

ICS 35.100.70
L78



中华人民共和国国家标准

GB/T XXXX-XXXX

信息技术 数据溯源描述模型

Information technology— Data provenance descriptive model

点击此处添加与国际标准一致性程度的标识

（征求意见稿）

在提交反馈时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言 II

1 范围 1

2 术语和定义 1

3 ProVOC 模型综述..... 1

4 各类构件描述信息 4

 4.1 一级类 4

 4.2 二级类 5

 4.3 三级类 6

前 言

本标准按照GB/T 1.1-2009给出的规则起草。

本标准由全国信息技术标准化技术委员会（SAC/TC28）提出并归口。

本标准起草单位：

本标准主要起草人：

信息技术 数据溯源描述模型

1 范围

本标准定义了一种数据溯源描述模型(以下简称ProVOC模型),规范了ProVOC模型构件的描述信息。
本标准适用于数据采集、发布、分析、处理等系统设计人员以及数据使用者。系统设计人员和数据使用者可根据本标准声明的表达模型规则描述数据溯源信息。

2 术语和定义

下列术语和定义适用于本文件。

2.1

数据溯源 data provenance
数据在整个生存周期内(从产生、传播到消亡)的演变信息和演变处理内容的记录。
数据溯源也称为“数据族系”、“数据系谱”。

2.2

ProVOC 模型 provenance vocabulary model
本标准定义的一种简便数据溯源描述模型。

3 ProVOC 模型综述

ProVOC模型由执行实体、活动和数据三个基本类(以下称为一级类)构件组成。执行实体包括人类执行实体(如,人、组织和公司)和非人类执行实体(如,数据提供之类的服务)。人类执行实体和非人类执行实体是执行实体下的子类构件(以下称为二级类构件)。数据包括参数和数据集两个二级类构件,其中,“参数”包括时间参数、空间参数和条件参数三个子类构件(以下称为三级类构件)。时间参数的例子有:与数据集相关的传感器读取时的时间戳,与活动相关的蛋白质分析过程中的时间持续性,与执行实体相关的传感器正常工作持续时间。空间参数的例子有:海洋上浮标的地理位置。条件参数的例子有:温度、湿度等。二级类构件“数据集”与数据溯源应用领域密切相关,可能按特定应用领域进一步分解出若干三级或更多层级的子类,这些层级的子类及其定义不在本标准范围内。
ProVOC模型的各类构件见表1。

表1 ProVOC 模型各类构件

一级类构件	二级类构件	三级类构件
数据	数据集	-
	参数	时间参数
		空间参数
		条件参数
活动	-	-
执行实体	人类执行实体	-
	非人类执行实体	-

说明：

“活动”、“人类执行实体”、“非人类执行实体”和“数据集”的子类不在本标准中定义，表中用“-”表示。

ProVOC模型各类构件之间存在相应的关联关系，主要包括两大类：从属关系和其他关系（以下称为交互关系）。

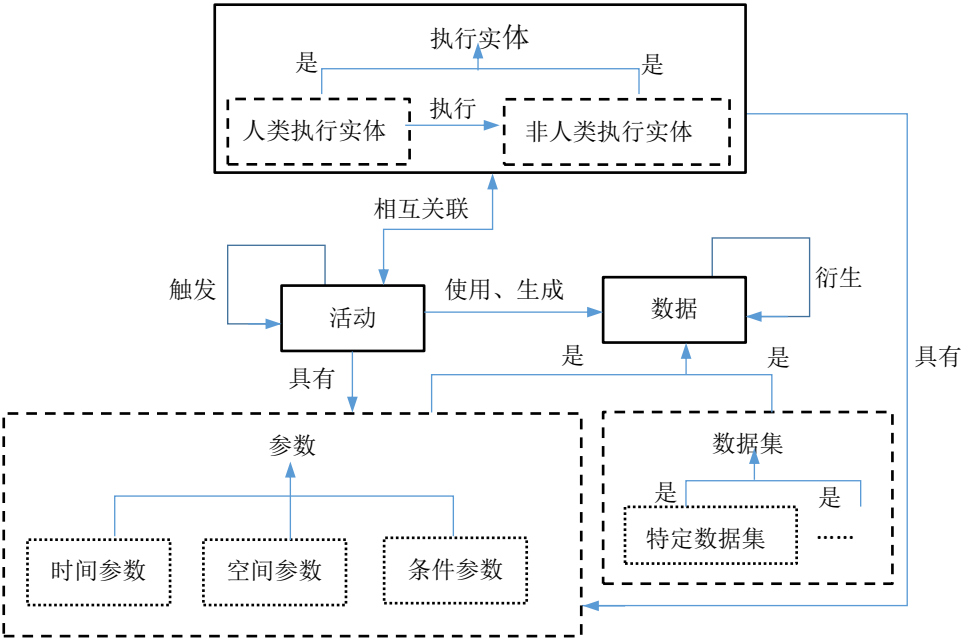
从属关系包括：

- “人类执行实体”和“非人类执行实体”是“执行实体”的子类构件；
- “参数”和“数据集”是“数据”的子类构件；
- “时间参数”、“空间参数”和“条件参数”是“参数”的子类构件。

