

IP-MAP: Representing the Manufacture of an Information Product

G. Shankaranarayanan {gshankar@acs.bu.edu}
Richard Y. Wang {rwang@acs.bu.edu}
*Department of Management Information Systems
School of Management, Boston University*

Mostapha Ziad {mziad@suffolk.edu}
Department of Computer Information Systems, Suffolk University

Abstract

Organizations have recognized the need for high quality information and academics have proposed several methods to measure and improve information quality. One such method is to manage information as an information product (IP). Many modeling methods for information manufacturing systems have been described. Almost all of these lack the ability to systematically represent the manufacturing processes and are deficient in constructs offered to explicitly represent manufacturing details. In this paper we propose the information product map (IP-MAP) as a method to systematically model the manufacture of an IP. The IP-MAP is an extension of the Information Manufacturing System (IMS) proposed earlier. It offers several advantages including the ability to visualize the manufacture, implement continuous improvement and quality-at-source, and measure the quality of the IP using appropriate quality dimensions. We submit that the IP-MAP would serve as a foundation upon which a suite of quality dimensions may be identified and implemented for managing the IP quality.

1. Introduction

More than ever before, organizations and the decision-makers within understand the importance of high quality information and its impact on the decisions made. Information quality and methods to improve it have been the focal point of research during the past decade. Organizations have defined high quality standards for their information and are striving to achieve and maintain these standards. It is now recognized that information may be treated as a *product* and the steps involved in creating it as a set of *manufacturing* processes. Several methods have been proposed to increase and continuously improve the quality of information. Some important ones include defining quality measures for an information product [1], defining the principles for managing information as a product [7], specifying practices for continuously improving the processes involved [5], and identifying benchmarks for information quality [6].

Academicians in the field of information quality have investigated managing the quality of an information product (IP). Methods for Total Quality Management (TQM) have been applied successfully in manufacturing environments. To apply these methods and processes to

manage the quality of an IP it is helpful to view the manufacture of an IP as a sequence of processes that must be represented accurately. Manufacturing an IP is akin to the manufacture of a physical product. Raw materials, storage, assembly, processing, inspection, rework, and packaging (formatting) are all applicable to an IP as well. Components and /or processes of a physical product may be subcontracted (outsourced). This is comparable to processing performed on an IP by a different agency or even a different department within an organization using a different set of computing resources. While the list of comparisons above is not exhaustive, one can still confidently say that modeling the flow of processes for an IP can be accomplished along the same lines as those for a physical product. Examples of information products include

- A Birth/Death Certificate
- A Hospital Bill
- A Student Transcript
- A Prescription for grinding eye glasses
- A Monthly Bank Statement

Although there have been attempts at developing models of an information manufacturing system, certain critical aspects of the stages that an IP goes through within the manufacturing system have not been addressed. Specifically, almost all models lack a systematic method for representing the processes involved in manufacturing (or creating) the IP. In this paper, we refer to this systematic representation as an information product map (IP-MAP). We extend the information manufacturing system model proposed by Ballou et al. [1] to develop a formal modeling method for creating an IP-MAP. This representation offers several advantages. First, using this representation, the IP manager will be able to visualize the most important phases in the manufacture of an IP and identify the critical phases that affect its quality. Second, the conceptual representation would allow IP managers to pinpoint bottlenecks in the information manufacturing system and estimate the time to deliver the IP. Third, based on the principles of continuous improvement for the processes involved, the IP-MAP representation would not only help identify ownership of the processes at each of these phases but would also help in implementing quality-at-source. Fourth, the representation would permit IP managers to understand the organizational (business units) as well as information system boundaries spanned by the different processes used in the manufacture of the IP. Finally, it permits the measurement of the quality of the IP at the various stages of the manufacturing process using appropriate quality dimensions.

The remainder of this paper is organized as follows. Section 2 describes the framework and the constructs used in the IP-MAP representation. The metadata associated with each construct is also described here. The application of the framework using a real world case is presented in Section 3. The conclusion and directions for further research including suggestions for incorporating quality dimensions are described in Section 4.

2. A Framework for IP-MAP

The proposed framework is aimed at creating a systematic representation for capturing the details associated with the manufacture of an IP. The objectives of this representation are:

- To provide a set of constructs that facilitate the representation of the steps involved in the manufacture of an IP. These constructs help model the various steps of the production process and assist the modeler in visualizing and representing these steps. The representation hence serves as a conceptual model for the manufacture of the IP.
- The representation allows the modeler to critically examine the steps in the manufacturing process. These steps include the arrival of the data elements (raw data), the locations for storing these data elements, the processes involved in creating, converting, and/or assembling the existing (or new) data elements, and the procedures for evaluating the IP and the work-in-progress for quality and correctness. Hence, the modeler/user can locate potential sources of information quality problems and more importantly, design procedures to rectify these problems thus ensuring a high quality IP.
- Further it allows the modeler to implement information quality-at-source (similar to quality-at-source in manufacturing). It permits the modeler to assign the responsibility of ensuring the quality of the IP - including the work-in-progress that comes into and leaves a “work area” - to individuals performing one or more steps in the manufacturing process within that “work area”.
- Finally, it provides a formal representation that can be used to assess the quality of an IP based on the selected information quality dimensions.

