Developing data production maps: meeting patient discharge data submission requirements

Bruce Davidson*

Resource and Outcomes Management, Cedars-Sinai Health System,
8700 Beverly Boulevard, Suite 2227, Los Angeles, CA 90048, USA
E-mail: bruce.davidson@cshs.org
*Corresponding author

Yang W. Lee

MIS, College of Business Administration, Northeastern University,
214 Hayden Hall, Boston, MA 02115, USA
Fax: 617-373-5052 E-mail: ylee@mit.edu
E-mail: y.lee@neu.edu

Richard Wang

MIT Information Quality Program, E40-209A,
Center for Technology, Policy, and Industrial Development,
1 Amherst Street, Cambridge, MA 02139, USA
Fax: 617-739-9367 E-mail: rwang@mit.edu

Abstract: Recent research in information quality management concluded that firms must manage information as a product, and that the entire information manufacturing system must be managed in order to enable firms to assure delivery of information products with high quality. In this paper, we report a longitudinal case study in a major hospital on how data production maps are developed and used to improve the quality of information products. Specifically, we focus on data production maps for patient discharge data. These data production maps have enabled this hospital to model, analyse, and improve the quality of patient-level data that must be submitted to the State Department of Health Services annually. Implications and lessons learned are also discussed.

Keywords: data production map; data quality; information product; information production map; information quality; information quality improvement.

Reference to this paper should be made as follows: Davidson, B., Lee, Y.W. and Wang, R. (2004) 'Developing data production maps: meeting patient discharge data submission requirements', *Int. J. Healthcare Technology and Management*, Vol. 6, No. 2, pp.223–240.

Biographical notes: Bruce Davidson is the Director of the Cedars-Sinai Resource and Outcomes Management Department. Dr. Davidson is responsible for guiding the development and implementation of initiatives designed to promote cost-effective, high quality medical care. He has more than 20 years

of experience in health services delivery and evaluation. He has particular expertise in health services research, organisational strategy, performance improvement, information management, medical treatment effectiveness, quality of care assessment, patient outcome analysis (clinical, economic, satisfaction, and functional status), and the design of information to support decision making by healthcare providers and consumers. He also teaches regularly at UCLA, USC, and MIT, being invited to guest lecture on all aspects of performance improvement and managing for quality. Dr. Davidson holds a PhD in health services research from the UCLA School of Public Health and was a Pew Health Policy Fellow at the RAND/UCLA Center for Health Policy Study. He also holds a Master's degree in public health from UCLA, and a Master's degree in urban planning from USC. He received his BS from MIT.

Yang W. Lee is an Assistant Professor in the College of Business Administration at Northeastern University. Dr. Lee's publications have appeared in leading journals such as *Communications of the ACM*, *Journal of Management Information Systems*, *Sloan Management Review*, *Information & Management*, and *IEEE Computer*. Her books include *Quality Information and Knowledge* (Prentice Hall, 1999), *Data Quality* (Kluwer Academic Publishers, 2000) and *Journey to Data Quality* (MIT Press, forthcoming). Her research interests include data quality, IT-mediated institutional learning, and systems integration. She was a Visiting Professor at MIT Sloan School of Management, where she taught e-systems integration and conducted research on data quality. She co-founded a firm, Cambridge Research Group, specialising in information quality management. She was also a Conference Co-Chair for the International Conference on Information Quality (ICIQ). She received her PhD from MIT.

Richard Wang is Director of the 'MIT Information Quality' (MITIQ) programme and Co-Director for the 'Total Data Quality Management Program' at the Massachusetts Institute of Technology. He had served as a Professor at MIT prior to heading the MITIQ programme and served on the faculty of Boston University and was a visiting Professor at the University of California, Berkeley. Wang has put the term *Information Quality* on the intellectual map with myriad publications and conferences. In 1996, Professor Wang organised the premier 'International Conference on Information Quality', which he has served as the general conference chair, and currently Chairman of the Board. He also co-founded the 'Workshop on Information Technologies and Systems' (WITS) in 1991. Wang's books on information quality include *Quality Information and Knowledge* (Prentice Hall, 1999), *Data Quality* (Kluwer Academic, 2001) and *Journey to Data Quality* (MIT Press, forthcoming). Dr. Wang received his PhD from MIT.

1 Introduction

A key lesson from recent research in information quality is that information[1] must be managed as a product (Kahn *et al*., 2002; Langley *et al*., 1996; Wang *et al*., 1998, 2003; Wang, 1998). To manage information as a product, the entire information manufacturing system must be understood and managed accordingly. The data production map has been conceptualised (Ballou *et al*., 1998; Lee *et al*., 2002; Pierce, 2001; Wang *et al*., 2001, 2003) as a critical tool to operationalise managing the entire information manufacturing system. To the best of our knowledge, this study is one of the first attempts to develop

and use the data production map in an actual organisational setting. We report how these concepts and models are applied in a major hospital, which we will refer to in this paper as the *Academic Medical Center*, or simply *the Center*. Specifically, we show how data production maps are developed to understand, analyse, and improve the quality of the patient-level data that must be submitted to the State Department of Health Services.

We refer to an information manufacturing system as the entire information system that produces information products (IP). The notion of information as product implicitly incorporates the concept of value-added in the sense that as data flows through the information manufacturing process, value is added. Further, it implies that the output of the information manufacturing system, namely, the information product, has value to a user of the information.

We define an Information Product (IP) as a collection of data element instances that meet the specified requirements. A data element is a basic unit that has meaning in the context of the operational environment. A data element could be, for example, an attribute in an entity or a form. Date of birth, social security number, and name are examples of data elements, while 1 March 1985, 026-345-6723, and John Doe are instances of each of the three data elements. An example of an information product can be an individual's birth certificate.

