Using Control Matrices to Evaluate Information Production Maps

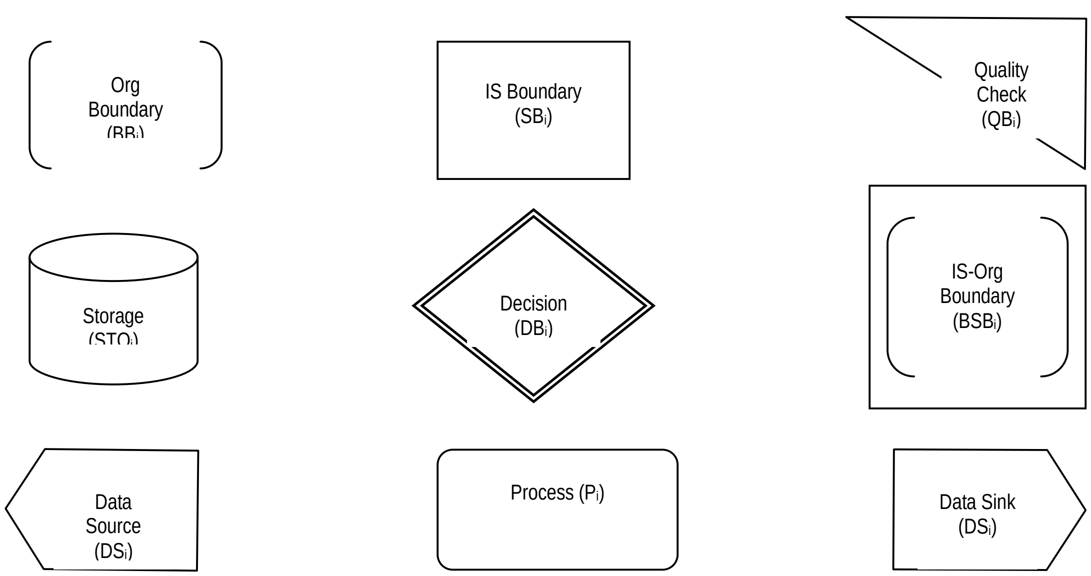
Elizabeth M. Pierce
Eberly College of Business
Indiana University of Pennsylvania
Indiana, PA 15705 USA
empierce@grove.iup.edu

Abstract: An information production map (IP-MAP) is a graphical model designed to help people to comprehend, evaluate, and describe how a data product such as an invoice, customer order, or prescription is assembled. In this paper, an information product control matrix is introduced in conjunction with the IP-Map to help data administrators assess the overall quality of their information products.

1. Information Production Maps

Information Production Maps (IP-Maps) are based on the idea that information outputs produced by information systems are analogous to products produced by manufacturing systems. Just as the quality of products produced in a manufacturing systems can be analyzed and tracked, the same can be done for information products. IP-maps are a means by which one can model the creation of the information product from its original raw data sources through the various processing, quality checking, and storage stages to its final form. Ballou et al. (1998) used this approach to model the information production process using symbolic blocks similar to those used in data flow diagrams. To distinguish this model, which emphasizes terminology and parameters based on manufacturing concepts as well as introducing the data quality construct, Ballou et al. (1998) referred to these diagrams as information manufacturing models. Shankaranarayanan et al. (2000) further enhanced the work done by Ballou et al. (1998) by adding three new blocks to model decision points, IS boundary points, and organization boundary points. Shankaranarayanan et al. (2000) also added metadata describing the department/role, location, business process, data composition, and base system to each of the blocks so that the IP-Map would contain relevant descriptive information that would allow the data administrator to more easily retrieve details about each of the IP-Map constructs.

Figure 1: List of IP-Map Constructs



2. Information Product Control Matrices

An information product such as a bill or prescription is prone to a variety of errors such as missing fields, incorrect information, or improper formatting. These errors may come from either the data sources or be introduced through the data process itself. Correction controls in the form of manual inspection or clean up programs may prevent, detect, or correct these errors, but not necessarily with 100% results. Although Information Production Maps give a good overall picture of the process by which data is transformed into the final information product, these diagrams do not highlight data quality parameters that describe the sources of error or the reliability rate of the various process components in a way that easily facilitates the assessment of the overall data quality of the information product.

To aid the data administrator in evaluating the information contained in the IP-Map and the resulting ramifications on the quality of the information product, this paper will borrow on ideas that IT auditors have used since the 1970's. Control matrices are a concise way to link data errors and irregularities to the quality controls that should detect and correct these data problems during the information manufacturing process. The columns of the matrix show sources of data errors or irregularities that occurred during the information manufacturing process. From a data administrator's perspective, the columns of the matrix list the data quality problems that can afflict the information product and identify the location in the IP-MAP model where the data administrator believes these data quality problems originate.

