可选
  <concept>.md        # 文件，一个概念，可选。
  <datasets>/         # 子目录，数据集目录，可选
    index.md          # 文件，目录索引（Agent 优先读取），可选
    user_dataset.md   # 文件，具体用户数据集文件，可选
  <tables>/           # 子目录，表目录，可选
    index.md          # 文件，目录索引（Agent 优先读取），可选
    user_ table.md    # 文件，具体用户表文件，可选
  <metrics>/          # 子目录，指标目录，可选
    index.md          # 文件，目录索引（Agent 优先读取），可选
    user_ metric.md   # 文件，具体用户指标文件，可选
  <references>/       # 子目录，参考模板文档目录，可选
    index.md          # 文件，目录索引（Agent 优先读取），可选
    user_reference.md # 文件，具体用户参考模板文档文件，可选
  <scripts>/          # 子目录，脚本目录，可选
    index.md          # 文件，目录索引（Agent 优先读取），可选
    user_ script.md   # 文件，具体用户脚本文件，可选
  <examples>/         # 子目录，案例文档目录，可选
    index.md          # 文件，目录索引（Agent 优先读取），可选
    user_example.md   # 文件，具体用户案例文档文件，可选
  <subdirectory>/     # 子目录，可选
  ...
```

图2 SOSF 标准包结构目录

# 5 标准概念文件

## 5.1 概念文件

### 5.1.1 概念文件组成

每个概念都是一个UTF-8 Markdown 文件，包含两部分：
——YAML 头部元数据块，由文件开头单独一行的 `---` 和结束的单独一行 `---` 分隔；
——Markdown 正文，包含自由格式内容。

### 5.1.2 头部元数据

头部元数据如图3所示。

```
```
YAML
---
type: <类型名称>                    # 必填
title: <可选显示名称>
description: <可选单行摘要>
resource: <可选底层资产的规范 URI>
tags: [<标签>, <标签>, ...]          # 可选
# ... 信任、生命周期、来源和计算系列
# ... 其他生成者定义的键/值对
---
```
```

图3 OKF 头部元数据结构图

以上元数据涉及的详细说明见表1所示。

表1 OKF 概念文件头部元数据的字段

| 字段 | 名称 | 数据类型 | 说明 |
|---|---|---|---|
| type | 类型，标识概念种类的短字符串 | 示例值：`BigQuery Table`、`BigQuery Dataset`、`API Endpoint`、`Metric`、`Playbook`、`Reference`、`Attested Computation`、`Document` | 必填 |
| title | 题目，人类可读的显示名称 | 字符串 | 可选 |
| description | 描述，总结概念的一句话 | 字符串 | 可选 |
| resource | 唯一标识概念所描述的底层资产的 URI | 字符串 | 可选 |
| tags | 标签，用于跨领域分类的短字符串的 YAML 列表 | 字符串 | 可选 |
| resource | 概念的来源 | 字符串 | 可选 |
| generated | 记录当前内容是如何产生的。 | 字符串 | 可选 |
| verified | 记录谁或什么已经根据其来源或resource确定了内容，例如级别：未验证、机器确认或人工审核。 | 字符串 | 可选 |
| status | 与生命周期有关的状态，例如：draft，stable，deprecated | 字符串 | 可选 |
| computation | 认证计算概念的字段 | 字符串 | 可选 |

#### 5.1.3 正文

正文是标准 Markdown。生成者应倾向于结构化Markdown（标题、列表、表格、围栏代码块）而非自由形式的散文，因为结构有助于人类阅读和智能体检索。

没有必填的正文部分。

### 5.2 标准概念文件

#### 5.2.1 头部元数据

表1中的头部元数据type取值为“Document”，title取值为“standard”，tags标签参见GB/T 22373-2021选取：Number、Name、Announcing_body、publishing_date、effective_date、ICS_number、CCS_number、industry、drafting_body、drafting_person、scope、patent等字段，见表2。

**表2 SOSF 标准概念文件头部元数据的字段**

| 字段 | 名称 | 数据类型 | 说明 |
|---|---|---|---|
| type | 类型 | 字符型，选：Document | 必选 |
| title | 题目 | 字符型，选：Standard | 可选 |
| Number | 标准号 | 字符型 | 必选 |
| Name | 标准名称 | 字符型 | 必选 |
| Announcing_body | 发布机构 | 字符型 | 必选 |
| publishing_date | 发布日期 | 字符型:YYYY-MM-DD | 必选 |
| effective_date | 实施日期 | 字符型:YYYY-MM-DD | 可选 |
| ICS_number | 国际标准分类 | 字符型 | 可选 |
| CCS_number | 中国标准分类 | 字符型 | 可选 |
| industry | 国民经济行业分类 | 字符型 | 可选 |
| drafting_body | 起草单位 | 字符型 | 可选 |
| drafting_person | 起草人 | 字符型 | 可选 |
| scope | 范围 | 字符型 | 可选 |
| patent | 是否包含专利信息 | 字符型 | 可选 |

#### 5.2.2 正文

在概念文件正文的基础上，规定标准的封面、目次、前言、引言、范围、规范性引用文件、术语和定义、其他章、参考文献、索引、附录等部分为一级标题，相应其下的子标题为二级标题，最多设置三级标题，其余设置为正文。