A key contribution of this paper is a longitudinal case study of seven years on how a healthcare organisation applied and adapted the concept of the data production map in the context of information quality and process improvement initiatives. The case study reports the comprehensive contexts involved in applying these concepts to a real-world problem in a healthcare institution.

This paper is structured as follows. Section 2 describes the background of the Medical Center and research methods used. Section 3 describes the process of developing data production maps in the Medical Center. Section 4 reports on how the data production maps enabled the improvement of data quality in the Medical Center. Finally, Section 5 concludes the paper with lessons learned and their implications.

2 Study site and research method

The Academic Medical Center is made up of an 875-bed hospital and affiliated physician groups, outpatient clinics, and post-acute rehabilitation services. It is one of the largest hospitals in the western United States, with an annual volume of about 50,000 inpatient admissions, including deliveries, and 150,000 outpatient and emergency room visits. As in most major hospital-centred healthcare delivery systems, the data derived from operational transaction systems (Admission, Discharge, and Transfer; Medical Records; Patient Billing; Laboratory; Pharmacy; etc.) feeds into strategic analyses and to external review agencies. This data has essentially been regarded as a byproduct of the provision of care to patients and, as such, of secondary importance to management in comparison to direct patient care services. Moreover, in many ways, the 'gold standard' means of data transmission in healthcare is essentially an illuminated manuscript in the form of the patient's medical record. Now, however, many strategically important initiatives, such as clinical quality and efficiency improvement, business development analyses, required regulatory and governmental reports, are seen to be at risk because the data used to monitor and support these organisational processes have been recognised to be incorrect or incomplete or otherwise faulty to some degree.

Not only could the unacceptable levels of data quality compromise the ability of such an organisation to operate successfully in a competitive managed care environment, but it could also place it at risk of being subject to audit by external agencies such as the Office of the Inspector General and the Joint Commission on Accreditation of Healthcare Organisations. Recognising that data is a valuable corporate asset, leadership concluded that the strategic use of data from diverse internal transaction system sources demands intelligent and disciplined management of data quality. Administrative leadership has been successful in introducing and nurturing a customer-oriented accountability system with a main focus on measurable continuous improvement in the last twelve years. Addressing these information quality issues was therefore framed in this Continuous Quality Improvement (CQI) context, and information quality goals have been included as part of the annual planning process since the beginning of the 1997–1998 fiscal year.

2.1 Research method

We used the longitudinal case study method (Eisenhardt; Yin, 1994) to analyse and report development of data production maps in the Academic Medical Center. This research reports the authors' investigations of the Center's attempts on developing data production maps in the context of process and information quality improvement. The general observation of the Center's initiatives on process and information management for seven years since 1996 set the environment for our focused investigation on data production maps in a context. During the investigation, we conducted at least three rounds of interviews with all stakeholders including data collectors, data custodians, and data consumers, who cover the entire scope of the information manufacturing system. Besides the annual interviews with the stakeholders, over 50 participants, we interviewed the key 12 informants regularly in developing the data production maps. We collected the Center's existing data flow maps, systems diagrams, and meeting memos on information and process management. We visited the site regularly, and participated and observed the meetings, and conducted e-mail and phone conferences to collect and verify data.

2.2 Information quality at the Medical Center

Administrative leadership for the Center's information quality initiative has been assigned to the Data Provider Group (DPG). The Data Provider Group was chartered as a multi-departmental working group by the CEO, during the latter part of the 1996–1997 fiscal year, to address ongoing discrepancies observed in analyses produced by different departments using different databases that supposedly contained the same information. The DPG is chaired by the Senior Vice Presidents of Medical Affairs and of Finance. The membership includes the Vice Presidents of Medical Affairs, Information Systems, Finance, and Contracting, as well as representatives from the following Departments: Information Systems, Resource & Outcomes Management, Health Information/Medical Records, Business Development, Patient Accounting, Cost Accounting, Budget, and Reimbursement.

Since March 1997, in an effort to create a more customer-oriented and efficient Information Systems (IS) function, the Senior Vice President for Medical Affairs (to whom the Vice President for Information Systems reports) had been restructuring that function and communicating regularly with the management group about these changes. These

changes included: a new oversight system for IS governance made up of multi-departmental committees of IS customers; a new, systematic approach to evaluating all Information System acquisitions; the systematic replacement of outdated hardware and software, including making personal computers more widely available and establishing a common e-mail and calendar-scheduling system; the construction of an institutional intranet for the online distribution of various kinds of information; the construction of a new ORACLE data warehouse, and the implementation of a new generic data query/report-writing tool for data analysis.

The development and implementation of a basic set of online management reports was established as a DPG goal for the 1997–1998 period. The work plan adopted by the DPG included responsibilities for six subgroups, one of which focused specifically on data quality issues. The measure of success for the on-line management reports project was established as 'an increase in the level of satisfaction with the management reports' throughout the health system. The DPG decided to adapt the MIT Total Data Quality Management (TDQM) Information Quality Survey to measure information quality, and moved to establish a baseline assessment for all 'data consumers' throughout the organisation with the first administration of the modified Information Quality Survey in December 1998. A follow-up assessment was administered in May 1999. The adaptation of the MIT TDQM Information Quality Survey for this purpose, and the results of the baseline and follow-up survey administrations are described elsewhere (Chun and Davidson, 1999, 2002).

Upon the completion of the online management reports project, the data quality subgroup was chartered by the DPG to continue its work more broadly, and hence became the Center's Data Quality Management Working Group (DQMWG). The DQMWG was charged to establish a plan to systematically identify, monitor, and address data quality problems in order to assure that the Center's data is 'fit for use' for its decision-making requirements. The DQMWG is made up of a combination of representatives from departments that are data providers, data custodians, and data consumers. Its basic functions are to establish routine processes to identify, track, and resolve data quality problems (Table 1).