The rows of the matrix are the IP-Map constructs such as the quality checks or corrective processes that were exercised during the information manufacturing process to prevent, detect, or correct these data errors or irregularities. The elements of the matrix are some rating of the effectiveness of the IP-Map construct at reducing the level of data errors or irregularities. These ratings could take several forms:

- ! o*r X* - A quality check exists to prevent certain error(s) from appearing in the information product. In this case, the data administrator has examined the information production process and has identified that there is a corrective or detective process in place that should prevent that type of error from appearing in the final information product. Notice the ! or X provides the lowest level of assessment information since this only indicates a quality check is present, and does not address how well the quality check performs its function.
- *Category* - A quality check exists in the IP-MAP to prevent certain error(s) from appearing in the final information product and the data administrator is able to describe its effectiveness at error prevention, detection and/or correction as Low, Moderate, or High. This categorical assessment provides more information than a simple ! or X since it captures the data administrator's belief as to how reliably the quality check performs its function.
- *Number* - A quality check exists in the IP-MAP to prevent certain error(s) from appearing in the final information product and the data administrator is able to describe its performance as 95% effective at removing the data irregularity. To obtain a precise numerical assessment of the quality check's effectiveness, the data administrator would need to devise a test in order to evaluate the effectiveness of the IP-MAP control. For example, the data administrator may create several test data sets seeded with known errors. The quality check is then applied to the test data set. If the quality check is able to correct on average 95% of the known errors then the quality check can be considered to be 95% effective at preventing those types of data irregularities from appearing in the information product.
- *Formula* - A quality check exists in the IP-MAP to prevent certain errors(s) from appearing in the final information product; however, its reliability rate depends on a relationship between itself and some other variables. Under this scenario, the data administrator has obtained through process experimentation an understanding of how the reliability of a quality check may fluctuate and is able to describe that fluctuation through a mathematical function. For example, clerks who take orders over the phone may be less reliable in checking the data quality of the orders as the day wears on and this behavior can be modeled using a function incorporating the time of day.

Note that not every IP-MAP construct will be included in the Information Product Control Matrix. Only those quality checks or corrective processes that impact the data quality of the information product are included. While the IP-MAP is designed to model the overall process and is an important first step in understanding how the various components of the information production fit together, the IP-Control Matrix is designed to focus only on those parts of the IP-MAP that prevent, detect, or correct data irregularities in the final information product. It is also important to note that not every quality check will detect every type of error. It is quite possible that multiple quality checks are employed during the information manufacturing process, each one designed to catch different types of data problems.

Once the information production control matrix is complete, the data administrator examines each data error columns of the matrix in order to weigh up the effect of the various data quality controls and to determine whether the quality of the information product is at an acceptable level. Acceptable levels of quality will depend on the organization's commitment to data quality as well as to the cost and benefits of maintaining the information product at a given quality level.

In addition, cost and frequency information can be added to the Information Product Control matrix to estimate the dollar impact of an unreliable information product. To help illustrate the assessment process, Table 1 shows a sample information product control matrix. For simplicity, it is assumed that each transaction corresponds to a sales order (information product).

Table 1: Sample Information Product Control Matrix

	Information Product					
	Source of Data Errors or Irregularities That Occurred					
	Duplicate data in Component Data produced during Process 1	Data became obsolete during Storage	Typos from Data Source 1	Missing data from Data Source 1	Bad data from Data Source 1	Wrong format used in Process 2 to produce Component Data
Estimated Frequency of Error	2% of transactions	3% of transactions per month	5% of transactions	10% of transactions	6% of transactions	4% of transactions
Estimated Cost of Data Error per Information Product	$1	$1	$2	$5	$3	$4
Reliability Ratings of IP-Map constructs						
Quality Check 1	98% of transactions					
Corrective Process 1		90% of transactions per month				
Quality Check 3			85% of transactions	95% of transactions	88% of transactions	
Quality Check 4						97% of transactions
Overall Quality = Error Rate x (1 - Reliability Rate of IP-Map Construct)	.04% of transactions lead to IP's that are duplicates	.3% of data in storage lead to IP's that contain obsolete data	,75% of transactions lead to IP's that have typos	.5% of transactions lead to IP's that have missing data	.72% of transactions lead to IP's that have wrong data.	.12% of transactions lead to IP's that have the wrong format

For each column (i.e. data irregularity), the data administrator assesses how many errors would still show up in the information product after the quality checks have been performed. His assessment depends on his understanding of the information production process as described in the IP-MAP. To get an overall assessment of the quality of the information product, the data administrator must then combine the individual data irregularities rates into an overall rating. In the simplest case where the error rates are independent and in the same units, the data administrator can apply basic probability rules to determine that the probability that a given information product is free of any defects:

$\prod$(1-Error Rate$_i$) = (1 - .04%)(1 - .3%)(1 - .75%)(1 - .5%)(1 - .72%)(1 - .12%) = 97.6% of IP's are correct.

