

开发信息产品图

Ballou, Donald

(美国纽约州立大学奥尔巴尼分校 纽约奥尔巴尼 12222)

苏颖 [校译]

(中国科学技术信息研究所, 北京 100038)

摘要:任何组织如果想要采取信息产品(IP)方法,不仅需要一套支持理论,而且需要模型、工具和技术。尽管已经有很多文章探讨了组织如何培养数据质量意识以及识别根本原因的方法,但这些方法只是信息产品管理员和数据质量分析师工具箱中的一部分,只有与信息产品图的构想相结合才能真正为信息质量项目提供支持。文章提出信息产品图,然而开发信息产品图的标准方法正在修订,信息产品图的构成还在不断完善。一个由来自学术界和数据质量专家组成的特别委员会(在本文中称之为标准组织)正在制定绘制和维护信息产品图的标准。

关键词:信息产品图;数据质量;信息质量;信息产品

中图分类号: G203 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2009.02.004

1 基本定义、概念与符号

信息产品是指符合数据消费者特定需求的数据元素实例的集合。数据元素是命名数据的最小单位。

信息产品的这一定义符合以前研究中使用的信息产品的直观意义或隐含意义,是可被信息产品图领域所有从业者使用的一般定义。

信息产品图委员会作为一般组织提出了一套基本的符号和构件,并以此来开发信息产品图。这个委员会是基于文献[1]、文献[2]和文献[3]的研究来开展工作的。本文将对这些研究进行阐述。

信息产品图就是对信息产品制造或创建流程的系统化描述,它使用了一套标准的符号和规则。该标准组织把信息产品分为4种类型:

(1) **标准型:**这类信息产品拥有事先指定的格式,按计划周期生成,或在必要的情况下随时生成。如典型的业务报告、账单及工资单。

(2) **特设型:**这类信息产品为格式灵活的实体,没有预先确定的格式,在需要的时候可随时创建和订制。

(3) **存储型:**这类信息产品由具有物理意义的记录、文档及数据库集构成。

(4) **自由格式型:**此类信息产品包括有意义的数据集,其格式并没有预先严格限定。它们包括来自视频、音频等多媒体数据,也包括来自书籍和期刊等印刷媒体数据。

提高不同类型信息产品的质量,需要采取不同的策略。尽管类型不同,但是它们都是信息产品,可以采用本文提出的方法和绘图规则。

尽管目前有一套标准符号和规则还在不断完善之中,但仍可用来描述信息产品的创建。表1列出了用于绘制信息产品图的基本符号。每个表格都给出了信息产品图的结构定义及其符号,并列举了实例。随着信息产品图的应用不断增加,以及获取的新知识不断增多,我们将对这套符号进行改进。

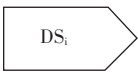








信息产品图和数据流程图有一些类似的地方,

作者简介:Ballou, Donald(1943-),男,奥尔巴尼大学商学院名誉教授,研究方向是各个领域的建模,包括非线性流体动力学、生产质量控制、区域与交通规划、信息资源管理,特别是数据和信息的质量。

