



# 中华人民共和国国家标准

GB/T 34945—2017

---

## 信息技术 数据溯源描述模型

Information technology—Data provenance descriptive model

2017-11-01 发布

2018-05-01 实施

---

中华人民共和国国家质量监督检验检疫总局 发布  
中国国家标准化管理委员会

## 目 次

前言 .....	I
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 .....	3
4.1 一级类 .....	3
4.2 二级类 .....	4
4.3 三级类 .....	6
5 实例 .....	7

## 前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

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这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位:中国科学院计算机网络信息中心、中国电子技术标准化研究院、中科院地理科学与资源研究所、中国科学技术信息研究所、上海计算机软件技术开发中心、北京科技大学、电子科技大学、北京软件和信息服务交易所有限公司、浪潮软件集团有限公司、智云天地农业信息技术(北京)有限公司、无锡中科恒源信息科技有限公司、银联智惠信息服务(上海)有限公司、四川长虹佳华数字技术有限公司和北龙泽达(北京)数据科技有限公司。

本标准主要起草人:胡良霖、黎建辉、陈希、王静、李娜、郭晓峰、赵辉、尹海清、周园春、李冰、于铁强、蔡立志、龙凯、夏琦、黄先芝、崔运鹏、陈曙东、周力、姜雪、高瑞鑫、王建华。

# 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, provenance vocabulary model), 规范了 ProVOC 模型构件的描述信息。

本标准适用于数据的采集、发布、分析和处理等系统的设计人员以及数据使用者。

## 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**

数据族系

数据系谱

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。

## 3 ProVOC 模型综述

ProVOC 模型由数据、活动和执行实体三个基本类(以下称为一级类)构件组成。数据包括参数和数据集两个子类构件(以下称为二级类构件), 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)。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(特定数据集)。活动是由执行实体发起或受执行实体控制、影响的一个或一系列动作。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二级类构件。

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	—
注：“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: 从属关系和交互关系。

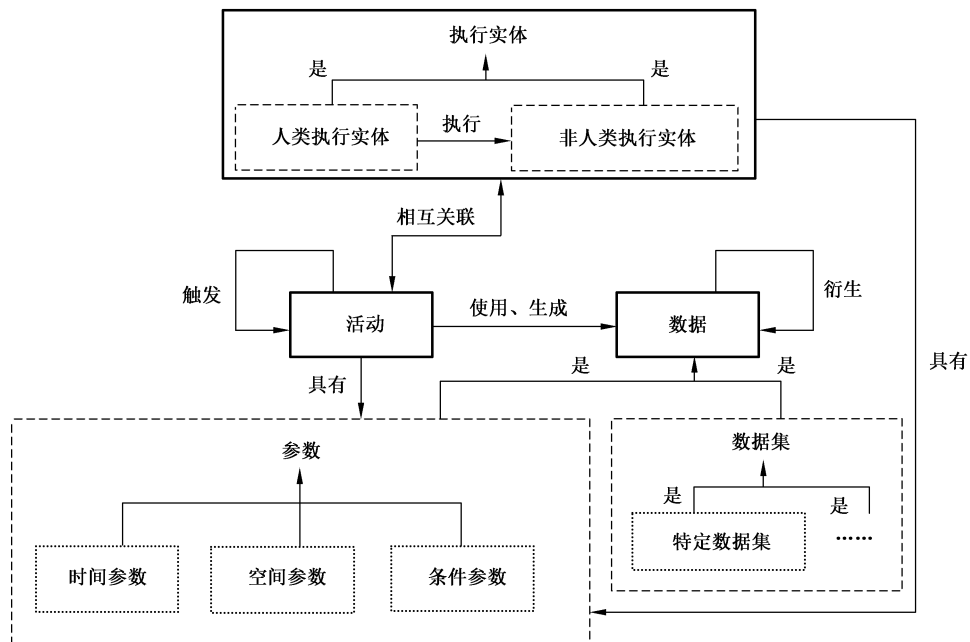
从属关系包括:

- a) “人类执行实体”和“非人类执行实体”是“执行实体”的子类构件；
- b) “参数”和“数据集”是“数据”的子类构件；
- c) “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：

- a) “执行实体”执行“活动”，它们之间彼此关联；
- b) “非人类执行实体”本身的活动由“人类执行实体”执行；
- c) “活动”中使用“数据”；
- d) “活动”中生成“数据”；
- e) “活动”中可能触发新的“活动”；
- f) “数据”可能衍生出新的“数据”；
- g) “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”；
- h) “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。



注：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

下位构件		上位构件						
		执行实体		活动	数据			
		人类执行实体	非人类执行实体		参数			数据集
					时间参数	空间参数	条件参数	
执行实体	人类执行实体		被执行	相互关联	使用于	使用于	使用于	—
	非人类执行实体	执行						—

表 2 (续)

下位构件		上位构件								
		执行实体		活动	数据					
		人类执行实体	非人类执行实体		参数			数据集		
时间参数	空间参数			条件参数	—					
活动		相互关联		触发		使用于	使用于	使用于	—	
						使用于				
						生成于				
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	—	—							

注 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”。

注 2: “—”表示构件之间关系不在本标准中定义。

注 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

表 6 (续)

信息项	说明
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	使用于执行实体
	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	使用于执行实体
	使用于活动

#### 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	使用于执行实体
	使用于活动

#### 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	使用于执行实体
	使用于活动

## 5 实例

Alice 应用数据发布软件发布一条数据 d2, d2 的生成过程中有一条由数据 d1 输入到该过程中, d1 是由传感器 Sensor1 在 gps 位置信息为(经度 geod2, 纬度 geo1, 高度 geo3)采集得到。

说明:

- 本标准中定义的数据溯源描述模型是一个抽象概念模型, 为便于理解, 举出此实例, 并将模型中描述的关联关系具体为: 生成-output; 使用-input; 相互关联-wasPerformedBy; 具有空间参数-locatedIn。
- 本示例采用 rdf 进行描述。

示例:

```
<>rdf:type ProVOC:Activity;
  rdfs:comment:Alice 应用数据发布软件发布一条数据 d2, 其中输入一条数据 d1
  ProVOC:output _:d2;
  ProVOC:input _:d1;
  ProVOC:wasPerformedBy<http://example.cn/Alice>
  ProVOC:wasPerformedBy<http://example.cn/PublishingSoftware1>.
```

```
<>rdf:type ProVOC:Activity;
  rdfs:comment:Sensor1 采集到的数据 d1
  ProVOC:wasPerformedBy<http://example.cn/Sensor1>
  ProVOC:output _:d1;
  ProVOC:locatedIn_:geod1
  ProVOC:locatedIn_:geod2
  ProVOC:locatedIn_:geod3
  _:geod1 rdf:type rdfs:lat
  _:geod2 rdf:type rdfs:long
  _:geod3 rdf:type rdfs:alt
```