

# 基于 E-R-P 建模体系的政务资源元数据表示模型

许永涛, 王延章, 陈雪龙

(大连理工大学管理学院信息与决策技术研究所, 大连 116024)

**摘要:** 针对电子政务中信息资源元数据的表示问题, 在研究实体—关系—问题 (E-R-P) 建模体系的基础上, 以数据仓库数据模型的星型模式为例, 给出了基于 E-R-P 建模体系的政务信息资源元数据表示模型及相应的系统概念模型。实践证明, 该模型能较好地实现政府信息资源的高效发现和共享。

**关键词:** 政务信息资源; 元数据; 表示模型

## Metadata Representation Model for E-government Resources Based on E-R-P Modeling System

XU Yongtao, WANG Yanzhang, CHEN Xuelong

(Institute of Information and Decision Technology, College of Management, Dalian Univ. of Technology, Dalian 116024)

**【Abstract】** To solve the problem of metadata representation for E-government information resources, this paper proposes entity-relation-problem (E-R-P) modeling system for metadata modeling and offers metadata representation model based on E-R-P modeling system. It takes metadata expression of star schema model of data warehouse as example. Corresponding system conceptual model is also proposed. And the practice proves the improvement of the model.

**【Key words】** E-government information resources; Metadata; Representation model

### 1 概述

由于各级政府基础信息的“三自”现象(自采、自用、自成体系), 政府各部门之间、政府与公众之间的各种系统难以兼容, 政府各部门的信息化建设条块分割, 形成“信息孤岛”, 给政府的信息共享带来巨大的障碍。主要体现在: 不同的信息系统使用多种术语表示同一概念; 同一术语在不同的信息系统中表达不同的含义; 各信息系统使用不同结构表示相同或相似的信息; 各信息系统中的概念间存在着各种联系, 但各种信息源的分布自治性导致这种隐含的联系无法体现。

元数据是关于数据的数据, 是面向特定应用的、用于描述信息资源属性的、机器可理解的信息, 它通过规范语法结构和语义结构, 使机器能够无二义性地表现和获取信息<sup>[1]</sup>。元数据表示模型的建立是解决信息系统之间异构问题, 实现政府信息共享的重要环节。目前, 国际上比较流行的元数据标准有CDWA、DC、EAD、GILS、TEI等<sup>[2-4]</sup>, 在实践中应用于政府信息资源描述、发现、管理的元数据标准主要有2种: GILS<sup>[5]</sup>和DC-government<sup>[6]</sup>。由于政府应用需求的多样化和开发主体的多元化, 存在着多种元数据模型, 这就使得各种元数据模型之间难以实现互操作, 给元数据的管理带来困难。目前, 很多文献对元数据进行了研究<sup>[7-9]</sup>, 但对元数据的表示研究并不多。元数据的表示(编码)通常采用XML方式, 但XML方式存在一个缺陷, 就是建模原语不够丰富, 很难表达逻辑关系的复杂概念及元数据的语义。为此, 本文针对电子政务中信息资源元数据的表示问题, 通过对实体-关系-问题(E-R-P)的建模体系的研究, 以数据仓库各种逻辑数据模型中被广泛接受的星型模式为例, 给出了基于E-R-P建模体系的政务信息资源元数据的表示模型及相应的系统概念模型。基于E-R-P建模体系的元数据表示模型可以有效描述政务信息

资源, 将元数据中的术语含义、术语间的逻辑关系更加明确地表达出来, 给元数据管理带来方便, 并较大地实现了政务信息资源的高效发现和共享, 具有很强的通用性。

### 2 E-R-P 建模体系

#### 2.1 E-R-P 建模体系概述

按哲学观点, 一个客观事物总是可分的, 可以分成单元、元素或下一个层次的事物, 这些单元、元素、下一层次事物(简称为子事物)是一种客观实在并构成客观事物的内在因素活动, 称这些客观实在为实体(Entity)。