基金项目:国家自然科学基金资助项目(70772021,70831003)。

收稿日期:2009年3月18日。

表 1 信息产品图指南

信息产品	原始数据
<p>定义 :通过人工、机器或电子方式生成,被数据消费者利用的最终数据集。例如 :出生/死亡证明、医院账单、学生成绩单、眼镜处方、月度银行对账单、信用报告、邮寄标签。符号/缩写 :</p> <p style="text-align: center;">$\xrightarrow{IP_i}$</p>	<p>定义 :预先定义的一套数据单元,在预先定义的流程中充当原材料,最终将生产出信息产品,是来自信息产品图边界以外的数据或信息。例如 :单一的数字、档案、文件、电子表格、报告、图像、动词短语。符号/缩写 :</p> <p style="text-align: center;">$\xrightarrow{RD_i}$</p>
<p>成分数据</p> <p>定义 :生产信息产品必须的一组临时的半加工信息。可能在信息产品图中生成,被用于创建最终信息产品的数据。例如 :文件摘录、中间报告、半加工数据集。符号/缩写 :</p> <p style="text-align: center;">$\xrightarrow{CD_i}$</p>	<p>源(原始输入数据)块</p> <p>用途 :用来表示每个原始输入数据项的来源。为了生产消费者期望的信息产品,这些数据项必须可以利用。这个块的其他名称还有数据源、数据商、起始点。例如 :患者、消费者及校友的信息。符号/缩写 :</p> <p style="text-align: right;"></p>
<p>消费者(输出)块</p> <p>用途 :用来表示信息产品的消费者。这个块中的消费者详细说明构成最终信息产品的数据元素。在这个信息产品图模型中,消费者事先定义了这个信息产品。这个块的其他名称还有数据宿、消费者块、目的地。例如 :经理、客户、病人、保险公司、邮政局。符号/缩写 :</p> <p style="text-align: right;"></p>	<p>数据质量块</p> <p>用途 :用来表示对那些关键数据项的数据质量的检查,这对生产无缺陷信息产品来说至关重要。与这个模块相关的是对特定成分数据项进行的一组数据质量检查。这个块有两种可能的输出,即“正确”流(概率 p)和“错误”流(概率 1-p)。所输入的数据为原始输入数据项和可能的一些成分数据项。这个块的其他名称还有质量块、评估块、检查块。例如 :人工检查、电子协调、错误检查、匹配、确认。符号/缩写 :</p> <p style="text-align: right;"></p>
<p>处理模块</p> <p>主要用途 :用来表示与生产信息产品所必需的一些或所有原始输入数据项或成分数据项有关的操作、计算或组合。 次要用途 :用作数据修正块。当在一组进入数据质量块的数据元素中发现错误时,需要采取某种修正行为。这个模块代表一个不在标准处理顺序之列,但可在特定环境下应用的处理过程。任何通过这个修正块的原始输入数据项或成分数据项可以被认为是净化的,无须回到数据质量块便能够被后续模块使用。这个块的其他名称还有处理模块、数据修正块。例如 :资料更新、校订、数据捕捉、上传、下载、报告生成、文件生成。符号/缩写 :</p> <p style="text-align: right;"></p>	<p>决策块</p> <p>用途 :在一些依赖特定数据项价值的复杂信息制造系统中,为了进一步的加工处理,可能需要把一些数据项导入一组不同的下游块。在这种情形下,可利用决策块获取被评估的不同条件,以及依据这些评估来处理输入数据项的相应程序。例如 :相同的一组与出生有关的数据项可能被用来开出生证、撰写人口统计摘要报告,或撰写附在检查新生儿先天疾病血样上的报告。符号/缩写 :</p> <p style="text-align: right;"></p>
<p>数据存储块</p> <p>用途 :用来表示在存储文件或数据库中获取数据项,以便对它们进行进一步的加工。存储块可被用来代表那些等待进一步加工,或取作信息目录一部分的数据项(原始数据项或成分数据项)。这个块的其他名称还有数据块、信息存储块。例如 :数据库、文件系统。符号/缩写 :</p> <p style="text-align: right;"></p>	<p>信息系统边界块</p> <p>用途 :用来反映原始输入数据项或成分数据项在从一种信息系统类型向另一种信息系统类型移植时的变化。这些系统变化可能是在业务单位内部或业务单位之间。例如 :从纸质到电子形式的数据输入。符号/缩写 :</p> <p style="text-align: right;"></p>
<p>业务边界块</p> <p>用途 :用来表示一个组织向另一组织移交原始输入数据项或成分数据项的情况。用来说明跨部门或组织边界的信息产品(或原始输入数据项或成分数据项)的移动。这个块的其他名称还有组织边界块。例如 :从手术室到病房的数据转移。符号/缩写 :</p> <p style="text-align: right;"></p>	<p>信息系统-业务边界块</p> <p>用途 :定义原始输入数据项或成分数据项经过业务边界和系统边界变化的情形。这个块的其他名称还有信息系统-组织边界块,信息系统-组织结合块。例如 :从病人护理部门的 Windows NT 到管理部门的 IBM RS/6000 集群的数据转移;从住院处的 RDC Ultrix 到管理部门的 IBM RS/6000 集群的转移。符号/缩写 :</p> <p style="text-align: right;"></p>

都是处理数据流程和数据转换的图表,因此图表技术存在类似之处。不过,信息产品图并不仅仅是一个数据流程图。与数据流程图或实体关系图相比,信息产品图包含了更多的信息,有信息收集者、管理者和消费者的信息,还有在信息生产加工过程中利益相关方的参与度及角色的特定信息。系统基础设施、组织基础设施以及特定职能和责任应该说明清楚。信息产品图的一个重要方面是它们加入了数据质量维度信息。Ballou等人已将数据质量的维度及时加入到信息产品图的一些模块中^[1]。