Thus, the work described in this paper occurred in an institutional context in which the need for active management of information products was recognised, and in which administrative leadership had been established for that purpose. This work, therefore, builds on several years of efforts to gradually introduce these concepts, and develop data production maps to improve information quality.

2.3 Patient discharge data submission requirements

As in many states, the Department of Health Services in this State requires every acute-care hospital to submit data describing each patient discharge. In this State, 31 specific data elements are required to be submitted for each patient discharge every six months (Table 2). The State establishes specifications for the quality of each of the 31 required data elements, and provides those specifications as public information. Data that does not meet the specifications, which are set by the State, is rejected and the offending hospital is required to correct the data and resubmit it within a specific time frame.

Table 1 Structure of data quality management working group

Data quality management working group membership	
Department/Area	Data Quality Role(s)
Resource and Outcomes Management	Chair, User
Health Information	Producer, User
Information Systems	Custodian
Patient Financial Services	Producer
Cost Accounting	Producer, User
Performance Improvement	User
Business Development	User
Finance	Producer, User
Materials Management	Producer, User
Operating Room Services	Producer, User
Medical Network Services	Producer, Custodian, User
Managed Care Contracting	User
Pathology and Laboratory Services	Producer, User

Table 2 Required data elements

Data elements required by the State	
1. Patient's Type of Care	17. Other Diagnoses at Admission
2. Hospital ID	18. Principal Procedure Code
3. Date of Birth	19. Principal Procedure Date
4. Sex	20. Other Procedure Codes
5. Race/Ethnicity	21. Other Procedure Dates
6. Race/Race	22. Principal E-Code
7. Zip Code	23. Other E-Codes
8. Admission Date	24. Patient's Social Security Number
9. Source of Admission/Site	25. Disposition of Patient
10. Source of Admission/Licensure	26. Total Charges
11. Source of Admission/Route	27. Abstract Record Number (Optional)
12. Type of Admission	28. DNR Order
13. Discharge Date	29. Expected Source of Payment/Payer Category
14. Principal Diagnosis	30. Expected Source of Payment/Type of Coverage
15. Principal Diagnosis at Admission	31. Expected Source of Payment/Plan Code Number
16. Other Diagnoses	

About eight years ago, changes in legislation resulted in the imposition of a daily monetary fine for each day that a facility remains in non-compliance with established specifications beyond the deadlines for correction. There are, then, two types of immediate costs associated with poor quality data submitted to the State: the internal cost of the rework

needed to make the required corrections; and the externally imposed costs associated with the monetary fines. But there are additional costs as well, since these 31 data elements are at the heart of almost all internal efforts to analyse patient care for management and strategic purposes. These data quality problems could compromise institutional decision making based on such analyses.

Historically, Academic Medical Center's data submissions to the State always required corrections for one or more of the data elements. Since the level of rework was not observed to diminish over time, systematic improvements in the production of the data submission as an information product were not occurring despite the corrections of the immediate problems for each submission. It was estimated that at least 1500 to 2000 person-hours may have been required annually to make the needed corrections. In light of this continuing 'wasted' effort in rework, made more onerous due to the imposition of monetary fines by the State, an effort was initiated to systematically evaluate and improve the quality of the data submitted to the State, under the auspices of the DQMWG. Thus, the Patient Discharge Data Submission Improvement Project was launched.

3 The patient discharge data submission improvement project

The Patient Discharge Data Submission Improvement Project (the Project) was framed initially as a three-month effort to study how to improve the quality of the patient-level data submitted to the State. The Project adapted the classical Performance Improvement approach that is currently in use for all our clinical and administrative performance improvement activities. Its model for improvement is based upon the general notion that 'all improvements are changes, but not all changes are improvements'. In order to know whether a change is an improvement, three principles must be addressed: a goal must be clearly articulated that unambiguously describes what one is trying to accomplish; a measure must be established that objectively allows a determination that the change has had an effect in the appropriate direction; and a series of changes that are hypothesised to be improvements are tested and evaluated using 'Plan – Do – Study – Act' (PDSA) cycles. Those changes that result in improvements are then implemented on an ongoing basis (Carey and Llyod, 1995; Langley *et al*., 1996). In the early stages of a performance improvement project, the initial PDSA improvement cycles focused on evaluating the work process to be improved. Thus, the objective of the Project was to clearly delineate and evaluate the work process, which is, the information manufacturing system used to produce the information product, the required patient discharge data set that needed to be improved.

A three-month Project period was set from April through June 2000. The deliverables for the project were defined as:

- to prepare 'data production flow maps' for all required data elements
- to identify where breakdowns occur in the production process for each data element
- to recommend improvements to prevent breakdowns from occurring in the future.

It was felt that the benefits would go beyond the avoidance of monetary penalties and the elimination of costly rework needed to correct the errors, contributing as well to overall strategic decision making and compliance with other external oversight initiatives by the

federal government and the independent Joint Commission on Accreditation of Healthcare Organisations. The Project itself, then, can be seen to follow the classical Performance Improvement approach from an initial analytic phase in which the work process itself is defined and described, and viable measures for evaluating improvements are proposed, through to the point of producing a list of changes hypothesised to be improvements, which could subsequently be tested following the project period and implemented if successful.