Expected Cost Per 1,000 IP's = Σ Cost$_i$ * 1000 * Error Rate$_i$ = $69.80 per 1,000 IP's.

In the case where numerical assessments of reliability were not obtained, the data administrator would need to subjectively combine the categorical or ! / *X* ratings to get an overall feel of the quality level of the information products. In addition, if the error rates are expressed in terms of functions, a spreadsheet or simulation program may be needed to gauge the overall reliability of the information product. In particular, where probability density functions are involved, the overall quality of the information product may be better expressed as a graph rather than a single number such as an average.

3. An Illustration of the Steps Used to Construct IP-Map Control Matrix

This section describes the steps a data administrator would follow in order to assess the reliability of an information product using an IP-Map and Information Product Control Matrix. For this simple example, a fictitious scenario is set up based on the author's experience in studying the operations of alumni affairs. For the purposes of this fictional case, the information product is a mailing label. A school called Big State University uses these mailing labels to send out publications to its alumni. Incorrect or out-of-date mailing labels are a problem for the university. Undeliverable mail costs the school money in terms of unrecoverable printing and mailing costs, as well as potential lost donation revenue from missing or disgruntled alumni.

After the end of each academic year, data (including current address information) about graduating seniors are taken from the Big State University's active student database and transferred to the Alumni database. Alumni are encouraged to send name/address corrections and changes to Alumni Affairs so that their address information can be kept up to date. Alumni may choose to phone in the information, send email via the Big State University's alumni web site, or send updates via regular mail. The secretary at Alumni Affairs records this information into the Alumni database on a weekly basis.

Unfortunately, only about 1 in 10 alumni remember to inform Big State University of their changes name and address changes. To track down moving alumni, every quarter Alumni Affairs sends a list of its mailing labels to a Change of Address Service, which compares the address of alumni against its master list and identifies those addresses that have changed. This service has demonstrated it is able to detect and correct 98% of the changed addresses and can also identify misspelled addresses and names. In other words, 2% of the changed addresses remain undetected by the Change of Address Service.

Besides the problem of incorrect or out-of-date addresses, there is also the problem of identifying deceased alumni. While relatives may contact Big State University about 20% of the time to stop the mailings to deceased alumni, in many cases, there is no notification. To help it identify deceased alumni, once a year, Big State University sends its active mailing list to an Obituary Service, which compares the list to its master list and can identify about 80% of the alumni who are now dead.

When it is time for Big State University to send out an alumni publication, Alumni Affairs runs a program to create a list of mailing labels, which are then pasted onto the outgoing publication by the University Mail Service. To minimize the number of out-of-date mailing labels, Big State University has embarked on a data quality campaign to better manage the production of its alumni mailing labels.

3.1 Create the IP Map

Big State University's Alumni Affairs Division begins its data quality campaign by constructing an information production map to describe its production of mailing labels. The IP-MAP is described by both a graphical depiction of the mailing label production process as well as a table of metadata that completes the representation by capturing information about each of the blocks and the data elements included in each flow in the IP-Map model.

