

ICS 点击此处添加 ICS 号

CCS 点击此处添加 CCS 号

NY

中华人民共和国农业行业标准

NY/T XXXXX—XXXX

农业本体网络构建技术规范

Technical specification for the construction of agricultural ontology network

(点击此处添加与国际标准一致性程度的标识)

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

中华人民共和国农业农村部 发布



# 目 录

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 构建原则 .....	5
5 农业本体数据标准 .....	5
5.1 农业本体数据类别 .....	5
5.2 农业本体数据格式 .....	6
5.3 农业本体数据关系 .....	7
6 农业本体数据推理 .....	8
6.1 基础关系推理 .....	8
6.2 分级相关性关系推理 .....	8
7 农业本体网络管理及可视化 .....	9
7.1 农业本体网络系统构建 .....	9
7.2 农业本体自动化导入 .....	9
7.3 农业本体网络可视化 .....	9
7.4 农业本体网络查询及编辑节点 .....	9
7.5 农业本体网络查询及编辑关系 .....	10
参 考 文 献 .....	11

## 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。

文件和其他标准、其他分部分标准和国际文件均无关联。

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。

本文件由农业农村部市场与信息化司提出。

本文件由农业农村部数据标准化技术委员会归口。

本文件起草单位：

本文件主要起草人：

（本文件历次版本发布情况）

# 引 言

随着农业科学研究的不断深入与数据量的快速增长，农业本体在数据管理、知识共享及科学研究中的重要性日益凸显。现有的农业科学数据本体虽具备一定的理论与实践基础，但在标准化、互操作性及自动化处理方面仍面临诸多挑战。为提升农业科学数据本体的规范性和实用性，指导农业领域的本体构建与网络化管理，编制本文件。

**编制原因：**农业科学数据涉及领域广泛、概念复杂，为实现跨学科领域数据的共享和应用，需要一套明确、标准化的农业本体构建技术标准。同时，农业科学数据的可视化、分析和推理等应用需求，进一步推动了农业本体网络化管理的必要性。

**编制目的：**本文件旨在为农业科学数据本体网络的构建、管理、推理及可视化提供标准化指导，确保农业本体的准确性、一致性、可扩展性和互操作性；同时规范农业本体网络的构建流程，提高其在科研和生产实践中的应用效率。

本文件根据农业本体网络构建与应用的实际需求，分为以下几个部分：

术语和定义部分界定了农业本体及相关术语的定义，奠定了文件的基础概念框架。

构建原则部分明确了农业本体在设计和开发中的核心指导原则，为后续技术内容提供理论依据。

农业本体标准部分规范了农业本体的分类、格式及关系定义，为农业本体的构建提供技术支撑。

农业本体网络管理及可视化部分提出了基于图数据库技术的网络化管理方法及可视化方案，为农业本体的实际应用提供操作性指导。

农业本体推理部分定义了基于逻辑推理和相关性分析的推理规则，进一步扩展了农业本体的功能和应用范围。

这些部分相辅相成，共同形成了农业本体网络构建及应用的完整技术体系。

本文件的发布将为农业科学数据本体的标准化及农业本体网络的建设提供重要技术支撑。

# 农业本体网络构建技术规范

## 1 范围

本文件界定了农业本体、农业本体网络及相关术语的界限。本文件规定了农业本体的要求、特性及分类标准，并明确了农业本体关系的类型及其逻辑规则。本文件确立了农业本体网络构建的原则、程序及管理体系，包含图数据库搭建、数据导入工具、以及可视化分析的总体原则。本文件描述了农业本体网络的管理及可视化方法，包括节点和关系的查询、编辑、以及推理过程中的路径和规则。

本文件适用于农业科学数据本体网络的构建、管理及分析任务。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**本体** *ontology*