It was clear that the development of data flow diagrams to describe the creation of the required submission dataset as an information product, would be a critical contribution to achieving lasting improvements in the quality of the data. This observation came from the recognition that the attempts to 'just fix' the data quality problems, that had resulted in the rejection of the Center's data submission by the State, were not resulting in the presence of fewer data quality problems from submission to submission. There are multiple information systems involved in the creation of the required submission dataset, and it was apparent that there were various causes of poor data quality in each information system involved. These causes included both inaccurate data entry as well as unanticipated data transformations in interface programmes, which were multiplied by the number of information systems involved to create a complex environment in which no one participant had any ability to understand and resolve any particular data quality problem.

3.1 Tracking multiple sources of information product

Prior to this Project, there was no commonly accepted and unambiguously articulated vision of the overall work process that yielded this information product as its output. The initial challenge of the Project was to gather sufficient background information to unambiguously describe the information product 'supply chain' from source to final output. This effort represents the first PDSA cycles of the improvement project, in which drafts of 'information product flow maps' were produced and reviewed for revision and correction until a commonly accepted version was finally produced. The information, in this case, was gathered through a series of repeated open-ended interviews with representatives of each of the departments involved.

The 31 data elements required for submission to the State were evaluated to identify the information system from which they each originated. In some instances, data elements required for submission were not available in the required format from source systems, and those elements had to be constructed from other precursors that were available. In those cases, the precursor data elements were evaluated to identify their source systems. It was determined that all of the required data elements originated in three source systems: the patient management system, the medical records system, and the patient accounting system. On this basis, representatives of the departments responsible for producing the required data elements were identified and interviewed to capture and record the unique knowledge each could contribute about how each of the data elements was produced and managed in its source system, and the issues and concerns from each participant's point of view.

Interviewees were asked to describe the information manufacturing system, as they understood it, for the data elements produced by the information systems in their areas,

and interviewers recorded their responses in diagrams and texts. The interview process yielded a collection of partial descriptions which could then be pieced together to produce a complete description of the overall process. Each of the disparate departments involved was able to articulate only their own roles (usually in an incomplete fashion), and each believed that errors were being introduced into the production process by other departments.

Through these interviews, additional information was also gathered regarding the intermediate transfer, processing, and storage of these data elements after they were extracted from the source systems, and before they were consolidated into the final information product. Departments and individuals responsible for these intermediate steps in the information product supply chain were thus also identified and interviewed in a secondary round of interviews. In a similar fashion, information was gathered about departments and individuals responsible for the final steps of data consolidation and the production of output files for submission, and a third round of interviews was conducted with them. In some cases, the same departments and individuals were responsible for multiple steps in the production process, and not necessarily sequentially. Multiple interviews were conducted with most departmental representatives in order to be certain that the information was being recorded correctly, and in order to answer additional questions that arose as more information was collected.

Finally, in addition to describing the current process by which the required submission was produced, it was also necessary to gather sufficient information that would allow assessment of the potential impact of new information system developments already underway. Two related developments were planned that would have a big impact on this production process. First, within 18 months two of the three source systems currently used (the patient management system and the patient accounting system) would be replaced by a new combined information system currently in development. Second, the replacement of one of those systems, the patient accounting system, would result in the termination of the vendor relationship through which the submission to the State had historically been handled. This latter development would mean that the submission dataset would need to be produced in-house, rather than simply managing a third party to accomplish this step of the production process.

During all these interviews, any available background documentation was also identified and collected. This background documentation was analysed as well, and used to confirm the information gathered in the interviews. The background documentation consisted of data entry protocols for the three source systems, data definitions from the State, edit criteria from the State, tape layout parameters for the required submission, and the State's rejection reports. The rejection reports were used to quantify and summarise the error history.

3.2 Developing information product maps

Four Information Product Maps were developed to represent each of the three source systems, as well as the anticipated information systems transitions noted earlier. These maps reflect a cross-functional collaboration and include a multi-level analysis of the data flow on the basis of departments, physical location, information system, and business process, where applicable. The maps of each of the three source systems show the flow of

data elements as they originate in the source system and as they move through various interfaces, ultimately ending up in the submission dataset on the required system medium. The fourth map depicts a migration plan that would accommodate the anticipated information systems transitions marked in colour (dark shades in black and white printout) to highlight each of three phases in the migration plan.

We now turn to describe each of the Information Product Maps that were developed. The first map (Figure 1) shows the flow of data from the medical records information system. The second map (Figure 2) shows the flow of data from the patient management information system. The third map (Figure 3) shows the flow of data from the patient accounting information system. And the fourth map (Figure 4) shows the migration plan to accommodate the transition to a new, combined information system that would include both patient management and patient accounting data, as well as the use of the in-house Data Warehouse to produce the submission dataset rather than through the vendor.