依据哲学中“事物是普遍联系的”观点, 构成一个客观事物的所有实体都是相互联系的。这种联系体现在实体之间的相互依赖、相互作用的关系(Relation)之中。不难理解, 任何客观事物的运动或某一领域的知识都可以用实体和相应关系加以描述。这样, 一个客观事物系统S就可以用相应的实体集E和关系集R来描述。

问题是人的主观愿望或要求与客观事物系统状态的差异, 因此, 问题P(Problem)是关于人们主观目标(目标的概念来源于客观事物系统)和客观事物系统的集合。

论域(Domain)是人们研究一个客观事物系统S时, 所涉及到的研究范围。如果研究的是某个系统的内部问题, 那么论域可以是这个系统; 如果研究的是几个系统间的问题, 那么论域就是这几个系统的集合。

由此可分别从实体、关系、问题和论域的角度对客观事物系统S进行表示, 从而形成描述客观事物系统的E-R-P建

**基金项目:** 国家自然科学基金资助项目(70271045, 60074038)

**作者简介:** 许永涛(1978-), 男, 博士生, 主研方向: 电子政务, 数据仓库; 王延章, 教授、博导; 陈雪龙, 博士生

**收稿日期:** 2006-07-03 **E-mail:** xuyongtao1978@163.com

模体系。系统 S 可以用模型  $S = \int_p (E, R, P)$  表示, 其中, D 表示系统的论域 (或作用空间); E 表示实体; R 表示关系; P 表示问题;  $\int$  表示系统的建模。

为此, 本文对政府信息资源这个论域的元数据描述分别从实体 E、关系 R 和问题 P 的表示着手, 形成了能够描述政府信息资源元数据的 E-R-P 模型体系。

### 2.2 E-R-P 建模体系中的数学运算

E-R-P 建模体系中蕴含着丰富的数学运算, 如实体集中的并、交、差、补差、对移差、析取等运算, 关系集中的关系到实体的转换、析取、集结、并合、连接等运算, 问题系统中的子问题划分、可析判定等运算。例如, 实体集中并的数学运算表示如下: 给定论域 E, 设它的 3 个实体子集分别为  $E_1$ 、 $E_2$  和  $E_3$ , 相应的属性集合为  $A_1$ 、 $A_2$  和  $A_3$ , 那么实体集  $E_1$ 、 $E_2$  的并运算的结果集  $E_3$  定义为  $E_3 = E_1 \cup E_2$ ,  $E_3 \triangleq \{e | e \in E_1 \text{ 或 } e \in E_2\}$ , 相应属性  $A_3 = A_1 \cap A_2$ 。

### 2.3 E-R-P 建模体系的主要特点

根据理论分析和实际应用, E-R-P 建模体系有如下特点:

(1) 是一种通用化的表示方法, E-R-P 表示体系的形成源于一般的哲学思想方法而非某一领域或具体问题的处理, 因此它不依赖于某一领域, 具有很强的通用性, 适用于数据量大, 涉及面广的政务信息资源元数据表示。

(2) 能很好地与一般化模型方法和算子算法相衔接, 适用于定量和定性情形。

(3) 是一种模块化体系, 与它相应的实体、关系和问题的表示都可拆分和可组合, 因此, 在应用上具有很大的灵活性和可扩展性。

(4) 具备通用性, 可构造相应的通用化应用平台、应用工具、数据库和数据仓库系统的基础, 促进了系统和政府应用的一体化发展。

(5) 是关于客观事物系统和相应问题描述的自然表示, 同时由于它的通用性强, 可以用于开各发层次的形式化的转换工具。对政府信息资源元数据的形式化表示, 具有很强的针对性和实用性。

### 3 基于 E-R-P 建模体系的政务资源元数据表示模型

在当前技术条件下, 可应用数据仓库的思想和有关技术对政务信息资源元数据进行表示。本文以大型电子政务工程——辽宁省党政信息网综合政务数据仓库的实际开发为例给出基于 E-R-P 建模体系的政务信息资源元数据的表示模型, 见图 1。