Figure 2: IP-Map for Mailing Labels

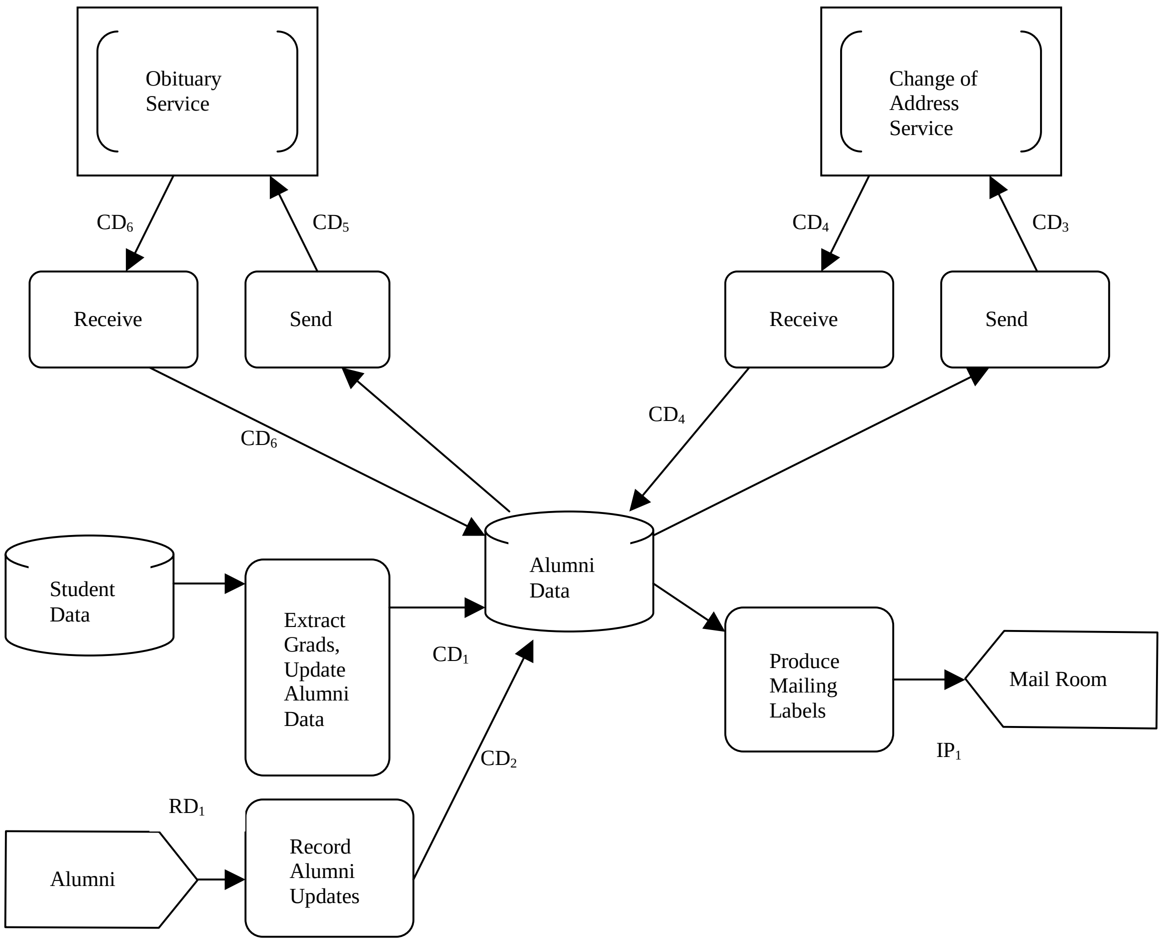


Table 2: Metadata for IP-Map in Figure 1

Name/Type	**Dept/Role**	**Location**	**Business Process**	**Composed Of**	**Base System**
Student (STO_1)	BSU IS Academic Division	222 Carpenter Hall	IS Academic maintains data on active students	Raw data produced by another system.	Oracle Database
Extract Grads, Update Alumni Data (P_1)	BSU IS Academic Division	222 Carpenter Hall	IS Academic extracts list of graduates each academic year.	Extract from Student Database (STO_1)	Cobol Program
Alumni (DS_1)	Alumni	Alumni	Alumni send in their mailing label updates.	Name, Address, and Status Changes	Mail, phone, or electronic correspondence.
Record Alumni Updates and Corrections (P_2) (Corrective Process)	BSU Alumni Affairs	202 Breezedale Hall	Secretary records updates.	RD_1 (Raw Data)	Entered via electronic form into Oracle Database.
Alumni Data (STO_2)	BSU Alumni Affairs	204 Breezedale Hall	Alumni Affairs maintains data on alumni.	Alumni Mailing Information, Other Alumni Data	Oracle Database
Send to Address Service (P_3)	BSU Alumni Affairs	204 Breezedale, Hall	Alumni Affairs prepares an extract file to send every quarter.	Extract from Alumni Data (STO_2)	COBOL Extract program, transmit tape.