2 创建信息产品图的步骤

创建信息产品图的程序是基于标准组织的最初提议。该组织的提议已经作为鉴别、绘图、优选以及改进组织信息产品质量的方法。在本文中,我们把创建信息产品图的步骤归纳如下:

步骤1 选择要绘制的信息产品。选择构成信息产品基本构件的数据元素。依据不同的情况,采用不同的方法完成。一种情况是通过检查和分解信息产品来获取数据元素。另一种情况可能无法清楚地鉴别或描述一个明确的信息产品,这时就可以运用自下而上的方法,选择对该信息产品看似关键的一个或多个数据元素。之后,在信息产品图的绘制过程中,对这组数据元素进行筛选(增加或删除元素)。

步骤2 确定数据收集者、数据管理者和数据使用者。确定谁创建、收集和输入数据,谁负责维护数据以及谁将使用这些数据是非常重要的。

步骤3 通过掌握数据元素的流向及其转化,以及数据元素流程之间和内部的连接,描述信息产品。

步骤4 确定职能角色。确定相关系统。确定所涉及的人员及其责任。

这些信息现在可以被合并到信息产品图中。一般来说,这个顺序将会是:描绘客观流程/工作流程、描绘数据流程、描绘系统基础设施、描绘组织基础设施和角色。

此外,信息产品集目录能够加强对信息的管理^[4]。下面讨论信息产品图的符号规则,并举例说明。

3 信息产品图

本节我们将描述用于医院重要信息产品IP图的产生过程。该案例来源于以前的研究,详见文献[2]。我们仅限于讨论一所大医院的一小部分操作和流程,仅包括住院病人的入院、治疗和出院部分。

其中有5个产品使用了从两个重要来源收集的信息:病人和医院工作人员(医生、护士、化验室技师、放射科医生、临床医生以及管理人员)。他们直接或间接参与了病人入院、治疗或出院的过程。每个人都利用了这个大信息集的子集。这5个产品分别是:第一个产品(IP₁)是每天、每周和每月提交给医院管理层的住院报告,报告记述了住院病人的数量,预计住院的天数以及病人的信息,它是一种监测工具,可以反映这些单位繁忙的程度。第二个产品(IP₂)是每天撰写并附在病人记录表上的病人治疗报告。护理人员(护士/医生)利用这个报告监测病人对治疗和治疗步骤的反应。第三个产品是出生/死亡报告(IP₃),是交给人口统计登记部门的。第四个产品是健康报告(IP₄),是公共卫生部门要求提供的有关治疗和出院病人、疾病、治疗及出院原因的分类报告,要求每年提交两次。第五个产品(IP₅)是交给保健组织(HMO)报销的医疗费账单,是一份详细列明向病人提供的服务、器材使用(如果有)、药物治疗、化验及治疗程序的清单。前2个产品都是医院内部使用的信息产品,后3个产品是送给外部机构的。

图1显示了描述撰写病人住院报告(IP₁)的信息产品图。有3个地方接收病人住院:住院处、急诊部和妇产科。病人(或陪伴人员)填写一张表格,提供病人的信息(从数据源DS₁获取的原始数据RD₁)。住院处工作人员利用格式界面(处理模块P₁)把这些数据输到接诊的医生办公室系统中(图2)。

在这个过程中,数据从纸质系统变成了电子化系统,系统边界块SB₁显示了这种情况。与这个界面相关的软件模块检查表格的完整性,并审核保证人/保健组织。这个审核被显示为QB₁。这些审核过的原始数据元素连同批准文件被送去存储,成分数据CD₁显示了这种情况。