本体是共享概念模型的形式化规范说明。

### 3.2

**农业本体** *agricultural ontology*

农业学科领域中一套得到认同的、关于概念体系的明确、正式的规范说明。

### 3.3

**农业本体关系** *agricultural ontology relations*

关系用于定义实体之间的各种类型的连接方式，这些连接可以是单向的或者双向的，它表明两个实体在逻辑上是如何关联的。

### 3.4

**农业本体网络** *agricultural ontology network*

通过将农业临近领域的多套本体中具有一定关系的术语间进行新关系的构建，从而将多张术语关系图链接为术语网，这张整体性网络即为农业本体网络。

### 3.5

**本体可视化** *ontology visualization*

是在标准化语言构造的层次上显示本体的概念和结构，方便本体开发人员或工程师以及其他希望分析和理解本体的用户将其用于数据注释和其他任务。

## 3.6

**本体推理 ontological reasoning**

利用本体已经蕴含的语义和逻辑，来对实体类型以及实体之间的关系进行推理。

## 3.7

**农业本体网络系统 agricultural ontology network system**

是一个基于农业本体数据构建的知识管理与应用平台，旨在通过对农业领域的本体知识和数据进行语义化建模与关联，形成高效的知识表示和推理体系。该系统以农业本体为核心，通过描述领域概念、属性及其相互关系，实现农业信息的标准化表达和语义互操作。

## 4 构建原则

**准确性：**确保本体中的概念和关系准确反映农业科学的实际状况。

**一致性：**本体内部的概念和关系应该保持逻辑上的一致性。

**可扩展性：**设计时应考虑到未来可能新增的概念或关系，使得本体能够灵活地适应变化。

**互操作性：**确保本体能够与其他系统或本体有效交互。

## 5 农业本体数据标准

## 5.1 农业本体数据类别

根据名称，将农业本体数据分为3类，**生物信息类：**包括基因、蛋白质、细胞等与生物研究信息相关的农业本体；**农业实体类：**包括环境，食品，农业活动等与农业实体相关的本体；**农业性状表型类：**包括农业表型或性状等相关的农业本体。各类应符合表1的规定。

表1 农业本体数据分类标准

英文名称	中文名称	本体格式	类别
OBO Metadata Ontology	OBO 元数据本体(OMO)	OBO	生物信息类
Sequence Ontology	序列本体(SO)	OBO	
Gene Ontology	基因本体(GO)	OBO	
ChEBI Ontology	生化实体本体(CHEBI)	OBO	
Protein Ontology	蛋白质本体(PRO)	OBO	
Cell Ontology	细胞本体(CL)	OBO	
Anatomical Entity Ontology	解剖实体本体(AEO)	OBO	
EMBRACE Data and Methods Ontology	EMBRACE 数据和方法本体 (EDAM)	OBO	
Relation Ontology	关系本体 (RO)	OBO	
Bioinformatics Web Services ontology	生物信息学 Web 服务本体 (OBIWS)	OWL	

Plant Ontology	植物本体(PO)	OBO	农业实体类
Environment Ontology	环境本体(ENVO)	OBO	
Plant Experimental Conditions Ontology	植物实验条件本体(PECO)	OBO	
Agronomy Ontology	农学本体(AGRO)	OWL	
Small-scale fisheries & aquaculture ontology	小规模渔业和水产养殖本体(SSFAO)	OWL	
Plant Trait ontology	特质本体(TO)	OBO	
Food Ontology	食品本体(FOODON)	OWL	
Compositional Dietary Nutrition Ontology	组合膳食营养本体(CDNO)	OBO	
SEONT	SEONT 农业住户调查本体	OBO	
An ontology of core ecological entities	核心生态实体的本体(ecocore)	OBO	
Soil food web ontology	土壤食物网本体(SFWO)	OBO	
Environment Ontology for Livestock	畜牧业环境本体(EOL)	OWL	
Livestock Breed Ontology	牲畜品种本体(LBO)	OBO	
Flora Phenotype Ontology	菌群表型本体(FLOPO)	OWL	
Phenotype And Trait Ontology	表型和性状本体(PATO)	OBO	
Animal Trait Ontology for Livestock	牲畜的动物性状本体(ATOL)	OWL	
Plant Stress Ontology	植物胁迫本体(PSO)	OBO	
Livestock Product Trait Ontology	畜产品性状本体(LPT)	OBO	

