



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
中国国家标准化管理委员会 发布

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业交易所有限公司、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、银联智慧信息服务(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 and 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: