



**Data Management  
for Higher Education**

---

---

**Data Management Recommendations**  
*California State University – Channel Islands*

---

---

*Prepared by Scott Flory and Aaron Walker  
December 3, 2015*

## Table of Contents

<b>Executive Summary.....</b>	<b>4</b>
<b>Background.....</b>	<b>5</b>
<b>Approach .....</b>	<b>6</b>
<b>Decision Support, Data Governance, and Data Management .....</b>	<b>7</b>
<b>Interview Findings.....</b>	<b>9</b>
Themes.....	9
Other Observations.....	10
Analysis of Issues.....	11
<i>Issue 1: Latency.....</i>	<i>12</i>
<i>Issue 2: Data Knowledge.....</i>	<i>13</i>
<i>Issue 3: Communication .....</i>	<i>15</i>
<i>Issue 4: Consistency .....</i>	<i>16</i>
Summary of Findings .....	17
<b>Recommendations .....</b>	<b>18</b>
Overview .....	18
Recommendations for Data Governance .....	20
<i>Establish a Formal Data Governance Framework .....</i>	<i>20</i>
<i>Identify Data Stewards.....</i>	<i>21</i>
<i>Review and Update Security around Data Sources.....</i>	<i>21</i>
Recommendations for Improved Access to Data .....	22
<i>Fill Open Positions.....</i>	<i>22</i>
<i>Increase Expertise among Current Staff.....</i>	<i>22</i>
<i>New Report Request Process.....</i>	<i>22</i>
<i>Complete the Blackboard Analytics Data Warehouse .....</i>	<i>23</i>
<i>Business Intelligence Audit.....</i>	<i>23</i>
Recommendations for Data Knowledge.....	25
<i>Business Glossary (Data Cookbook) .....</i>	<i>25</i>
<i>Central Repository for Information about Data and Reports (Data Cookbook).....</i>	<i>25</i>
<i>Data Asset Inventory.....</i>	<i>25</i>
Recommendation Roadmap.....	26
<b>Appendix A: Fishbone Diagrams .....</b>	<b>28</b>
<b>Appendix B: Interview Schedule .....</b>	<b>32</b>
<b>Appendix C: Data Systems, Reporting Tools, Integrations .....</b>	<b>35</b>
Data Systems .....	35
Reporting Tools.....	35
Integrations.....	35

Appendix D: Data Governance Roles and Responsibilities .....36

Executive Sponsor ..... 36

*Level of Effort* ..... 36

*Responsibilities*..... 36

*Qualities*..... 36

Project Leader ..... 37

*Level of Effort*..... 37

*Responsibilities*..... 37

*Qualities*..... 37

Appendix E: Glossary .....38

## Executive Summary

In order to meet its strategic objectives, CSUCI must be able to take advantage of the vast amount of data available to it, particular the data it collects and generates from its operations. Maximizing these data assets requires a collaborative effort from everyone who generates, works with, or consumes data. Building an infrastructure to support data-driven decision making is difficult without participation by leaders in institutional research, administrative offices, academic departments, and information technology.

Globally, particularly in for-profit situations but increasingly in higher education, we have observed a growing interest in data governance and data management as a way to build such an infrastructure. Data management—in a higher education setting—is the core of our corporate practice, and IData has been hired to assist with establishing a data governance framework and improving data management practices at CSUCI.

IData is now working with CSUCI to do the following: 1) conduct an assessment of its data management needs and resources; 2) provide an integrated set of recommendations for improvement; 3) develop a timeline and roadmap for acting on these recommendations; and 4) help to pilot an initial set of projects.

This document is intended to provide a plan of action for CSUCI, as well as to begin the process of establishing formal data governance and improving the quality and management of data across the campus. This document includes an assessment of existing strengths and weaknesses as a well as a set of recommendations for future action.

The data management issues in play at CSUCI are primarily procedural, and to some extent cultural, rather than technical. Broadly speaking, we recommend establishing formal data governance bodies and protocols, committing to an environment that fosters and demands data fluency, and developing a shared business glossary that will support better communication about and wider understanding of institutional data.

The recommendations that follow are centered on strong institutional data management practices in the areas of data governance and stewardship, data knowledge, and improved access to institutional data. We believe that acting on our recommendations will allow for better operational and analytical use of CSUCI's institutional data, which ultimately can help the university more effectively achieve its strategic objectives.

## Background

In March 20015 California State University – Channel Islands (CSUCI) issued a request for proposals to provide data management and governance consulting. CSUCI is a young and rapidly growing campus, and in many ways its abilities to manage its data have not kept up with its data management needs. As noted in the RFP, the University “developed solutions to data needs that were idiosyncratic to division, department, program, organizational need.”

The university recognized that these solutions are not sustainable, for a number of reasons, including the sheer volume of data now generated on campus, the difficulty meeting external reporting requirements, and the growing desire to utilize data systematically as part of a rigorous, evidence-based decision structure. The systematic utilization of data is difficult without a set of comprehensive data management practices, since utilization will be affected by capture, storage, disposal, and many other issues.

The campus also recognized that it lacked a workable data governance structure, which was resulting in uncertainty regarding reporting processes and reported data, no “codebook” for definitions of data, and an inability to take advantage of best practices surrounding data management, integration, quality, and so on.

IData has been working with institutions of higher education on data management issues since 2004, and because we kept seeing a remarkably similar set of problems and challenges we developed a set of tools, including software, and practices to assist clients with these challenges. The issues facing CSUCI are not unique to the institution, although of course their particular manifestation is unique and requires special attention.

IData responded to the RFP with enthusiasm, and recommended a three-phase process: first, assess the situation and make targeted recommendations; second, pilot some of these recommendations with CSUCI; third, coach CSUCI as it addresses other recommendations independently.

This document represents the culmination of phase one, our approach to which is described below. It uses an assessment protocol that has been successfully employed with numerous other clients, but the recommendations offered, and the steps we suggest in order to address those recommendations, come directly from our experiences with CSUCI staff and documents.

## Approach

IData proposed a set of services in order to assess existing data management practices, recommend changes and/or additions to those practices, and then work with CSUCI personnel to begin to implement a series of organizational and process improvements.

To assess, IData came on campus in September 2015 to interview representatives from all administrative areas as well as academic representatives. These interviews were used to gain an understanding of how various offices are utilizing data and what issues they have related to the production and consumption of data. We also interviewed both the Institutional Effectiveness office and the office of Communications and Technology regarding the analytical use of data and their issues providing and managing quality data.

Respondents were quite candid in these interviews, and the resulting discussions were illuminating and helpful. Many issues were raised, and others became apparent, issues we believe are symptoms of a larger problem that requires an institutional approach to solve.

These interviews were then summarized, reviewed, and analyzed to see which common themes emerged. This document is the final element of the first phase of this effort – assessment. The following sections contain a summary of the interviews and analysis that leads to a set of recommendations. These recommendations are then presented in a general timeline.

The next phases of effort will be implementing selected recommendations and coaching CSUCI staff on the continuation of these processes. This will be a coordinated effort between IData and CSUCI.

## Decision Support, Data Governance, and Data Management

Institutions of higher education are increasingly looking to use data to make better decisions, and CSUCI is no exception. Indeed, we were pleasantly surprised by the desire from all corners of campus to get access to, and make better use of, institutional data. Some institutions refer to this as being data-driven, and others prefer terminology like “evidence-based.” While these preferences probably reflect meaningful differences at some level, we generally lump these phrases and other similar terms under the rubric of *decision support*, by which we mean using data from one or more sources to inform decision making in some manner.

In order to use data to drive decision support, the following data features—at a minimum—must exist:

- 1) Clarity about the source of data
- 2) Understanding of the accuracy and completeness of data
- 3) Transparency of data extraction and methods of analysis
- 4) Regular access to structured data
- 5) Sufficient data fluency within the organization

Moreover, shifting to an evidence-based decision structure often requires considerable cultural change and learning.

Absent reliable, accessible, widely-understood data, we believe most business intelligence and analytics efforts, no matter how well-resourced, will be unsuccessful. Many institutions embark on high-profile data warehousing programs, or purchase expensive “business intelligence” software, in hopes of taking advantage of exciting new technology. However too many find that their data is essentially in an ungoverned state, and no technology or technologies are going to solve that problem.

Governed data is data that exists in an environment where policies and standards have been enacted regarding data quality, integrity, access, storage, disposition, and usage, among other dimensions. The work of creating these policies and developing these standards, broadly speaking, is the core of *data governance*. Data stewardship—defining and classifying data, supervising its use and presentation, etc.—is a key part of data governance, while data operations—building and maintaining a technical infrastructure, backing up or warehousing or integrating data systems, managing tools that enable data discovery—is typically not considered a component of data governance. So improvements to data operations that might come with new tools may not have any real impact on data stewardship or governance.