交互关系包括：

- “执行实体”执行“活动”，它们之间彼此关联；
- “非人类执行实体”本身的活动由“人类执行实体”执行；
- “活动”中使用“数据”；
- “活动”中生成“数据”；
- “活动”中可能触发新的“活动”；
- “数据”可能衍生出新的“数据”；
- “执行实体”具有相应的“参数”，包括“时间参数”、“空间参数”和“条件参数”；
- “活动”具有相应的“参数”，包括“时间参数”、“空间参数”和“条件参数”。

ProVOC模型结构如图1所示。



说明：
图中，实线方框是一级类构件，短划虚线是二级类构件，点划虚线是三级类构件。

图1 ProVOC模型结构

ProVOC模型各类之间的关联关系见表2。

表2 ProVOC 模型构件之间的关系

关系		上位构件		活动		数据			
		人类执行实体	非人类执行实体			参数			数据集
下位构件	执行实体	人类执行实体	非人类执行实体	相互关联		时间参数	空间参数	条件参数	
	非人类执行实体	执行	被执行			用于	用于	用于	-
活动		相互关联		触发		用于	用于	用于	-
						使用于			
数据	参数	时间参数	具有	具有	使用	生成	衍生		
		空间参数	具有	具有					
		条件参数	具有	具有					
	数据集	-	-	-					

说明：

1) 为表达各类构件之间的关系，表2中使用了“上位构件”和“下位构件”两个术语。“上位构件”是指存在关联关系的两类构件中作为主体的构件（即主宾结构句子中的主语），“下位构件”是指两类构件中作为客体的构件（即主宾结构句子中的宾语）。表2中构件间关系有两种描述方式，例如，在描述“人类执行实体”与“非人类执行实体”的关联关系时，一种是：“（上位构件中的）人类执行实体执行（下位构件中的）非人类执行实体”，另一种是：“（上位构件中的）非人类执行实体被（下位构件中的）人类执行实体执行”；又例如，在描述“活动”与“时间参数”、“空间参数”和“条件参数”的关联关系时，有两种模式方式，一种是：“（上位构件中的）活动具有（下位构件中的）时间参数、空间参数和条件参数”，另一种是：“（上位构件中的）时间参数、空间参数和条件参数用于（下位构件中的）活动”。

- 2) “- ”表示构件之间关系不在本标准中定义。
- 3) 空白格表示该关系不适用。

4 各类构件描述信息

4.1 一级类

4.1.1 数据

构件“数据”的描述信息见表3。

表3 数据的描述信息

信息项	说明
名称	数据
标识	Data
子类	数据集
	参数
关联关系	使用于活动
	生成于活动
	衍生另一个数据

4.1.2 活动

构件“活动”的描述信息见表4。

表4 活动的描述信息

信息项	说明
名称	活动
标识	Activity
子类	在具体数据溯源应用中定义
关联关系	与执行实体相互关联
	触发另一个活动
	具有时间参数、空间参数和条件参数
	使用数据
	生成数据

4.1.3 执行实体

构件“执行实体”的描述信息见表5。

表5 执行实体的描述信息

信息项	说明
名称	执行实体
标识	Agent
子类	人类执行实体
	非人类执行实体
关联关系	与活动相互关联
	具有时间参数、空间参数和条件参数

4.2 二级类

4.2.1 数据集

构件“数据集”的描述信息见表6。

表6 数据集的描述信息

信息项	说明
名称	数据集
标识	DataCollection
父类	数据
子类	在具体数据溯源应用中定义
关联关系	使用于活动
	生成于活动
	衍生另一个数据

4.2.2 参数

构件“参数”的描述信息见表7。

表7 参数的描述信息

信息项	说明
名称	参数
标识	Parameter
父类	数据
子类	时间参数
	空间参数
	条件参数
关联关系	用于执行实体
	用于活动
	使用于活动
	生成于活动
	衍生另一个数据

4.2.3 人类执行实体

构件“人类执行实体”的描述信息见表8。

表8 人类执行实体的描述信息

信息项	说明
名称	人类执行实体
标识	HumanAgent
父类	执行实体
子类	在具体数据溯源应用中定义
关联关系	执行非人类执行实体

4.2.4 非人类执行实体

构件“非人类执行实体”的描述信息见表9。

表9 非人类执行实体的描述信息

信息项	说明
名称	非人类执行实体
标识	NonHumanAgent
父类	执行实体
子类	在具体数据溯源应用中定义
关联关系	由人类执行实体执行

4.3 三级类

4.3.1 时间参数

构件“时间参数”的描述信息见表10。

表10 时间参数的描述信息

信息项	说明
名称	时间参数
标识	TemporalParameter
父类	参数
子类	时间段
	时间点
关联关系	用于执行实体
	用于活动

4.3.2 空间参数

构件“空间参数”的描述信息见表11。

表11 空间参数的描述信息

信息项	说明
名称	空间参数
标识	SpatialParameter
父类	参数
子类	在具体数据溯源应用中定义
关联关系	用于执行实体
	用于活动

4.3.3 条件参数

构件“条件参数”的描述信息见表12。

表12 条件参数的描述信息

信息项	说明
名称	条件参数
标识	ConditionParameter
父类	参数
子类	在具体数据溯源应用中定义
关联关系	用于执行实体
	用于活动