Figure 1 Data production map I, medical records system data elements

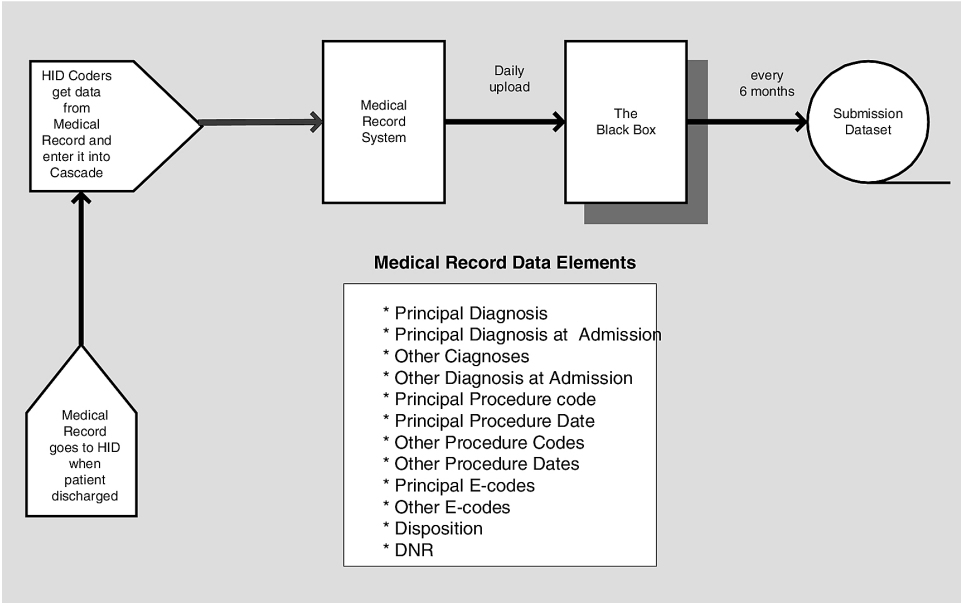


Figure 2 Data production map II, patient management system data elements

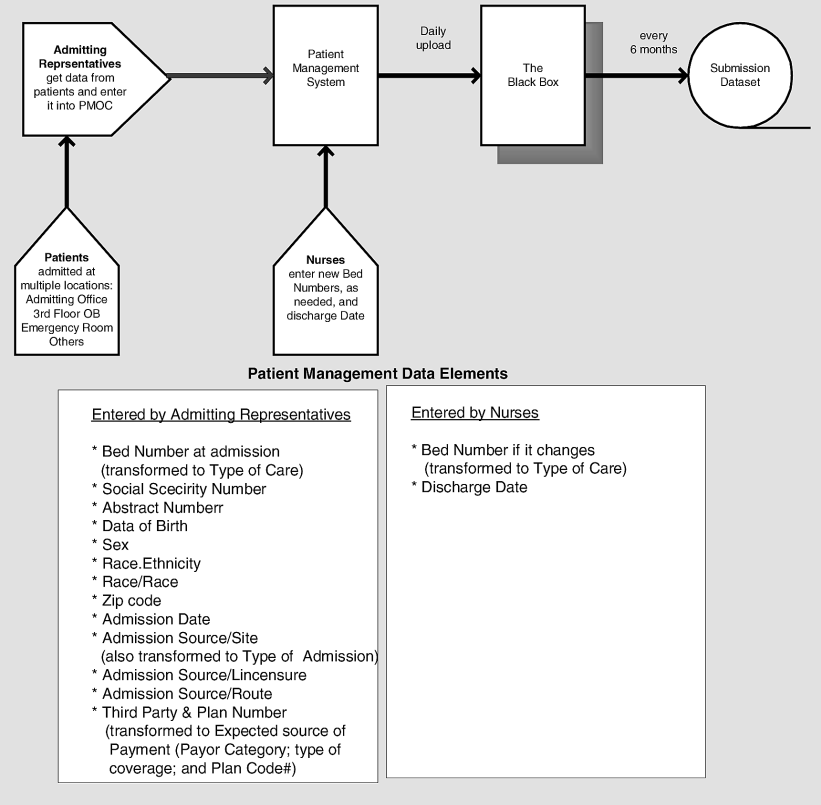


Figure 3 Data production map III, patient accounting system data elements

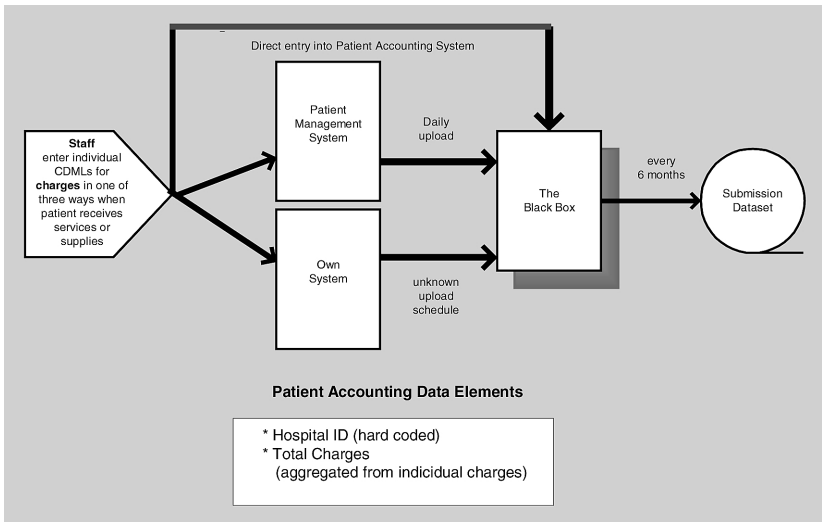
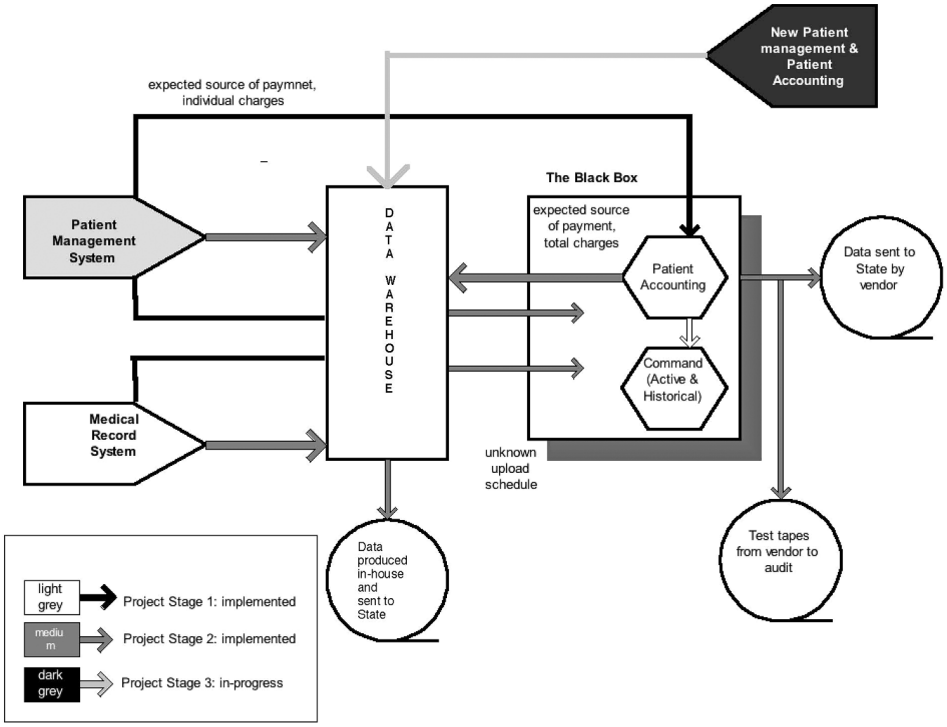