In our view, *data management* encompasses the set of tools, techniques, and tactics employed to actualize data governance. Data management also encompasses tools used to establish and take care of data architecture and ongoing operations. Data management and data governance are closely related but not interchangeable concepts. Changes to governance policies, for example, will affect data management methods; changes to data

management procedures ought to be reflected in (and in many cases driven by) updates to governance.

The Data Management Association ([www.dama.org](http://www.dama.org)) includes data governance as one of ten components of data management in its body of knowledge. Other components include architecture, data development, data operations, and security, among others. We prefer a simplified conceptual framework for data management, in which data governance (including security, quality, and metadata management) and data operations (including architecture, technology, tools, and skills) form the two halves of the whole.

CSUCI is currently engaging in both data management and data governance activities. It has integrated many different systems, it has developed a reporting architecture, and it has employees who have knowledge of data, systems, and tools. But this work has been piecemeal and idiosyncratic. There is little formal data governance in place, but an informal network of data definitions, rules regarding security and access, and at least at the departmental level quality standards has grown up. The recommendations that follow are designed to make the informal formal, and to extend locally successful practices across the institution.

Of course neither properly governed data nor smoothly running operations represents the end of the work. Institutions must still decide on metrics that matter, for example, and then execute strategies to respond to those metrics. These and other tasks form some of the ongoing work of data management. As will become more apparent later in this document, our recommendations begin with addressing governance issues in the context of a data management overhaul.



## Interview Findings

### Themes

The most common sentiment expressed during these interviews was that CSUCI had large amounts of “good data,” but that people were unable to make use of that data. Typically this was expressed as data requests taking too long to fulfill, so data couldn’t be used when it would be most useful. We discuss this delay in greater detail below.

Another overarching theme, although not always expressed consistently, was the lack of a common language to discuss data. Data consumers and data providers work together well, and in good faith, but they have trouble understanding each other’s specific needs. The gathering of business requirements when providing information is a critical step, and we are pleased to see that it is part of standard CSUCI practice. But if requirements gathering goes on too long, or involves too much back and forth, or starts to seem repetitive, frustration is likely to develop. Frustration and dissatisfaction can jeopardize productive working relationships, they can lead to unauthorized and often ineffective workarounds, and they can eventually diminish or eliminate the commitment to data-driven decisions.

This lack of a common language is closely related to another problem involving a lack of communication about data, particularly decisions about data. The Chancellor’s Office collects, validates, and republishes data drawn directly from CSUCI’s systems, and for many purposes this data set should be considered official. Yet some people seem unaware of it, and others are uncertain about its validity or usefulness.

Several power users have become the go-to source for data, both for reports and for questions about usage. While we have no reason to believe that these users provide erroneous information, the kind and amount of institutional knowledge that they have accumulated is in critical need of better documentation.

Moreover, there is no decision or review structure in place to accommodate or communicate data changes, whether something as prosaic as adding a value to a code table or something with further-reaching effects. More than one office mentioned finding out about a data change only when a report they relied on suddenly showed new values.

Not surprisingly, a common concern was having “one version of the truth,” or as one respondent put it, making it so that the University could “speak with one voice.” To a certain extent, having one version of the truth means reporting the same facts and figures the same way, and that is a worthwhile goal to the extent it is possible. However the idea of speaking with one voice resonates more powerfully with us, as it is evidence of a more sophisticated and nuanced way of thinking about data and its use.

### Other Observations

In an atmosphere where data is hard to extract and utilize, we were nevertheless encouraged by much that we observed. Despite its youth, the campus has a strong institutional memory, and while they are scattered the pockets of knowledge are deep. (A benefit to being young is that there is no history of failure to overcome.) The technical staff are skilled, and they keep a surprisingly large number of systems and processes operational.

With regard to leadership, having executive support for making better use of data is a positive indicator, and while the relationship with the Chancellor's Office regarding data collection could be better, membership in the CSU system looks to be beneficial overall. CSUCI employees show a collegial respect and a cooperative attitude toward each other and toward their sister institutions.

We also observed a number of best practices, mostly at the departmental level. Sometimes best practices are perceived as a roadblock, and this may be somewhat the case at CSUCI. Cutting corners is not sustainable over the long term, and while not every best practice is right for every office, we generally observed employees and offices doing well, but wanting to do better.

While data quality was not raised as an issue directly, except in one high-profile instance regarding student race and ethnicity data, we did hear that sometimes the data that were reported seemed incorrect, or incomplete, or inappropriate for the task at hand—the “good data” in the system didn't turn into good aggregate data. Data sets seem unstable, so facts and figures change frequently, without explanation, and sometimes without notice.