病人入院后,接收病人的病房护士负责安排

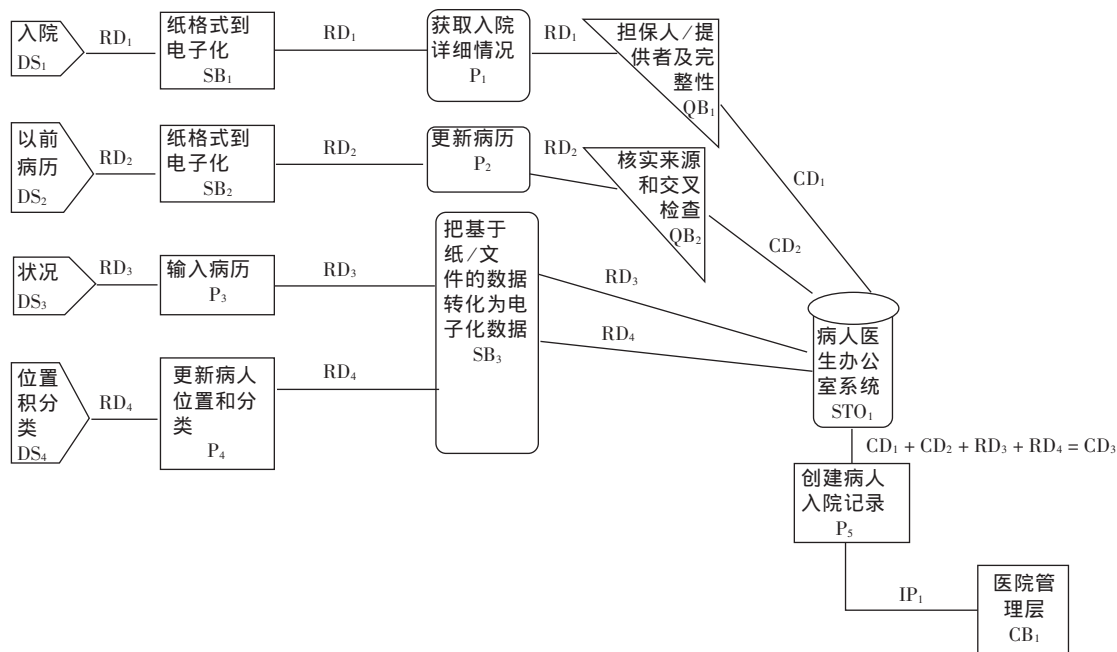


图1 病人入院记录的信息产品图

相应病房的床位号(心脏病重症加护病房、普通病房),此外还要检查病人的总体情况和性格。护士(被看成是两种数据源 DS₃ 和 DS₄,因为很多护士可能要完成这两项工作)在病人记录表上记录下这些信息(RD₃ 和 RD₄),并输入到电脑系统(处理模块 P₃ 和 P₄)中。由于底层系统有变化,因此让系统边界块(SB₃)来代表这种变化。调取病人以前的病历(源块 DS₂),获取的信息(RD₂)被用于更新病人在该系统中的病历(处理模块 P₂)。确定这些病历是来自病人认可的合适来源,如有必要,进行记录的医生/医生办公室可对这些信息进行核实。质量块 QB₂ 就代表这种核实。之后,由此产生的成分数据(CD₂)被送去存储。所有这些信息都可以在医生办公室系统的数据库中获得,存储块 STO₁ 表示了这种情况。处理模块 P₅ 产生的信息产品 IP₁,使用了被看作为 CD₃ 的成分数据项集。它被送至医院管理层,如消费者块 CB₁ 显示的那样。

一旦入院手续办完,就创建了一条推荐并可实施的治疗/处置记录,如图2中的信息产品图所示。专家医生和主治医生(数据源 DS₇ 和 DS₈)推荐可供实施的治疗和处置/化验过程。这些信息之后便被记录(RD₇)在病人记录表上。在填写记录前,主治医生要进行审核,并咨询专家医生进行修改(如有必要)。质量块 QB₄ 代表了这项审核。这样,

万方数据

审定的治疗/处置信息(CD₅)就可以通过处理模块 P₈ 在计算机系统中获取。主治医生还报告病人的治疗进展,并签字结束完成的推荐治疗/处置,如 RD₈ 所示,这些信息可以通过处理模块 P₉ 在计算机系统中获取。SB₅ 代表了系统从基于纸到电子化的变化。化验和放射报告(数据源 OS₅)被收集后,这些信息(RD₅)被输入电脑中。SB₄ 代表了系统的这种变化。处理模块 P₆ 获取这些信息,这个流程中的一个模块也要核实这些报告的来源。P₆ 生成的成分数据 CD₄ 与 QB₃ 所示的病人记录一致,并被送去存储。

手术室(不同于病人护理部门)的意见和报告通过处理模块 P₇ 被电子化上传。业务边界块 BB₁ 代表了跨业务部门的这种信息传输。所有这些信息都存储在病人护理系统数据库中,如存储块 STO₂ 所示。治疗报告(IP₂)通过处理模块 P₁₀ 创建,并被送至医疗人员手中(消费者块 CB₂)。