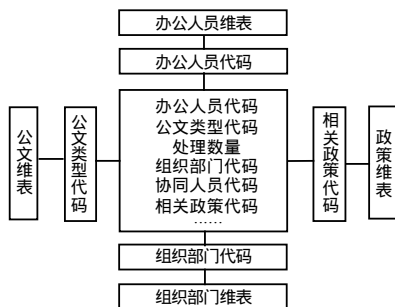


图 1 公文处理分析星型图逻辑模型

在数据仓库的各种逻辑数据模型中, 星型模型被广泛接受, 而且星型模型适用于产生大量报表的场合, 可以对用户的要求做出快速反应。星型模型的逻辑实体包括指标、维度和详细类别实体。指标实体位于星型图的中心, 是用户查询的中心。与指标实体相比, 维度实体是一个小得多的实体。维度和与它们相关的数据可以使数据仓库的用户轻松地浏览指标数据。为了进一步减少每次查询返回的数据行数或对维度实体细化, 引入了详细类别实体为用户提供更多的定性信

息, 以对其进行决策支持。这 3 种实体及它们之间的关系构成数据仓库星型逻辑模型。因为 E-R-P 建模体系强调实体间的关系, 所以基于 E-R-P 建模体系的政府信息资源元数据表示可以充分描述为星型图, 用 E-R-P 建模体系描述该逻辑模型是简洁的, 而且具有通用性。为此, 本文以图 1 政府信息资源中公文处理分析问题星型模型为例, 基于 E-R-P 建模体系, 从实体、关系、问题出发, 描述政府信息资源的元数据表示。

### 3.1 实体描述

作为星型模型的基本构成, 每个实体 (包括指标实体、维度实体、详细类别实体) 都对应一个属性集。为描述实体, 根据 E-R-P 建模体系的实体的定义和数学描述, 定义下列元数据表示:

```

实体 (Entity) {
  header {版本号, 实体生成时间}
  body {标识, 名字, 描述, 属性集,
    关联实体 {关联实体标识, 关联关系实体, ...}
  }
}
属性 (Attribute) {
  header {版本号, 属性生成时间}
  body {标识, 名字, 描述,
    属性集 {属性标识, 名字, 描述, 属性类型, 属性长度, 是否可空, 关键字, ...}
  }
}

```

基于这样的描述, 如图 1 所示的政府公文处理分析星型模型中, 例如, 指标实体可以描述如下:

```

BEGIN ENTITY
BEGIN HEADER
  Version "1.0"
  Date 2005/9/2
END HEADER
BEGIN BODY
  Identifier = "0100503"
  Name = "office"
  Description = "某办公人员公文处理情况事实表"
  Attribute-Set-ID = "0000503"
BEGIN REL-ENTITY //关联实体集
BEGIN REL-ENTITY1
  Rel-Entity = "0100504"
  Relation = "0110503"
END REL-ENTITY1

```

```

...
END REL-ENTITY
END BODY
END ENTITY

```

#### 属性集描述:

```

BEGIN ATTRIBUTE SETS
BEGIN HEADER
  Version "1.0"
  Date 2005/9/2
END HEADER
BEGIN BODY
  Identifier = "0000503"
  Name = "officer-attributes"
  Description = "办公人员维表属性集"
BEGIN ATTRIBUTE1 //实体属性集
  Identifier = "0010503"
  Name = "officer-key"
  Description = "办公人员代码"
  DataType = "CHAR"

```

```

DataLength = 10
Null = FALSE
Key = "foreign-key"
END ATTRIBUTE1
...