Address Service (BSB_1)	NCOA, Inc, BSU Alumni Affairs	NCOA HQ, MD 204 Breezedale Hall	Address Service performs a matching operation.	CD_3 (Component Data)	Vendor System
Receive Updates from Address Service (P_4) (Corrective Process)	BSU Alumni Affairs	204 Breezedale Hall	Alumni Affairs receives address updates from NCOA. Applies updates to Alumni database shortly after receiving.	CD_4 (Component Data)	Tape received. COBOL program used to load updates into database.
Send to Obituary Service (P_5)	BSU Alumni Affairs	204 Breezedale Hall	Alumni Affairs prepares an extract file to send every year.	Extract from Alumni Data (STO_2)	COBOL Extract program, transmit tape.
Obituary Service (BSB_2)	Obituary, Inc., BSU Alumni Affairs	Obituary Service HQ, VA 204 Breezedale Hall	Obituary Service performs a matching operation.	CD_5 (Component Data)	Vendor System
Receive Updates from Obituary Service (P_6) (Corrective Process)	BSU Alumni Affairs	204 Breezedale Hall	Alumni Affairs receives obituary updates. Applies updates to Alumni database shortly after receiving.	CD_6 (Component Data)	Tape received. COBOL program used to load updates into database.
Produce Mailing Labels (P_7)	BSU Alumni Affairs	204 Breezedale Hall	Alumni Affairs runs a program to produce a set of mailing labels.	Alumni Data (STO_2)	COBOL program used to extract data, produce mailing labels.
Mail Room (IP_1) (Information Product)	BSU Post Office	111 Sullivan Hall	Mailing labels are used to address alumni publications.	Name, Address, City, State, Zip	Set of Paper Labels

3.2 Identify the IP Components and their Potential Problems

Every information product can be viewed as a collection of parts in the same way that a manufactured product such as a car is itself a collection of different parts. These parts are often in turn a collection of smaller parts. For example, a car has many parts (wheels, engine, chassis, seats, etc.). Some of these parts (ex. wheels) can be divided into smaller pieces (ex. tire, hub cap, bolts, etc.). An information product like a mailing list can be viewed as a collection of individual labels. Each label can in turn be broken down into components (name, address, city, state, zip). Some of these components (ex. name) can be further subdivided into smaller pieces (i.e., first name, middle name, last name). Just as different manufactured parts have different problems and failure rates, the different components that make up the information product can have different types of data irregularities and error rates.

Before Big State University's Alumni Affairs Division can estimate the reliability of their mailing labels, they must determine the data components that make up their information product as well as the potential problems that may plague certain components (see Table 3). While some data irregularities affect the information product as a whole such as a duplicate or obsolete mailing labels, some data problems such as an alumni move to a different borough of the city may only affect certain parts of the mailing label. It will be up to the Alumni Affairs Division to decide at what level of detail to track the quality of the mailing label (i.e. mailing label as a whole, main components, or sub-components).

For this example, the Alumni Affairs group uses the standard three-line format used for domestic mail in the U.S. For the sake of simplicity, Alumni Affairs does not send publications overseas or to military address that require a four-line format. The problem that Alumni Affairs is most interested in tracking is out-of-date addresses, names, and deaths. Mistakes in spelling occur relatively infrequently and are of less consequence to the Alumni Affairs group. Note: Since this case is fictional, the error rates cited in Table 3 are meant only to illustrate the Control Matrix technique. In a real-life scenario, the data administrator would need to base the error rates on historical observations, U.S. Census or U.S. Postal data, or experimental study.

Table 3: Components that make up a mailing label

Mailing Label		
Components	**Sub-Components**	**Potential Problems**
Recipient Line	Name treated as a whole	Typos - Mistakes in the spelling of the name occur infrequently and typically do not interfere with the publication's delivery. Changed (most often occurs in Alumni Data (STO_2) when female alumni marry or divorce). Alumni Affairs estimates that about 6% of the alumni per year request a name change. Alumni Affairs also estimates that about 60% of the name changes also involve an address change as well.
Delivery Address Line	Address treated as a whole	Alumni Affairs estimates that a small percentage of alumni have mistakes in their address. Typically typos include incorrect or missing direction suffix, wrong street name or number, wrong or missing route number, and wrong or missing apartment number. The Address Change Service will generally catch and correct these mistakes. Changed (most often occurs in Alumni Data (STO_2) when alumni move). Alumni Affairs estimates that 16% of their alumni change addresses each year.
Post Office, State, Zip Code + 4 Line	i. Post Office (City)	Alumni Affairs estimates that a small percentage of alumni have mistakes in the City, State or Zip fields. The Address Change Service will generally catch and correct these mistakes. Changed (most often occurs in Alumni Data (STO_2) when alumni move to a new city). Alumni Affairs estimate that 1/2 of the moves involve a change of city.
	ii. State	Changed (most often occurs in Alumni Data (STO_2) when alumni move out of state). Alumni Affairs estimates that 1/4 of the moves involve a change of state.
	iii. Zip Code + 4	Changed (most often occurs in Alumni Data (STO_2) when alumni moves to a new zip code region). Alumni Affairs estimates that 3/4 of the moves involve a change in zip code.
Mailing Label as a Whole		Obsolete (occurs in Alumni Data (STO_2) when alumni dies). Alumni Affairs estimates that 8% of its alumni die each year. Duplicate Labels - This issue is not considered a serious problem by Alumni Affairs.