图3中的信息产品图描述了信息产品 IP₃、IP₄ 和 IP₅ 的生产过程。通过处理模块 P₁₁ 和 P₁₂,住院部系统和病人护理系统中的信息被上传到管理系统中。每个系统中的记录是一致的,确保了入院病人能对应正确的治疗(质量块 QB₆ 所示),所产生的成分数据 CD₁₀ 被存储到管理系统数据库中,STO₃ 代表了这种情况。由于这3个系统各不相同,

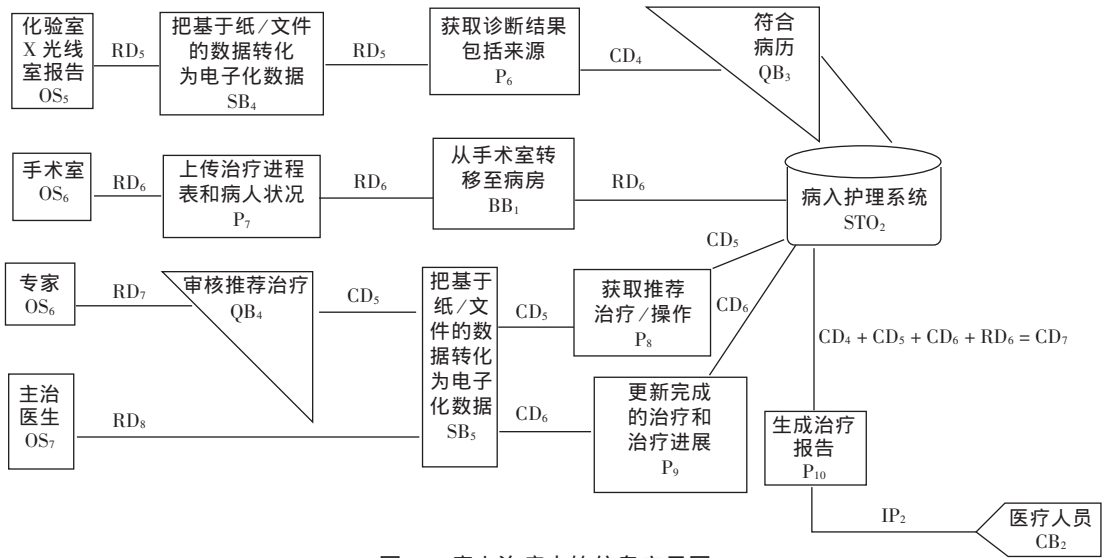


图 2 病人治疗史的信息产品图

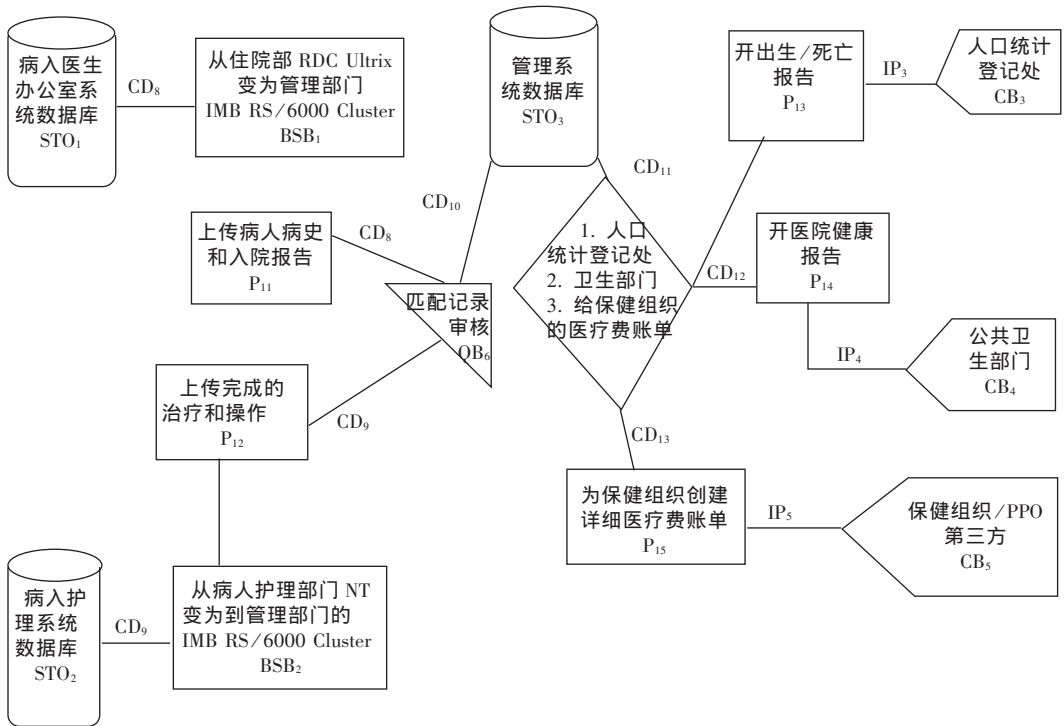


图 3 人口统计报告、医院健康报告和医药费账单的信息产品图

因此，我们必须指出在这种转变过程中底层系统的变化。我们还必须了解这样一个事实，即信息能改变业务边界。我们使用系统与业务边界组合块 BSB₁ 和 BSB₂ 代表这种变化。处理模块 P₁₃ 创建人口统计报告 (IP₃)，这些报告被送给消费者 (CB₃) 和人口统计登记部门。处理模块 P₁₄ 和 P₁₅ 分别创建医院健康报告 (IP₄) 和病人医药费账单 (IP₅)。州卫生部门 (CB₄) 和保健组织 (CB₅) 是这两个信息产品

的消费者。用来生产每个产品的数据项集是完全不同的，成分数据项 CD₁₁、CD₁₂ 和 CD₁₃ 表明了这种情况。

为了完成这种表示，我们必须获取模型中的每个数据流程所包含的每个块和每个数据元素的信息。这类似于数据流程图的数据字典，我们称之为与这个模型相关的元数据。元数据可以在数据库中获取。这个模型的全部元数据太大，无法在本

表 2 信息产品图的样本数据

名称/类型	部门/角色	地点	业务流程	基础系统构成
入院 - DS ₁	住院部/病人	住院部, 妇产科, 急救室	标准格式 (#1101P)	纸质病历
以前的病历 - DS ₂	住院部/住院部工作人员	住院部大楼, 病历室	病人认可的联系方式和请求	纸质病历

文表示 因此表 2 仅列举了一个示例。

4 结 语

在本文中,我们介绍了信息产品图(IP - Map)的概念。我们提出了一套绘制信息产品图的符号和制图规则。开发信息产品图是一项相对较新的工作,这些符号和规则仍需不断改进。不过,就评估、判断和向组织中的个人介绍数据质量来说,这些信息产品图的表述是必不可少的。这些信息产品图是判断数据质量好坏不可或缺的工具。在多数组织中,没有哪个人或团体掌握了与关键信息

产品相关的所有流程、系统、人员以及组织部门的全部知识。开发和利用信息产品图是一个促进有关收集、存储、维护和使用组织数据的组织知识管理的机制。

参考文献

- [1] D Ballou, R Wang, HPazer, G Kumar. Modeling Information Manufacturing Systems to Determine Information Product Quality[J]. Management Science, 1998, 44(4): 462-484.
- [2] G Shankaranarayan, M Ziad, RY Wang. Managing Data Quality in Dynamic Decision Environments: An Information Product Approach[J]. Journal of Database Management, 2003, 14(Oct. -Dec.): 14-32.
- [3] LL Pipino, YW Lee, RY Wang. Data Quality Assessment [J]. Association for Computing Machinery. Communications of the ACM, 2002, 45(Apr.): 211-218.
- [4] E M Pierce. Assessing Data Quality with Control Matrices[J]. Association for Computing Machinery. Communications of the ACM, 2004, 47(Feb.): 82-86.

Developing Information Product Maps

Ballou, Donald

(State University of New York at Albany, New York Albany 12222)

Assessor: Su Ying

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: To implement the information product(IP) approach, an organization needs not only a supportive philosophy but also models, tools, and techniques. Despite many papers have discussed a number of ways in which an organization can foster data quality awareness and identify root causes. Those methods are part of the tool kit of the information product manager and data quality analysts. Those in conjunction with the construct of information product maps provide a formidable set of tools to implement the IP approach and support a viable data quality program. It is important to make clear at the outset that the development of a standard methodology for producing IP - Maps is an ongoing project and that the ideas and constructs presented here are evolving. Nevertheless, the constructs are usable and useful. Work is ongoing to establish standards for the creation and maintenance of information product maps. An ad hoc committee of academics and data quality practitioners has been working to establish these standards. We refer to this committee as Standards Group throughout this paper.

Keywords: information product map, data quality, information quality, information product