## 5.2 农业本体数据格式

农业本体数据格式为19项属性，各农业科学数据本体格式应涵盖表1的规定。

表 2 农业本体数据格式

属性名称	说明
Vertex_ID	术语编号



ontology_id	原始编码
name	术语名称
subset	子集
is_obsolete	是否过时
def	定义
synonym	同义词
xref	外部参照
comment	注释
alt_id	可替代术语（可跨本体）
created_by	创建作者
creation_date	创建日期
consider	可参考术语
replaced_by	被术语替代
namespace	命名空间
namezh	中文名称
defzh	中文定义
property_value	属性值
disjoint_from	完全不重叠术语

### 5.3 农业本体数据关系

基于共现次数的关系筛选方法可以确保各个本体数据遵循相同的格式和标准，从而提高数据的一致性，减少本体的复杂度，实现对数据量有效的控制和管理。经过筛选，一共得到14个关系名称，结果如下：

继承关系（is-a）：表示一种具体概念是另一种概念的子类。

部分整体关系（part-of）：表示一个概念是另一个概念的一部分。例如，“根”是“植物”的一部分。

部分整体逆向关系（has-part）：表示一个概念拥有某个部分。

属性关系（has\_quality）：表示一个概念拥有某种性质或特性。

交集关系（intersection-of）：表示某个概念是两个概念的交集。

时空关系（occurs-in）：表示一个概念在特定时间和空间中发生，表示某个事件、过程或行为发生在某个特定的环境、位置或范围中。

负调节关系（negatively-regulates）：表示一个概念对另一个概念的调节作用是负向的，即减少或抑制某种活动或过程。

约束关系（only-in-taxon）：表示某个概念仅适用于特定的物种或类群。

正调节关（positively-regulates）：表示一个概念对另一个概念的调节作用是正向的，即促进或增强某种活动或过程。

调节关系 (regulates)：表示某个概念对另一个概念的调节作用，通常不指定是正向还是负向调节。

不相交关系 (disjointWith)：表示两个概念在本体中是完全互不相交的，即它们没有任何共享的实例。

继承关系 (subclassOf)：表示一个概念是另一个概念的子类，表示继承、包含或特性传递。

因果关系 (determined-by)：表示某个概念是由其他概念所决定或确定的。

演变关系 (develops\_from)：表示某个概念是从另一个概念演变、发展或衍生出来的。

## 6 农业本体数据推理

农业本体数据推理分为基础关系推理和分级相关性关系推理。基础关系推理根据开放生物学本体组织定义的生物学本体基础关系，除去各本体中自定义关系外，保留四项关系构建新三元组。分级相关性关系推理用于扩展跨本体的术语关系，借鉴临床医学中分级证据的方法，定义分级关系，以美国NCCN（美国国立癌症研究体系）为例，其在发布的权威性临床指南中，采用4级证据分类体系，国内测序企业在提供基因检测报告和报告解读的过程中，也大致依据类似的分级策略。因此，参考这种分级证据体系，同样制定四级相关性体系构建跨本体关系。

### 6.1 基础关系推理

推理上可用的基础关系为is a、part of、has part和regulates四种。调用Jena推理推理则遵循如下表3所示规则，例如，若存在三元组<A is a B>，同时存在<B part of C>，则可以推理出<A part of C>。在IntelliJ IDEA 2024.2.3中编写Java程序，使用Apache Jena提供的API，将上述规则转化为自定义的10条推理脚本，生成新的三元组关系。

表3 农业本体基础关系推理规则

关系	is a	part of	has part	regulates
is a	is a	part of	has part	regulates
part of	part of	part of	—	—
has part	has part	—	has part	—
regulates	regulates	regulates	—	—

### 6.2 分级相关性关系推理

首先推理五种相关性，即定义相关、语义相关、生物网络相关、统计相关和参考物种相关。定义相关，即术语A在定义过程中，直接标注为参考了术语B，或者认为与术语B相关。语义相关，即术语A的语义解释与术语B的语义解释类似；生物网络相关，即术语A在生物网络中的直接下游产物与术语B相关；统计相关，即通过WCGNA或其他分析手段，判断术语A与术语B相关；参考物种相关，即在某物种中已有实验证实术语A与术语B相关。基于五种相关性，制定四级分类体系如下：