The proposed model can be explicitly differentiated from the one described in Ballou et al. [1] by the following new/extended features:

- It adopts a "top-down" approach in that the design requirements of the final product drive the design of the IP-MAP. The IP designers and developers are required to precisely specify the raw or component data items that are needed to produce a particular IP. This specification provides the IP manager with the ability to look ahead and determine the feasibility of creating the IP. It is possible that some raw data items may not be available or cannot be produced in the current situation.
- Several extensions to the constructs (blocks) introduced by Ballou et al. [1] have been defined. These constructs facilitate the explicit representation of details in the manufacture of the IP, and include the *decision* block, the *organizational/business boundary* block, and the *information system boundary* block.
- A repository for capturing the metadata associated with the constructs in the IP-MAP is also defined. The metadata adds to the ability of the IP-MAP to comprehensively track and manage the information associated with the IP and serves to resolve issues concerning the quality of the IP.
- Our model permits checking every raw input data item for data quality problems before it can be used in the production of an IP. The Ballou et. al. model, on the other hand, allows raw input data items to be used without checking their quality first and uses a quality block only

when the raw input data items have “historically exhibited data quality deficiencies”. We have not explicitly addressed the incorporation of quality dimensions into the IP-MAP in this paper.

Prior to explaining the constructs and modeling procedures, let us first examine the composition of an IP and distinguish it from a physical product, both in terms of its composition and in terms of its manufacture. Unlike a physical product where the *overall product and its quality* are of interest to the consumer, for an information product it is the *data items that comprise the IP and the quality of each* that are of importance to the consumer. The IP must therefore be identified in terms of the data items used to manufacture it and must be specified by the final consumer of the IP. This breakdown drives the requirements of the raw information (*raw data items*) and the semi-processed information (*component data items*) needed to manufacture the IP. Raw data is defined as data (or information) that comes from outside the boundaries of the IP-MAP. Component data is defined as data that might be generated within the IP-MAP and used in creating the final IP.

Another distinct feature of IP manufacture compared to the manufacture of a physical product is that the raw data items and the component data items do not deplete when an IP is manufactured [1]. While the raw data items may be stored for long periods of time, the component data items are stored temporarily until the final product is manufactured. The component items are regenerated each time an IP is needed. The same set of raw data and component data items may be used in the manufacture of several different products. The set of information products that uses the same set of raw and component data items may be considered a family (or group).

An IP is created using data that is constantly being captured and stored. For example, the data on the birth of a child, the infant screening tests for congenital defects, etc. (raw materials) are being captured in anticipation of the fact that the parents may request a birth certificate (the IP) some time in the future. It is possible that this birth certificate (for a particular child) may never be requested and hence may never be created. The same raw data (material) could also be used for manufacturing a different IP such as a government report on vital statistics generated by the hospital to justify or request government grants. Also, the raw material is never depleted even after multiple information products have been manufactured using this raw material. It is therefore imperative that a good representation accurately tracks the details on *what* triggered the capture of this data along with *how*, *who*, and *where* it was captured.

To address the “what” question we define events (or triggering events) that initiate data capture. Associated with an event are entities. For example, the birth of a child is an event. The event is triggered by the actual birth of a child (an entity), delivered by a gynecologist or a midwife (another entity) in a hospital labor room, operating room, or a residence (yet another entity). The gynecologist is said to be a data creator (producer). The data creator may verbally inform the nurse (a data collector) who records the data on a piece of paper. In this case, the medium of recording is said to be a paper-based system, and the nurse is also the entity that enters the data into the system. The data creator, data collector, and the data recorder are all entities that answer the “who” question. In general, there are three kinds of roles under the data supplier category: data producer or creator, data collector, and data entry clerk. Often, however,

a single person may assume several or all of these roles. For example, a gynecologist may record the birth event into a computer system. In this case, the gynecologist is said to have assumed all the three roles as a data creator, data collector, and data entry clerk; and the medium is a computer-based system.

For an IP “where” could have two implications: the physical location (hospital emergency room, OBGYN department) and the system used to capture the information about the product (paper form, computer system using a word processor, computer system with a specific application data entry interface, etc.). Both are important when tracking the IP as they serve to identify the "source" of the data and more importantly, the factors that (may) affect the quality of the data associated with each source. The proposed model is designed to capture this data (metadata) about the source explicitly.

Once the raw data items are received from the data sources they go through a sequence of steps that result in the IP. The steps include storage, quality evaluation, and processing, before ending up at the destination (or sink) of the IP-MAP. Some of these steps result in the creation of component data items. To comprehensively capture and represent these steps and the information associated with each, the model supports constructs (blocks) and a repository for capturing the metadata associated with each block. The five constructs supported by Ballou et al. [1] that are retained in our IP-MAP are the source, sink, process, quality, and storage blocks. The IP-MAP extends these constructs with three new constructs; the decision, the information system boundary, and the departmental/organizational boundary blocks. As the IP-MAP attempts to capture more details and is more explicit in its representation, it provides a detailed description of the eight blocks along with the metadata associated with each. The constructs and the symbols used in to represent them are listed in Table 1.