3.3 Construct Information Product Control Matrix

Big State's Alumni Affairs Group is now ready to set up their Information Product Control Matrix. Their focus will be on out-of-date labels for now, but they could expand the matrix to include address mistakes as well. Alumni Affairs has also included frequency and cost information in their matrix along with an assessment of the reliability of current data quality checks in their process.

Information Product Control Matrix						
	Label	Recipient Line	Delivery Address Line	Post Office, State, Zip+4 Line		
	Alumni	Name	Address	City	State	Zip
	Dead	Changes	Changes	Changes	Changes	Changes
Error Frequency	.67% of alumni per month	.5% of alumni per month	1.33% of alumni per month	.67%% of alumni per month	.33% of alumni per month	1 % of alumni per month
Avg. Dollar Error Cost per individual mailing	$3	$0	$1*	$1*	$1*	$1*
IP-Map Constructs that control for that data error.						
Record Alumni Updates and Corrections (P_2).	20% of dead alumni detected per month	10% of name changes detected per month	10% of address changes detected per month	10% of city changes detected per month	10% of state changes detected per month	10% of zip changes detected per month
Receive Updates from Address Service (P_4)		98% of name changes detected per quarter	98% of address changes detected per quarter	98% of city changes detected per quarter	98% of state changes detected per quarter	98% of zip changes detected per quarter
Receive Updates from Obituary Service (P_6)	80% of dead alumni detected per year					

Legend & Notes: Percentages in the body of chart reflect the ability of a quality check or corrective process to detect/correct a particular type of error in the alumni records that it processes.

*2/3 of alumni use mail forwarding, 1/3 do not. The $1 represents the average expected loss of the publication and mailing fees.

3.4 Assess the reliability of the information product

From the information product control matrix, the data administrator can use several methods to assess the overall reliability of the information product depending on the type of data provided. The IT auditing literature cites examples using deterministic models, software reliability models, engineering reliability models, bayesian models, and simulation models to evaluate the use of controls on data integrity. For the data provided in this example, one can use a spreadsheet to analyze the data quality levels over the short term. Given that the both the error rates and corrective processes are time dependent, the results for measuring the reliability of the mailing labels are best displayed graphically over time rather than as a single number. From these results, the data administrator can make a determination based on the needs of the business as to whether or not the data quality levels are acceptable.

3.4.1 Reliability of the Information Product

Using the data supplied by the IP Control Matrix along with a spreadsheet, one can forecast the number of labels with out-of-date information using the formula:

$$\text{Num\_Bad\_Labels}_j = \sum_{i=1}^{n} (\text{Num\_Bad\_Labels}_{ij-1} + \text{New\_Bad\_Labels}_{ij} - \text{Corrected\_Labels}_{ij})$$

where

i = 1, 2, 3, ..., n reasons for out-of-date labels, j = 1, 2, 3, ..., m periods.
Num_Bad_Labels_j = the overall number of bad labels in period j.
$\text{Num\_Bad\_Labels}_{ij-1}$ = the number of labels containing a particular type of error i in period j-1.
$\text{New\_Bad\_Labels}_{ij}$ = the number of additional labels that are afflicted with a particular type of error in period j.
$\text{Corrected\_Labels}_{ij}$ = the number of labels containing a particular error i that were detected and corrected in period j.

	Time Period											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Name Changes												
% of alumni who change name each month	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
% of name changes self-reported & corrected each month	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
% of name changes caught by Address Service & corrected per month	0	0	1.35	0	0	1.35	0	0	1.35	0	0	1.35
Cumulative % of labels with undetected name changes each month	0.93 *****	1.38	0.48	0.93	1.38	0.48	0.93	1.38	0.48	0.93	1.38	0.48

******** Period one assumes a starting name error rate among alumni mailing labels of .48%.

	Time Period											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Address Changes												
% of alumni who change address each month	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33
% of address changes self-reported & corrected each month	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
% of address changes caught by Address Service & corrected per month	0.00	0.00	3.60	0.00	0.00	3.60	0.00	0.00	3.60	0.00	0.00	3.60
Cumulative % of labels with undetected address changes each month	2.47 ****	3.67	1.27	2.47	3.67	1.27	2.47	3.67	1.27	2.47	3.67	1.27