I级相关：定义相关。

II级相关：其他四种相关全部具备证据。

III级相关：其他四种相关中具备两或三种证据。

IV级相关：孤证，四种相关中仅具备一种证据。

最后,推理遵循如下表4所示规则,例如,若存在三元组<A I级相关 B>,同时存在<B II级相关 C>,则可以推理出<A II级相关 C>。使用 Apache Jena 提供的 API 或者R 4.4.1编写脚本,生成新的三元组关系。

表 4 农业本体分级相关性关系推理规则

关系	I级相关	II级相关	III级相关	IV级相关
I级相关	I级相关	II级相关	III级相关	IV级相关
II级相关	II级相关	III级相关	III级相关	IV级相关
III级相关	III级相关	III级相关	III级相关	IV级相关
IV级相关	IV级相关	IV级相关	IV级相关	IV级相关

## 7 农业本体网络管理及可视化

### 7.1 农业本体网络系统构建

农业科学数据本体网络系统实现以 HugeGraph 作为管理引擎,环境为 Linux 平台,Ubuntu20.04 的64位版本,以“HugeGraph-Server+Hugegraph-loader+HugeGraph-hubble+Gremlin 查询语言”为技术核心,图数据库后端存储方式 hugegraph.properties 配置为 RocksDB。系统前端基于Vue和Element-ui 前端框架开发操作界面,调用后端提供的API来获取数据,并利用百度开发的ECharts组件关联后端数据库,创建交互性强、可扩展的数据可视化图像,继而通过Axios前后端交互技术与后端数据交互,具有数据驱动、组件化、虚拟 DOM 三大核心特征。后端则采用“Spring+Spring Boot+Spring Cloud+HugeGraph+OAuth2”的技术栈组合向前端提供数据和服务,这个技术栈涵盖了从后端框架、数据库、缓存、身份认证到编程语言多个层次的应用开发需求。

### 7.2 农业本体自动化导入

OntoHugegraphImporter 智能导入工具旨在将农业领域的本体文件数据进行标准化处理,并实现全流程的自动化导入。通过预处理农学领域本体数据,工具能够自动生成规范化的顶点数据文件、边数据文件、图模型文件以及映射文件,并将这些文件上传到服务器上,最终一键上传到HugeGraph 图数据库中,形成农业本体网络。

### 7.3 农业本体网络可视化

农业科学数据本体网络系统前端利用百度开发的ECharts组件创建交互性强、可扩展的数据可视化图像,是基于Web的可视化管理平台,提供了一站式的可视化分析和管理工具,覆盖了从数据建模、数据导入到数据分析和图管理的全流程。在本体可视化子模块中,可直观查看本体网络全貌。

### 7.4 农业本体网络查询及编辑节点

在农业科学数据本体网络系统中可以更便捷的完成查询节点、新增节点和编辑节点的操作,步骤如下:

(1) 在本体可视化子模块中选择列表本体名称,点击列表/图形化切换,输入ontology id或其他属性内容可以检索要查找的节点。

(2) 点击新增实体，在弹出的用户输入框中按内容要求输入属性值，包括节点名称、节点类型、创建时间、alt id 、comment等。

(3) 选中需要更改的节点，跳出实体操作弹框，修改属性值或删除节点。

## 7.5 农业本体网络查询及编辑关系

在农业科学数据本体网络系统中边的查询、新增和编辑按以下步骤操作：

(1) 在本体可视化子模块中点击图形上方提供的关系类型选择框，可查看任意一种关系的所有边和起始、目标节点。

(2) 在新增实体弹出的用户输入框的下方，选择要创建的边类型，最后点击确定即可完成创建。

(3) 在实体操作弹框可修改上下位关系内容，点击右侧减号可以删除对应关系。

参 考 文 献

- [1] GB 3100 国际单位制及其应用
-