The IP-MAP allows us to track an information product as it goes through the various stages of an information manufacturing system. A single IP manufacturing system should be capable of manufacturing all the variants of a product P. In some respects the idea is similar to that of a manufacturing cell created based on group technology. As the raw data requirements for the system that manufactures product P includes the raw data needs of all its variants, it is conceivable that the initial processing for P and its variants may be the same. Decision blocks help address the differences in manufacturing requirements of each variant that can be manufactured in an information manufacturing system. We introduce the new blocks to capture these additional details.

2.1 The Building Blocks of IP-MAP

The modeling constructs in the IP-MAP consist of eight types of construct blocks. Any block used in modeling the IP-MAP corresponding to some IP must belong to one of these 8 block types. Each construct block is identified by a unique and non-null *name*. Each construct block is described by a set of attributes. The composition of this set varies depending on the type of construct block it is associated with. For instance, if the IP is a birth certificate, we might have a source block named *Gynecologist* with an associated set of attributes that includes *date of birth, time of birth, place of birth, sex, weight, and the mother's name*.

2.1.1 *Source (raw input data) block*

This block is used to represent the source of each raw (input) data that must be available in order to produce the IP expected by the consumer. Associated with this block are the business unit /role responsible for the raw input data, the process that was used in capturing it, the data element(s) that make up the raw data, and the underlying system in which it was stored and from which it originates.

Source Block =

{<Name>, <Department/Role>, < location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) creating the raw data provided by this source are captured in *Department/Role* field. *Location* refers to the physical location (building, floor, room, area etc.) where the source is located. The *Business process* is a description of the set of rules/procedures associated with the raw data. *Composition* captures the data items that the raw data is composed of. It is meaningful only if the raw data consists of more than one data element. The *base system* identifies the system (paper-based, electronic) that stores this raw data supplied.

2.1.2 *Customer (output) block*

This block is used to represent the consumer of the IP. The consumer specifies in this block the data elements that constitute the “finished” IP. Consequently, in our model, the consumer has to define the IP in advance. Associated with this block are the name of the business / organizational / departmental unit in charge of the IP, the name of the entity that will actually use the product, and the set of data items that make up the IP.

Customer Block =

{<Name>, <Department/Role>, < location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) consuming the information product are captured in *Department/Role* field. *Location* refers to the physical location (building, floor, room, area etc.) where the consumer is located. The *Business process* is a description of the set of rules/procedures associated with the consumption of the IP. *Composition* captures the data items that the IP is composed of. The *base system* identifies the system (paper, computerized) that receives the IP.

2.1.3 *Data Quality Block*

This block is used to represent the checks for data quality on those data items that are essential in producing a “defect-free” IP. Therefore, associated with this block we have a list of the data quality checks that are being performed on the specified component data items. The quality block has two possible outputs: the “correct” stream (with probability p) and the “incorrect” stream (with probability $1-p$). The inputs to the quality block are the raw input data items and possibly some components data items.

Quality Block =

{<Name>, <Department/Role>, < location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) responsible for verifying the quality of the set of data elements entering this block are captured in the *Department/Role* field. *Location* refers to the physical location (building, floor, room, area etc.) where the quality check is performed. The *Business process* is a description of the set of rules/procedures used for checking the quality of the data elements involved. In this block, it is possible that only a subset of the data elements is evaluated

for quality. *Composition* captures the data items that enter the quality block. The *base system* identifies the system (paper-based, computerized) that receives the data elements being checked for quality.

2.1.4 Processing Block

This block is used to represent any manipulations, calculations, or combinations involving some or all of the raw input data items or component data items required to ultimately produce the IP. We allow for the specification of the processing requirements to be associated with the block. The processing block is also described by a set of 6 items as follows:

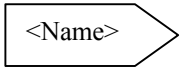
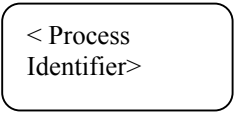
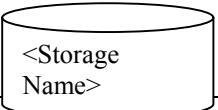
Process Block =


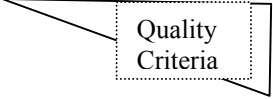
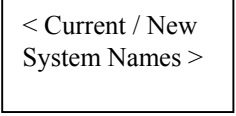
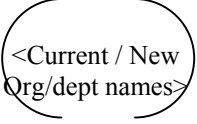
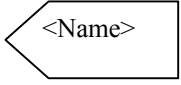
{<Name>, <Department/Role>, <location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) responsible for processing the set of data elements (raw and/or component) entering this block are captured in *Department/Role* field. *Location* refers to the physical location (building, floor, room, area etc.) where the processing takes place. The *Business process* is a description of the set of rules/procedures associated with the process. This is of importance in this block as it defines the processing tasks that act upon the set of data elements that arrive at this block and should include the subset of the data elements that are actively processed here. *Composition* captures the data items that enter the processing block. The *base system* identifies the system (paper-based, computerized) that receives data elements and where the processing occurs.

Data Correction Block: The data correction block is a specialization of the process block. When quality problems are identified in a set of data elements that enter the quality block, some corrective action is required to fix these errors. The block represents a process that is not part of the standard processing sequence but is utilized under special circumstances. Therefore any raw input data items or component data items that go through the correction block can be considered “cleansed” and used by subsequent blocks obviating the need to loop back into the quality block.

Table 1: IP-MAP Constructs and Symbols used to represent each

SYMBOL	REPRESENTS
	<p align="center">Data Source / Data Vendor / Point-of-Origin</p>
	<p align="center">Process</p>
	<p align="center">Data / Information Storage</p>

	Decision
	Quality / Evaluation / Check
	Information System Boundary - used when a data unit (raw data, component data) changes from one system (paper or computerized) to another (paper or computerized)
	Organizational Boundary - used when a data unit (raw, component) moves across departments or across organizations
	Data Sink / Consumer Block / Point-of-Destination.

2.1.5 Data Storage Block

This block is used to represent the capture of data items in storage files or databases so that they can be available for further processing. Storage blocks may be used to represent data items (raw and/or component) that wait for further processing or are captured as part of the information inventory in the organization. The representation of a storage block follows the standard used for representing other blocks in the IP-MAP.

Storage Block =

{<Name>, <Department/Role>, < location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) responsible for capturing the set of data elements (raw and/or component) entering this block are captured in *Department/Role* field. *Location* refers to the physical location (building, floor, room, area etc.) where the capture takes place and where the data elements reside in storage. The *Business process* is a description of the set of rules/procedures associated with the management of the stored data elements. In the case of the storage block it might include the set of procedures associated with how the data must be managed and authorizations involved in the release of the data elements stored. *Composition* captures the data items that are stored. The *base system* identifies the system (paper-based, computerized) that receives data elements and where the elements are stored.

2.1.6 Decision Block

In some complex information manufacturing systems, depending on the value of some particular data item(s), it may be necessary to direct the data items to a different set of blocks downstream for further processing. In such cases, we need a decision block to capture the

different conditions to be evaluated and the corresponding procedures for handling the incoming data items based on the evaluation. For example, the same set of birth-related data items may be used to generate a birth certificate, or a summary report of vital statistics, or a report that might accompany the blood sample from the newborn to test for congenital diseases. Each of the above is a unique IP using the same set of raw and component data. The final stages in the manufacture of each may be different while the initial stages where the raw data is received, checked, processed, and assembled together with other raw data elements may all be the same. The decision block at the end of the initial stages identifies what the final product is and directs the data elements to the appropriate set of final stages to be completed.

Decision Block =

{<Name>, <Department/Role>, <location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) responsible for evaluating the decision alternatives are captured in *Department/Role* field. *Location* refers to the physical location (building, floor, room, area etc.) where the decision process occurs. The *Business process* is a description of the set of rules/procedures associated with the decision process. In the case of the storage block it might include the set of "IF-THEN-ELSE" rules that define the different alternatives and the procedures for directing each. *Composition* captures the data items that enter the block. The *base system* identifies the system (paper-based, computerized) that receives data elements and where decision process occurs.

2.1.7 Business Boundary Block

We introduce the business boundary block to represent instances where the raw input data items or component data items are "handed over" by one business (or organizational) unit to another unit. The business boundary block is used to specify the movement of the IP (or raw / component data) across departmental or organizational boundaries. The role of the business boundary block in the IP-MAP is to highlight the data quality problems that might arise when crossing business unit boundaries and therefore assign accountability to the appropriate business unit. The business boundary block is described as follows:

Business Boundary Block =

{<Name>, <Department/Role>, <location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) responsible for authorizing the movement of the data elements involved are captured in *Department/Role* field. In this case, it consists of a pair of values one for each of the units involved. *Location* refers to the physical locations (building, floor, room, area etc.) of the units involved in the movement of the data elements. The *Business process* is a description of the set of rules/procedures associated with the movement. Oftentimes some process such as an EDI, Email, or FTP accomplishes the movement of data across units. *Composition* captures the data items that enter the block. The *base system* identifies the system (paper-based, computerized) that accomplishes the movement of the data elements involved.

2.1.8 Information System (IS) Boundary Block

This block is used to reflect the changes to the raw input data items or component data items as they move from one information system to another type of information system. These system changes could be intra or inter-business units. The information system boundary block is

used to specify the two information systems involved. Examples of information system changes include moving the data items from a particular type of database management system (DBMS) to another type of DBMS, or getting the data items from a particular data warehouse system to a DBMS, or transcribing the data items from a paper-based system to an on-line DBMS. Also, the type of media used in the transfer of data items between the two information systems is specified as part of the description of the IS boundary block. Examples of media include paper forms, magnetic tapes and disks, and CD-ROMs. The use of such types of media to transfer data items from the first IS to the second IS can adversely affect important data quality parameters such as timeliness and correctness. For instance, if the data items produced are in paper-based system, then there is a need for an agent, for example a data entry clerk, to extract (visually read) the paper-based data items and enter them in the second system (which could be a database). This manual operation is not only time consuming, adversely affecting timeliness, but can potentially introduce errors during the transcription / entry of the information. On the other hand, if the data items produced by the first system can be automatically captured by the second system (as in the case of an EDI-based system), the operation is not only faster improving timeliness, but also more accurate.

IS Boundary Block =

{<Name>, <Department/Role>, <location>, <Business Process>, <Composition>, <Base System>}

Name is the unique name by which the block is identified. The business unit and the role associated with the individual(s) responsible for moving the data elements is captured in *Department/Role* field. *Location* refers to the physical locations (building, floor, room, area etc.) of the systems involved in the movement of the data elements. The *Business process* is a description of the set of rules/procedures associated with the movement. Oftentimes some process such as an EDI, Email, or FTP accomplishes the movement of data across systems (and across units). *Composition* captures the data items that enter the block. The *base system* identifies the two (or more) systems (paper-based or computerized) that send and receive the data elements.

There are circumstances where the raw input data items or component data items go through both a business boundary and a system boundary change. We need a block to capture this situation. The **combined business-information system boundary block** is defined for this purpose. As an example, consider the real life scenario where the information about a child's birth is handed over in a paper form from a hospital to the state registrar of births. There, a clerk enters the information electronically in the state's DBMS. This example requires the use of a combined business-information system boundary block to capture the hospital to state business boundary transition and the paper-based to the electronic birth certificate information system boundary transition.

3. Constructing an IP-MAP

In the preceding section we described the constructs that are useful in creating a conceptual modeling method to represent the manufacture of an IP. We use an example to illustrate how these constructs are used in defining the IP-MAP. Let us examine some important reports (IPs) typically produced in a hospital. For the purpose of this illustration we will consider a small subset of the operations (and processes) of a major hospital including only the in-patient admissions, treatment, and discharge sections. There are five products associated with these

operations. All five use information that is gathered from two important sources: the patient and the team of hospital employees (doctors, nurses, lab technicians, radiologists, therapists, and administrative staff) involved (directly or indirectly) in the admission, treatment, or discharge of the patient. Each uses a subset of the large set of information. The first product (IP₁) is the *admissions report* submitted to the management of the hospital on a daily, weekly, and monthly basis. It provides a description of the number of patients admitted, expected duration of stay, along with patient information and serves as a monitoring instrument that reflects how busy the units are. The second product (IP₂) is the *patient treatment report* generated on a daily basis and appended to the patient's chart. Care providers (nurses/doctors) use it to monitor the patient's response(s) to treatments and procedures administered. These two are information products used internally within the hospital. The final three products are sent to external agencies. The *birth/death report* (IP₃) is submitted to the registry of vital statistics, and the *health report* (IP₄) is a bi-annual report required by the department of public health on the types of patients treated and released, ailments, treatments, and the reason for discharge. The final product (IP₅) is the *patient bill* submitted to the HMOs for payment. This is an itemized list of services, equipment charges (if any), medications, tests, and procedures provided to the patient.

The IP-MAP representing the manufacture of the patient admission report (IP₁) is shown in Figure 1. An inpatient may be admitted at any one of three locations: the admissions office, emergency room, or in the department of maternal and fetal medicine. The patient (or an accompanying adult) provides the patient information (raw data RD₁ from data source DS₁) by completing a form. The admissions clerk enters this data into the Patient Medical Office System using a form-based interface (process P₁). In this process the data changes from a paper-based system to an electronic system shown by the system boundary block SB₁. The software module associated with the interface checks the form for completeness, and verifies the guarantor/HMO and this check is shown as QB₁. The raw data elements examined along with the authorization is sent for storage and is shown by the component data CD₁.

Upon admission, the ward nurse responsible for admitting the patient assigns a bed number that specifies the type of ward (cardiac ICU, general, etc.) and also examines the general condition and disposition of the patient. The nurse (treated as two data sources DS₃ and DS₄ as the two tasks may be done by more than one nurse) records this information (RD₃ and RD₄) on the chart and subsequently enters it into the computer system (process blocks P₃ and P₄). As the underlying system changes a system boundary block (SB₃) is shown to represent this change. The patient's past medical records (source block DS₂) is obtained and the information (RD₂) is used to update the patient's medical record in the system (process P₂). The records are verified to ensure that they come from the right source authorized by the patient and, if necessary, the information is verified with the doctor/medical office that created the record. Quality block QB₂ represents this check. The resulting component data (CD₂) is then sent for storage. All of this information is captured in the data storage of the medical office system shown by the storage block STO₁. The information product IP₁, generated by process P₅, uses a collection of component data items cumulatively identified as CD₅. It is sent to the hospital management as shown by the consumer block CB₁.

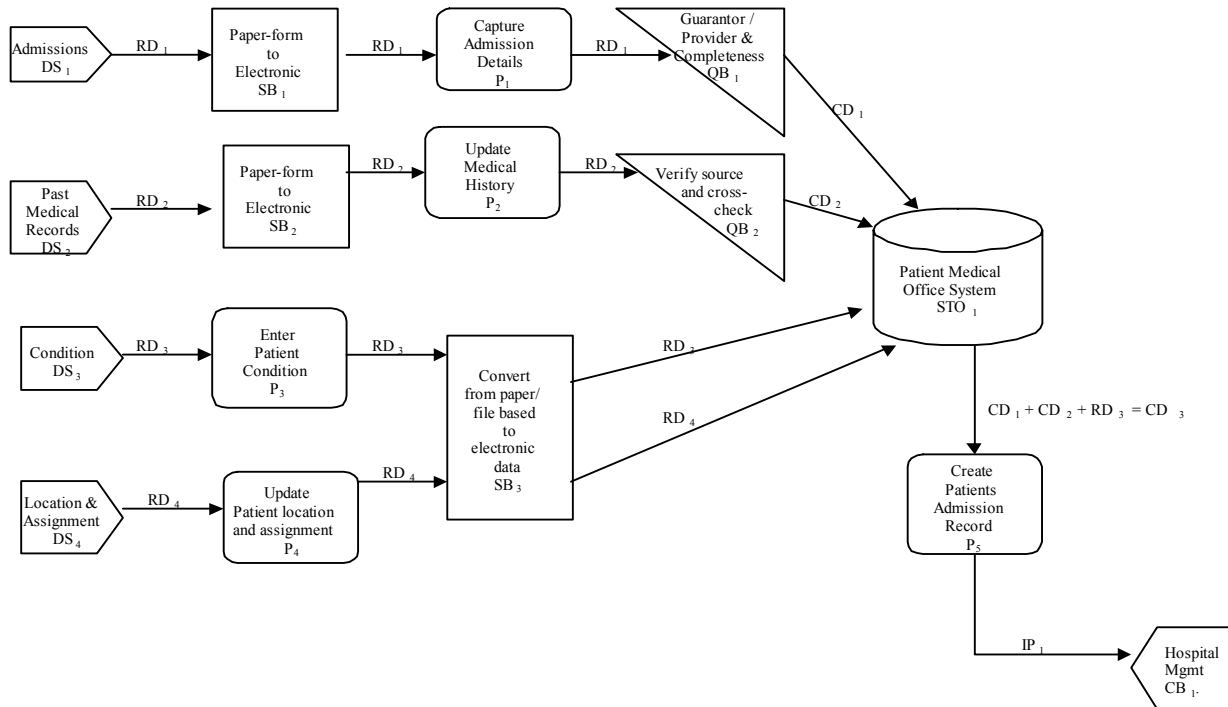
Once the admission is complete, a record of the treatments / procedures recommended and performed on the patient is created as shown by the IP-MAP in Figure 2. The specialists and

attending physicians (data sources DS₇ and DS₈) recommend the course of treatment and procedures/tests to be performed. This information is then recorded (RD₇) on the charts. Prior to its capture, it is verified by the attending physicians and modified (if needed) in consultation with the specialist. The quality block QB₄ represents this check. The resulting authorized treatments/procedure information (CD₅) is captured in the computer system by process P₈. The attending physicians also report on the progress of the patient and sign off on the recommended treatments/procedures completed as shown by RD₈ which is captured in the system by process P₉. The change of system from paper-based to electronic is represented by SB₅. The reports from the labs and radiology (data source DS₅) are collected and the information (RD₅) is entered into the computer. The change in system is represented by SB₄. Process P₆ captures this and a module in this process verifies the source of the report as well. The component data CD₆ generated by P₆ is matched with the patient's record shown by QB₄ and sent for storage.

The comments and reports from the surgical unit (different from the patient care facility) are electronically uploaded by process P₇. The business boundary block BB₁ represents the transfer of information across business units. The storage location for all the above information is the Patient Care System database shown by storage block STO₂. The treatment report (IP₂) is created by process P₁₀ and sent to care givers (customer block CB₂).

The manufacture of the information products IP₃, IP₄, and IP₅ is represented by the IP-MAP in Figure 3. The information in the admissions office system and the patient care system is uploaded into the administrative system by processes P₁₁ and P₁₂. The records from each are matched to ensure that the right admission is combined with the right treatment (shown by quality block QB₆) and the resulting component data CD₁₀ is stored in the administrative system database represented by STO₃. As all three systems are different, we need to show the change in the underlying system during this transfer. We also need to capture the fact that the information changes business boundaries as well. We use the combined system and business boundary blocks BSB₁ and BSB₂ to represent the transfer. Process P₁₃ generates the report on vital statistics (IP₃) which is sent to the consumer (CB₃) and the Registry of Vital Statistics. Processes P₁₄ and P₁₅ generate the hospital health report (IP₄) and the patient bill (IP₅) respectively. The state department of health (CB₄) and the HMO (CB₅) are the consumers of the two information products in that order. For each of the three products, the set of data items used to generate each is different and is shown by the component data items CD₁₁, CD₁₂, and CD₁₃.

Figure 1: IP-MAP for Patients Admissions Record



To complete the representation, we need to capture the information about each of the blocks and the data elements included in each flow in the model(s) above. This is akin to the data dictionary for a data flow diagram and we refer to this as the metadata associated with the model. The metadata is captured in a repository whose design is discussed further in Section 4. The complete metadata for the above model is too large for this paper and therefore only a sample is shown in Table 2.

Table 2: Sample metadata for IP-MAP in Figure 1

<i>Name/Type</i>	<i>Department/Role</i>	<i>Location</i>	<i>Business Process</i>	<i>Composed Of</i>	<i>Base System</i>
Admissions /DS ₁	Admissions Office/ Patient	Admissions , OB/GYN, Emergency	Standard Form (#1101P)		Paper-based - Patient File
Past Medical Records / DS ₂	Admissions Office / Admissions clerk	Admissions Bldg., Records Room	Contact source and request with patient authorization.		Paper-based - patient file

Figure 2: IP-MAP for Patient's Treatment History

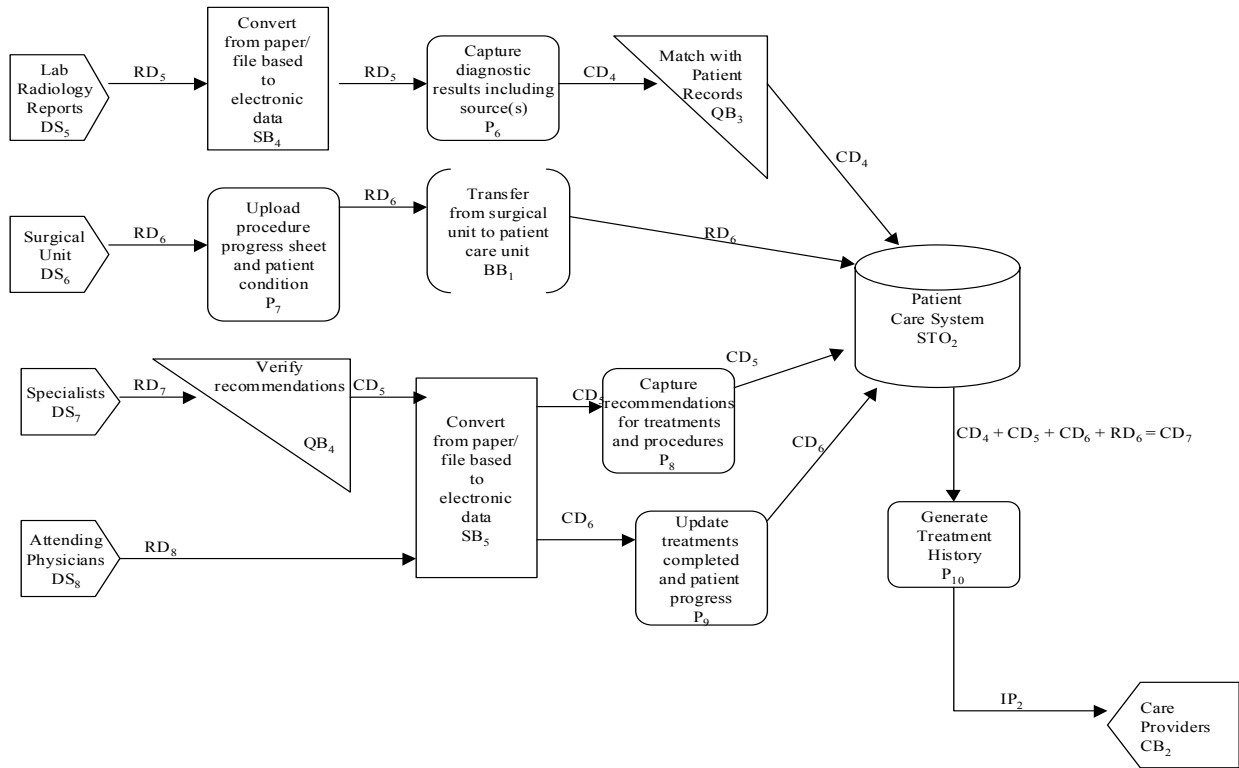
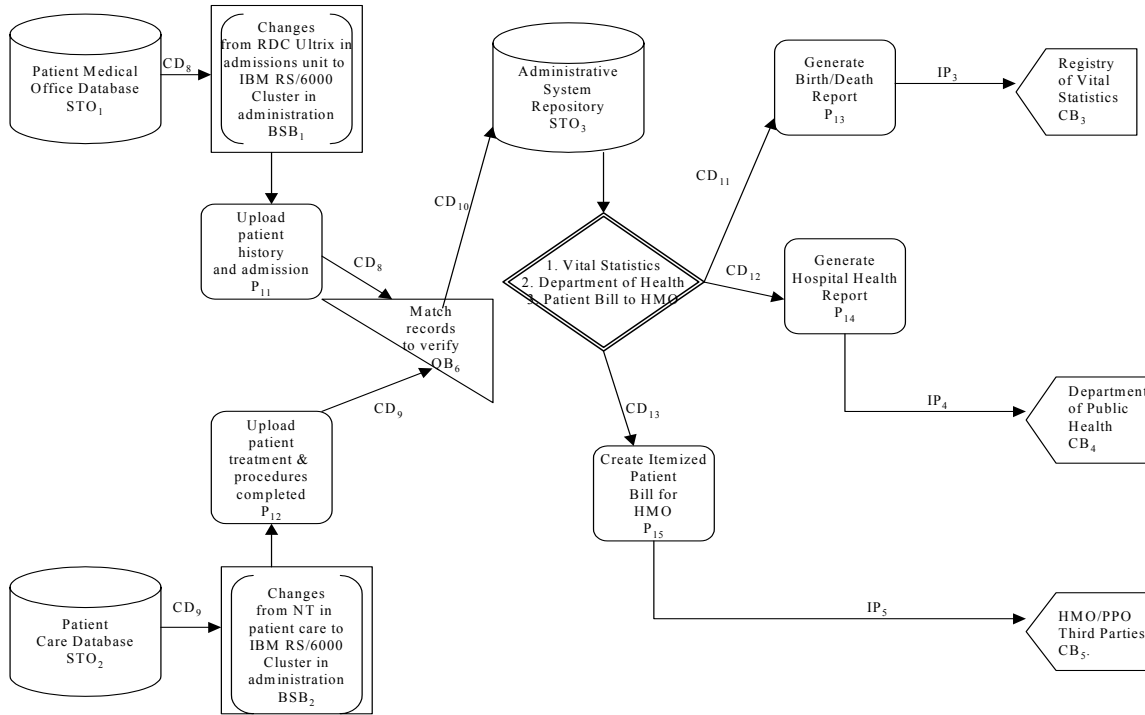


Figure 3: IP-MAP for Vital Statistics Report, Hospital Health Report, and Bill



4. Conclusion and Research Directions

In this research we have presented a modeling method for representing the manufacture of an IP, the IP-MAP. This representation offers several advantages. First, it allows the IP manager to visualize the most important phases in the manufacture of an IP and identify the critical phases that affect its quality. Second, using this representation, IP managers will be able to pinpoint bottlenecks in the information manufacturing system and estimate the time to deliver the IP. Third, based on the principles of continuous improvement for the processes involved, the IP-MAP representation would not only help identify ownership of the processes at each of these phases but would also help in implementing quality-at-source. Fourth, the representation would permit IP managers to understand the organizational (business units) as well as information system boundaries spanned by the different processes / stages in the IP-MAP. Finally, it permits the measurement of the quality of the IP at the different stages in the manufacturing process using appropriate quality dimensions. We hope that this model would serve as a foundation upon which a suite of quality dimensions may be identified and implemented for managing the IP manufacture.

The IP-MAP and modeling constructs we have introduced allow us to represent the manufacture of an information product. Two important issues require some discussion: a comparison of the IP-MAP with existing modeling methods and the utilization of information quality measures incorporated within the IP-MAP.

Several modeling techniques are available for conceptually representing information systems. The most prevalent ones are the data flow diagrams [8], the ER diagram [4] and its extensions, and the object modeling diagrams [2, 3]. The first issue to address is how does the IP-MAP compare with these methods and the niche IP-MAP occupies. We have omitted the details due to the paper length constraint.

An important objective in developing the IP-MAP representation is the ability to incorporate data quality dimensions in each block. The quality dimensions of timeliness, data quality, and cost have been described in detail in Ballou et al. [1]. They have further shown how these dimensions can be incorporated into some of the construct blocks in the IP-MAP. Using the IP-MAP, the IP manager can *trace* the source of a data quality problem in an IP to one or more preceding steps in its manufacture. We define the property of *traceability* as the ability to identify (trace) a sequence of one or more steps that precede the stage at which a quality problem is detected. Also, given two arbitrary stages in the IP-MAP, we must be able to trace the set of one or more stages, in progressive order, between the two. Using the metadata, the individual /role /department that is responsible for that task(s) can be identified and quality-at-source implemented. The second advantage of incorporating the time dimension is it allows the manager to estimate the *duration for delivery* of an IP. In a simplistic sense, time estimation techniques such as CPM/PERT may be applied to the product line to estimate *completion time* of the product. We have used a graph-based representation of the IP-MAP to prove that given any stage in the processing of an IP, it is possible to identify (or trace) the sequence of steps that precede that stage. It can also be shown that given any two stages in the IP-MAP for an IP, we can determine the time taken for the product to go from one stage to the next.

We are in the process of designing and implementing a prototype system for managing the IP-MAP. The system supports a GUI to help designers create the IP-MAP by dragging and dropping the constructs from a tool-bar that houses the icons for the constructs. Metadata on the constructs are captured from the designer at the time of creating each construct using data-entry interfaces. Metadata is stored in a repository maintained in a state-of-the-art relational DBMS. The system has the ability to query the repository and display the information about each construct when the designer/user clicks on a specific construct on the IP-MAP. The system would include the functionality for *traceability* and *duration for delivery* dimensions as well. We hope to be able to identify additional useful quality dimensions that could also be incorporated into the prototype system.

References

- [1] Ballou, D. P., R. Y. Wang, H. Pazer and G. K. Tayi, "Modeling Information Manufacturing Systems to Determine Information Product Quality," *Management Science*, 44(4), 1998, pp. 462-484.
- [2] Booch, G., *Object-Oriented Design with Applications*. ed. B. Cummings. Redwood City, CA, 1994.
- [3] Booch, G., I. Jacobson and J. Rumbaugh, "The Unified Modeling Language Notation Guide version 1.0," 1997, .
- [4] Chen, P. P., "The Entity-Relationship Model - Toward a Unified View of Data," *ACM Transactions on Database Systems*, 1, 1976, pp. 166-193.
- [5] English, L., "Plain English on Data Quality," *DM Review*, 2000, pp. 36, 78.
- [6] Kahn, B. K. and D. M. Strong. "Product and Service Performance Model for Information Quality: An Update", in *Proceedings of the 1998 Conference on Information Quality*. Cambridge, MA: pp. 102-115, 1998.
- [7] Wang, R. Y., Y. L. Lee, L. Pipino and D. M. Strong, "Manage Your Information as a Product," *Sloan Management Review*, 39(4), 1998, pp. 95-105.
- [8] Yourdon, E., *Modern Structured Analysis*. ed. Y. P. C. Series. Prentice Hall, Englewood Cliffs, NJ, 1989.