Figure 4 Data production map IV, overall data production map



In each of these maps, it should be noted that the current actual process by which the submission dataset is aggregated and placed on the required medium by the third-party vendor is labelled as the 'Black Box'. This is to highlight the undocumented complexity and the ambiguity of the vendor's internal computing environment. If what went into the 'Black Box' was known and what came out of it was known, then that should be sufficient to know whether data quality problems were being created within the Center or by the vendor. Moreover, it was not thought to be a worthwhile exercise to evaluate the vendor's computing environment further, since the migration plan would result in an elimination of the vendor relationship altogether.

3.3 Four information production maps developed

The first information product map (Figure 1) shows the 'supply chain' for the 12 required data elements originating in the medical records information system. These data elements are listed in Figure 1. The map shows that all these data elements are derived from the patient's medical record that are sent to the Health Information/Medical Records Department (HID) when the patient is discharged from the hospital. The Coders abstract the data elements when they receive the patient's medical record and they enter them into the medical record information system. Through a daily upload, this data is automatically sent into the 'Black Box' where it is used for various transactional purposes. Every six months, these data elements are aggregated with the other required data elements and placed on a tape or cartridge for submission to the State.

The second information production map shows the 'supply chain' for the 15 data elements originating in the patient management information system (Figure 2). These elements are listed in the figure. Four of these source system data elements undergo simple or complex transformations, as indicated, in order to create other data elements required for submission to the State. Fourteen of these 15 data elements are entered by representatives of the Admitting Department by interviewing patients at the time of admission in a variety of locations. The admitting representatives get the data from the patients, or their family members, and enter the data into the patient management information system. Also, one data element, bed number, can be changed during the patient's stay by nurses if appropriate. One additional data element, discharge date, is entered only by nurses. Through a daily upload, this data is automatically sent into the 'Black Box' where it is used for various transactional purposes and stored. Every six months, these data elements are also aggregated with the other required data elements and placed on a tape or cartridge for submission to the State.

The third information production map shows the 'supply chain' for the two data elements originating in the patient accounting information system (Figure 3). These elements are listed in the figure. One of these data elements, the hospital identification number, is hard-coded and never varies. The other, total charges, is aggregated from individual charges within the patient accounting system. The individual charges represent all the billable hospital services or supplies provided to the patient, and they are entered in the form of Charge Description Master List (CDML) codes by staff as those services or supplies are provided. Three different methods of entering these codes are used. In some instances, CDML codes are entered directly into the patient accounting system within the 'Black Box'. In other instances, CDML codes are entered into the patient management system and automatically uploaded daily into the 'Black Box'. Finally, some departments enter CDML codes into their own transactional systems (e.g. Operating Room, Pharmacy, Laboratory) that then have their own unique upload schedule by which these codes are transferred into the 'Black Box'. The individual charges are used for various transactional purposes. Every six months, these data elements are aggregated with the other required data elements and placed on a tape or cartridge for submission to the State.

The fourth information production map shows an overview of the current overall data production map and, through colour-coding, illustrates a migration plan (Figure 4). The migration plan both anticipates the new combined patient management/patient accounting system, and also anticipates the need to be able to produce the required output dataset for submission directly to the State, rather than doing so through the third-party vendor. The light grey shades with dark grey arrows show the status of the 'supply chain' at the first stage of the project. Data elements originating in each source system are described generically on this map. On this map, some detail is shown within the 'Black Box', primarily to permit interim plans for data capture to be illustrated.

The interim plan, the second stage of the project, is illustrated by medium grey shade symbols and arrows, and it covers the time period up until the point in time when the new combined patient management/patient accounting information system is fully implemented. In this interim plan, the 'supply chain' has been modified so that all required data elements can be captured in the Center's data warehouse from the three existing source systems. This would be necessary so that the required submission dataset can be produced in-house and sent directly to the State on the required medium. The ultimate implementation, stage three, of the combined patient management and patient accounting information system is

indicated in dark grey shade with light grey arrows. This shows the replacement of the old separate patient management and patient accounting systems, and the flow of needed data elements into the data warehouse from the new combined system. In this scenario, the medical records system would also continue to feed data elements into the data warehouse.

3.4 Improvements of data production process

As a direct result of these efforts to develop the information production maps described above, specific and systematic improvements in the production process for the required submission dataset were implemented, and future improvements were planned. The first specific improvement involved obtaining the audit criteria handbook and the audit programme source codes in COBOL from the State. These were then used to develop audit programmes in SAS for key problematic data elements. A 'test tape' was then ordered from the vendor in advance of the required submission, so that the submission dataset could be audited prior to the submission to the State. The results of the audit were shared with the appropriate 'owner' departments and they were able to make some corrections prior to submission in order to reduce the number of data quality problems identified by the State after receipt of the dataset. While this may seem like an obvious strategy, it had never before been conceptualised and implemented. It was only as a result of explicitly analysing the information manufacturing system that sufficient insight was gained to implement this strategy.