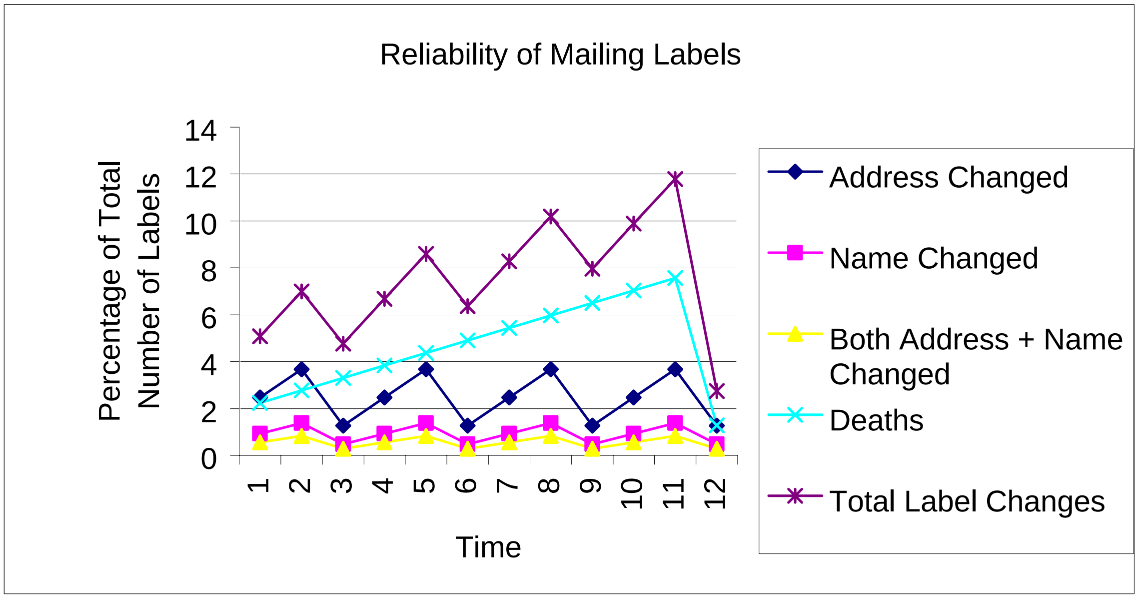
******** Period one assumes a starting address error rate among alumni mailing labels of 1.27%.

	Time Period											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Alumni Deaths												
% of alumni who die each month	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
% of alumni deaths reported & corrected each month	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
% of alumni deaths caught by Obituary Service & corrected per month	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.82
Cumulative % of labels addressed to dead alumni each month	2.24 ****	2.77	3.31	3.84	4.37	4.91	5.44	5.97	6.51	7.04	7.57	1.29

******** Period one assumes a starting death error rate among alumni mailing labels of 1.71%.

	Time Period											
Totals	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Cumulative % of labels with undetected name changes each month	0.93 *****	1.38	0.48	0.93	1.38	0.48	0.93	1.38	0.48	0.93	1.38	0.48
Cumulative % of labels with undetected address changes each month	2.47 ****	3.67	1.27	2.47	3.67	1.27	2.47	3.67	1.27	2.47	3.67	1.27
Adjustment for 60% Dual Name & Address Changes	-0.56	-0.83	-0.29	-0.56	-0.83	-0.29	-0.56	-0.83	-0.29	-0.56	-0.83	-0.29
Cumulative % of labels addressed to dead alumni each month	2.24 ****	2.77	3.31	3.84	4.37	4.91	5.44	5.97	6.51	7.04	7.57	1.29
Total % of labels with undetected errors each month	5.08	6.99	4.77	6.68	8.59	6.36	8.28	10.2	7.96	9.88	11.8	2.75

The graph below shows the mean percentage of mailing labels that have data quality issues for the next 12 periods.