```

```

END BODY
END ATTRIBUTE SETS

```

由于指标实体、维度实体、和详细类别实体在 E-R-P 建模体系中都表示为具有一定属性的实体，因此元数据描述完全一致，具有很强的通用性，这种统一的元数据表示方式有利于政府信息资源元数据的实现和管理。

### 3.2 关系描述

前面讨论了单个实体的元数据描述，下面讨论星型模型中实体间关系的元数据表示。根据 E-R-P 建模体系中关系的定义和数学描述，可以将关系的元数据表示如下：

```

关系 (Relation) {
header {版本号, 关系生成时间}
body {关系属性集 (标识, 名字, 描述,...), 施关联实体, 被关联实体
关联矩阵 {行号, 列号,...}
}}

```

基于这样的描述，如图 1 所示的政府公文处理分析星型模型中，指标实体与公文处理维度实体关联关系元数据表示如下：

```

BEGIN RELATION
BEGIN HEADER
Version "1.0"
Date 2005/9/2
END HEADER
BEGIN BODY
Identifier = "0110503"
Name = "officer-document"
Description = "办公人员和公文关联关系"
Rel_OEntity = "0100503" //被关联实体
Rel_IEntity = "0100504" //施关联实体
Rel_Matrix = "1000503"
Attribute-Set-ID = "0000504" //关系属性集
BEGIN MATRIX //关联矩阵
BEGIN ELEMENT1
ROW = 1
COLUMN = 1
END ELEMENT1
...
END MATRIX
END BODY
END RELATION

```

### 3.3 问题描述

问题 P 是 E-R-P 建模体系的核心，是关系集成、是综合组织运用各种方法和工具的基础，是由目标实体集等要素组成的，往往反映了系统的全部特性。研究或抽取系统可能所要解决的问题，并描述问题，构造问题集合，建立问题库。问题可以分为宏观、中观和微观，可以由问题来引出实体和关系，从而完成对信息资源元数据的建模。

根据 E-R-P 建模体系中问题的定义和数学描述，可以将问题实体的元数据表示如下：

```

问题 (Problem) {
header {版本号, 问题生成时间}

```

```

body {问题属性集 (问题标识, 名称, 描述,...), 目标实体集,
条件实体集, 关联实体集, 问题关系集}
}

```

基于这样的描述，图 1 所示的公文处理分析星型模型中，对政府信息资源公文处理分析问题实体进行元数据表示可作如下描述：

```

BEGIN PROBLEM
BEGIN HEADER
Version "1.0"
Date 2005/9/2
END HEADER
BEGIN BODY
BEGIN ATTRIBUTE //问题属性集
Identifier = "0503"
Name = "sale-database"
Description = "面向公文处理分析问题的数据仓库模式"
END ATTRIBUTE
BEGIN TARGET ENTITY //目标实体集
Target-Entity1 = "100503"
...
END TARGET ENTITY
BEGIN CONDITION ENTITY //条件实体集
Condition-Entity1 = "100509"
...
END CONDITION ENTITY
BEGIN RELATION ENTITY //关联实体集
Relation1 = "110504"
...
END RELATION ENTITY
BEGIN PROBLEM RELATION //问题关系集
Relation1 = "0511"
...
END PROBLEM RELATION
END BODY
END PROBLEM

```

## 4 计算机管理系统结构

在这个体系中，各种信息资源元数据的管理通过实体管理、关系管理、问题管理组成，如图 2 所示。

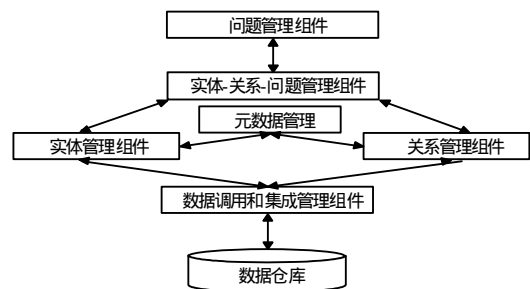


图 2 基于 E-R-P 建模体系的信息资源元数据管理

信息资源元数据通过“实体”及其属性描述，一个信息资源的元数据作为一个客观事物可以分解成许多实体，这些实体可以按其相互关系组成一个倒立的多叉树，树结构中每个节点代表一个实体。对实体的管理包括实体的输入、修改、删除、查询、运算等。同样，对实体的属性也可以进行如上的处理，关系的管理基本和实体类似。问题是决策和问题求解的出发点和归宿，由目标实体集等要素组成，往往反映了系统的全部特性。(下转第 97 页)