The second specific improvement involved identifying several constructed data elements that could be modified in order to simplify the transformations needed to produce certain required data elements from the data elements available in the source systems. Here again, conceptualising this improvement had not been possible prior to explicitly analysing the information manufacturing process involved. A third specific improvement was being able to conceptualise useful suggestions to the team building the new combined patient management and patient accounting information system. The development of the information production maps resulted in a deeper understanding of the information manufacturing process and it became possible to develop recommended specifications to ensure that all data elements required by the State were built into this new system. These recommendations were both for data entry standards as well as routine audit procedures.

Finally, the Center has continued to initiate improvement cycles to advance its abilities to manage this information manufacturing system more effectively. All these advances are based on the foundation established with the information production maps, which provide a common understanding of the complex processes involved and a common language to discuss them. Each required data submission is now proactively managed, and although new unanticipated issues continue to arise, the Center is much better at understanding them and addressing them. A pilot has been initiated to develop the capability for in-house production of the submission dataset so as to be prepared when the relationship with the current third-party vendor is discontinued.

4 Lessons learned: data quality improvement at the Medical Center

The opportunity to provide this epilogue arose as the case study was being prepared for publication, in order to describe how this improvement project at the Center has contributed to reaching various milestones in its journey along the road to improve data quality.

As intended, Academic Medical Center has been quite successful in finding and fixing the errors in the required data prior to final submission to the State, as a result of the project. Consequently, the data submissions by the Center have been accepted by the State upon the first official submission for every six-month submission. In addition, each of these successful submissions was made prior to the due date. In this respect, the improvement project helped the Center achieve the fundamental objectives: to meet the State's accuracy, completeness, and timeliness requirements of the data submitted.

However, the broader objective of providing input to the team, building the new combined patient management/patient accounting information/medical record coding system, has encountered many obstacles along the way. The importance of this broader objective should not be underestimated, since it alone offers the Center the capability of preventing the creation of the data quality errors, which require fixing. The implementation of the new information system is still in progress, more than four years into what was originally projected as an 18-month development project. Its development has faced more difficulties than originally anticipated, and its implementation has been repeatedly rescheduled. Its development challenges have resulted in many revisions to specifications. As a result, motivating the actual incorporation of input related to designing the new information system to prevent the creation of the errors that the Center now finds and fixes has turned out to require constant vigilance, continued advocacy, and learning.

This new information system will eventually become the single source for all the required patient discharge data elements, allowing the institution to exercise complete control over every aspect of producing the submission dataset. This is in contrast to the ongoing environment as was described above, in which the use of many proprietary vendor-driven applications, as well as archaic and poorly documented programming codes underlying the homegrown systems and interfaces prevented the exercise of any control over most upstream sources of errors. The Center's ability to recognise these shortcomings, and to plan explicitly to avoid them, does represent another successful use of the output of the improvement project.

Consequently, a variety of methods by which the new system could help to prevent the creation of these errors can thus be articulated and illustrated, and this contribution simultaneously supports the implementation of refined business processes that are themselves more efficient. These are profound improvements that would not be possible to imagine without the exercise of articulating and mapping the data production maps and understanding the information manufacturing system involved. Thus, carrying out the improvement project itself helped the Center to create new organisational knowledge that is being used more broadly than could have been anticipated at the time.

First, the new information system can help eliminate the need for wasteful rework by allowing the specification of the data elements to be collected as well as the permissible values, and populating them. By specifying acceptable values for the required data elements, no transformations prior to submission would be needed, and errors now introduced in these transformations will be eliminated. However, the new information system also provides the capability of directly performing any transformations that might be needed.

Second, built-in data quality management capabilities, such as routine audit and feedback of discrepant or incomplete data values for real-time correction by upstream staff responsible for data creation, will be included. This will reduce the creation of unacceptable data values at the source, and thus prevent the population of downstream repositories with erroneous or incomplete data as well. Since this new information system will be one unified system, it also eliminates the need for programmed interfaces to move data from various source systems into a single staging repository prior to making it available for submission or retrospective analysis. It therefore eliminates the potential for inadvertent transformation of data as an unintended effect of programming approaches or errors in programming interfaces. Each of these methods can clearly help prevent the creation of the errors that the Center now finds and fixes before data is submitted to the State.

But these methods also can be applied across the entire institution to a variety of other information manufacturing systems to bring profound benefits through the elimination of wasteful rework. Thus, the improvement project did provide an example that has been emulated for several additional performance improvement efforts focused on other information management activities. The development of information production maps to illuminate several other intractable information management tasks has been particularly useful. Ultimately, this work has helped the Center to better meet many of the hospital-accreditation standards in the Management of Information to which all hospitals are held accountable by the 'Joint Commission on Accreditation of Healthcare Organisations' (JCAHO).

A close reading of the Management of Information standards shows that JCAHO intends for hospitals to engage in explicit information management planning and performance improvement activities, such as those described here. Listed in Table 3 are the specific Management of Information (IM) standards that apply to Information Management Planning activities. Additional detail on these standards is available on the JCAHO website at www.jcaho.org.