3.4.2 Average dollar impact of erroneous information product

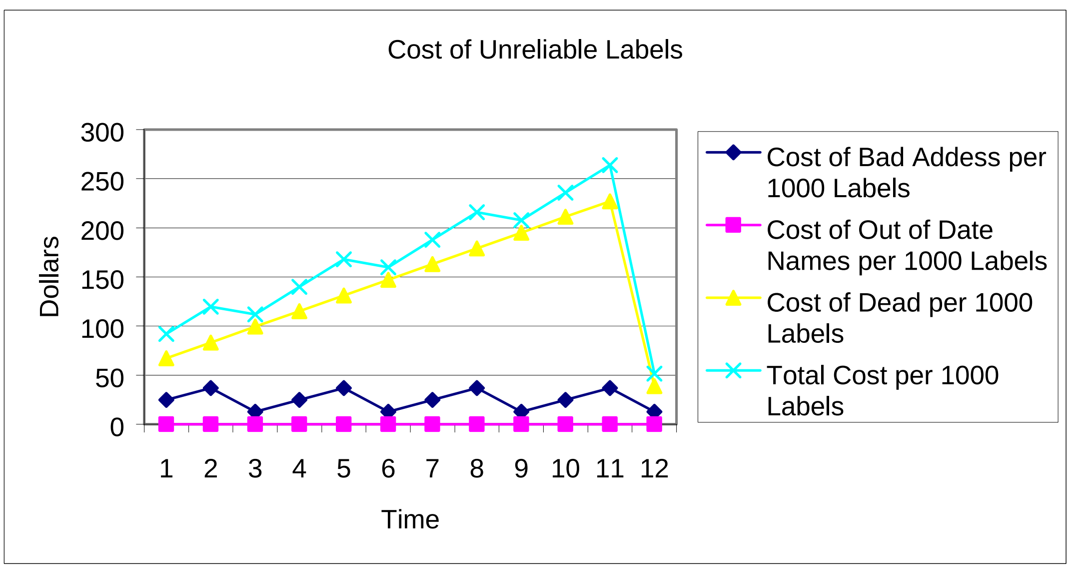
Using the cost information from the IP Control Matrix and the formula:

$$\text{Cost}_j = \sum_{i=1}^{n} \text{Cost\_Error}_i * \text{Num\_Bad\_Labels}_{ij}$$

where

Cost_j = Overall cost of bad labels mailed out in period j.
Cost_Error_j = Cost of a label containing a particular type of error i.

one can display a graph of the average cost of out-of-date mailing labels for the next 12 periods.



3.5 Discussion of Results

Once the data administrator has established the reliability level of the labels and the associated costs of bad labels over time, decisions can be made as to whether the quality is acceptable or if improvements should be made. In this fictional case, the Alumni group at Big State University will have to weigh the costs of additional quality checks or corrective processes against the perceived benefits of increasing the percentage of correct labels. One factor that was not explicitly considered in this case is the timing of the alumni publications. It may be that there is no problem with letting the quality of the mailing list deteriorate provided a clean up is performed prior to the commencement of a mass mailing.

If further data quality improvements are desired, the IP-Map is a useful tool for helping the data administrator to identify potential areas of improvement. Revisions in the information production control matrix and the subsequent what-if analysis using spreadsheets can help project the amount and value of the quality improvements as part of the cost-benefit analysis.

4. Summary and Future Research Directions

The information product control matrix is a tool that can be used to evaluate the reliability of an information product. Essentially, it is an application of control matrices, a tool that IT auditors have long used to help them to make an evaluation decision on how well a system safeguards assets and protects data integrity. The information product control matrix is designed to be used in conjunction with IP-Map in the same way that a traditional control matrix can be used in conjunction with a data flow diagram that shows the controls exercised over the data flows through a system.

An information product control matrix differ from the data manufacturing analysis matrix that Ballou (1998) used to track the data units through the various stages of the data manufacturing process insofar that it focuses just on the corrective processes and data quality checks that most influence the rate at which errors occur in the data elements most pertinent to the final information product. In some respects, the information product control matrix is a specialized version of the data manufacturing analysis matrix, which looks at all the data units and the processes they undergo in their entirety.

Further research and development must be done to refine the information product control matrix to ensure that this tool can adapt to all the complexities that an IP-Map can model. It should also be noted that this technique is only as good as one's understanding and knowledge of the information manufacturing process. The more detailed the measurements, the better the estimates of the reliability of the information product. In addition, the development of software that automates the process of creating, maintaining, and analyzing the IP-Map and the IP Control Matrix is greatly needed to encourage practitioners to adopt these data quality tools in the field.

References

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.

Shankaranarayanan, G., R. Y. Wang, M. Ziad, "IP-MAP: Representing the Manufacture of an Information Product," Proceedings of the 2000 Conference on Information Quality, Cabridge, MA.

Weber, Ron, "Information Systems Control and Audit," Prentice Hall, Upper Saddle River, NJ, 1999.