Table 3 Accreditation standards on information management

JCAHO Standards Related to Information Management Planning	
IM.1	The hospital plans and designs information management processes to meet internal and external information needs.
IM.2	Confidentiality, security, and integrity of data and information are maintained.
IM.2.1	Records and information are protected against loss, destruction, tampering, and unauthorised access or use.
IM.3	Uniform data definitions and data capture methods are used whenever possible.
IM.4	The necessary expertise and tools are available for the analysis and transformation of data into information.
IM.5	Transmission of data and information is timely and accurate.
IM.5.1	The format and methods for disseminating data and information are standardised, whenever possible.
IM.6	Adequate integration and interpretation capabilities are provided.

Besides technical solutions, management strategies can have a profound impact on eventual success (Davidson, 2001). Although the benefits of incorporating the lessons learned seem obvious, it is not enough to advocate vigorously for a logical argument at a single point in time. The factors that influence attention and priorities at every level of the organisation are varied and shifting, and so the approach to data quality must be flexible. With this toolkit of techniques and the accumulation of examples from which to learn, the Center is confident that it can eventually learn how to overcome all obstacles on this journey and achieve all the benefits that high quality data can provide.

5 Concluding remarks

We have presented a case study of how the Academic Medical Center developed its information production maps to gain an understanding of the complex information manufacturing system that creates, processes, and delivers patient-level data that must be submitted to the State. Furthermore, the Center has traced the causes of the problems that were preventing successful production of the required information. As a result, currently, the CEO no longer receives letters of non-compliance from the State. Although the penalty for non-compliance is small in this case, $100/day for non-compliance, the rework and frustration was costly. As a prestigious Center, that takes pride in being successful and exceeding expectations, the ability to submit data complying with the criteria established by the State is a critical event. This is particularly true given that the problem had persisted for many years before the MIT Total Data Quality Management methods were introduced and information production maps were developed.

The key lesson learned from this case study is that information product maps, when properly developed and supported by the organisational senior management group, can be effectively applied to solve complex, cross-functional problems that appear to be intractable. Equally important is that this case traces a real-world problem over a period of seven years to a successful conclusion. It is also noteworthy that this longitudinal case study also observed improved overall data quality in the Center, which is unusual compared to most successful projects that are short lived as projects ends.

Managing information as a product has generated much interest since its first introduction with four principles (Wang *et al*., 1992). Current research in this area focuses on formal mathematical properties of information product maps (IPMAP), such as tractability from one node of the map to another, standards for IPMAP notations and software development, and linkage between IPMAP to other aspects of the information manufacturing system, such as how input data can be processed and delivered as an information product to the consumer in such a way that meets or exceeds their expectations.

References

Ballou, D.P., Wang, R.Y., Pazer, H. and Tayi, G.K. (1998) 'Modeling information manufacturing systems to determine information product quality', *Management Science*, Vol. 44, No. 4, pp.462–484.

Carey, R. and Llyod, R. (1995) *Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control Applications*, New York: ASQ Quality Press.

Chun, A.T. and Davidson, B. (1999) 'Implementing the information quality survey: a case study at Cedars-Sinai health system', in *Conference on Information Quality*, Cambridge, MA, pp.266–284.

Chun, A.T. and Davidson, B. (2002) 'Measurement for IQ improvement: translating results from the IQ survey into management objectives', in *The Seventh International Conference on Information Quality*. Cambridge, MA, pp.31–35.

Davidson, B. (2001) 'Storytelling as a management tool: institutionalising the data quality function at Cedars-Sinai Medical Center', *Proceedings of the Sixth International Conference on Data Quality*, Cambridge, MA, pp.3–11.

Eisenhardt, K. 'Building theories from case study research', *Academy of Management Review*, Vol. 14, pp.532–550.

Kahn, B.K., Strong, D.M. and Wang, R.Y. (2002) 'Information quality benchmarks: product and service performance', *Communications of the ACM*, pp.184–192.

Langley, G.J., Nolan, K.M., Nolan, T.W., Norman, C.L. and Provost, L.P. (1996) *The Improvement Guide: a Practical Approach to Enhancing Organisational Performance*, San Francisco: Jossey-Baas.

Lee, Y., Allen, T. and Wang, R. (2001) 'Information products for remanufacturing: tracing the repair of an aircraft fuel-pump', *Sixth International Conference on Information Quality*, Cambridge, MA, pp.77–82.

Lee, Y., Strong, D., Kahn, B. and Wang, R. (2002) 'AIMQ: a methodology for information quality management', *Information & Management*, Vol. 40, No. 2, pp.133–146.

Pierce, E. (2001) 'Using control matrices to evaluate information production maps', *The Sixth International Conference on Information Quality*, Massachusetts Institute of Technology, pp.412–423.

Wang, R., Allen, T., Harris, W. and Madnick, S. (2003) 'An information product approach for total information awareness', *2003 IEEE Aerospace Conference*, Big Sky, Montana.

Wang, R., Ziad, M. and Lee, Y. (2001) *Data Quality*, Norwell, Massachusetts: Kluwer Academic Publishers.

Wang, R.Y. (1998) 'A product perspective on total data quality management', *Communications of the ACM*, Vol. 41, No. 2, pp.58–65.

Wang, R.Y., Lee, Y.L., Pipino, L. and Strong, D.M. (1998) 'Manage your information as a product', *Sloan Management Review*, Vol. 39, No. 4, pp.95–105.

Yin, R. (1994) *Case Study Research*, Thousand Oaks, CA: Sage Publication.

Notes

[1] We use the term 'data' and 'information' interchangeably in this paper, unless specified otherwise. Thus, information product and data product are used interchangeably, unless specified otherwise. The conventional wisdom typically suggests that information is data that has been processed, and that data is the raw material that feeds into a processing system (information manufacturing system in our paper). The information output from one processing system often feeds into another system, and therefore the information can be data as well. Given this and other reasons, we decided not to artificially differentiate data and information in this paper, unless they are used in a specific context.