We will remind readers that data quality includes not only accuracy, which can be described as the extent to which data in a system corresponds to known values in the real world, but also completeness, integrity, and timeliness, among other factors. Incomplete data, no matter how accurate, can complicate decisions; data that lacks integrity, that is, data that is inconsistent across systems, is almost inherently problematic; data that isn't captured or updated quickly can cause issues in areas that rely on real-time reporting.

It's not clear whether or to what extent data quality is an issue. Stories of unreliable reports are troubling. As we discuss in more detail below, there is clearly a lack of shared knowledge around institutional data, and that lack of knowledge manifests as a lack of trust in reported data. If not addressed, a lack of trust in data can turn into a lack of trust between colleagues, and we would hate to see that occur in an environment where collaboration and collegiality seem to carry the day.

### Analysis of Issues

We have grouped the most common and/or critical of the issues that were raised into four broad categories, and will explore them below. We used a fishbone diagram as part of our methodology for helping to unpack issues and to identify their root causes and critical inflection points

### How to read a fishbone diagram

In each diagram, the large box on the right is the main issue being analyzed. The center line is drawn from that and the “causes” for the issue become the lines going off of the center line. We keep asking “why” for these reasons to get additional nodes until the answer of “why” begins to repeat or become a common theme. These are our root causes to the main issues and appear in bold italicized text on the diagrams.

Full-size copies of the fishbone diagrams are included in the Appendix.

## Issue 1: Latency

The inability to get information in a timely fashion was one of the common themes. Requests for data are slow to be filled, and sometimes even slow to be acknowledged. The request process is opaque, as consumers don't always know which office to ask, what request format to use, or how to respond to follow-up questions.

The clear implication is that decisions are made in a relative vacuum: questions take too long to answer, and the data team (whoever that might be) is overtaxed just trying to present raw data.

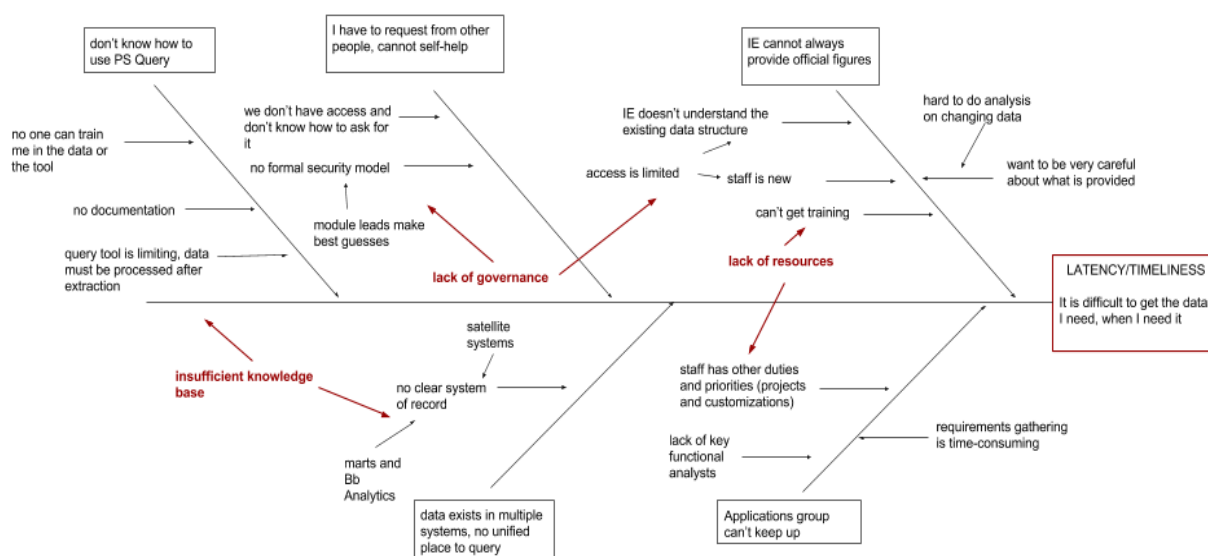
### Primary symptoms

- Application Services and Institutional Research Planning and Effectiveness (IRPE) cannot be as responsive as they'd like to be, bottlenecks result
- Repeated, common requests still require manual effort
- Process is slow and communication sparse – not sure when to expect results and what to do in the meantime
- Requirements gathering takes too long – much back and forth
- No certainty regarding which office or person to ask for which requests
- Too few users are able to write, alter, or even execute reports or queries

### Root causes

- Supply of reporting & analytical resources insufficient to keep up with the demand for information
- Process for requesting and providing data is inconsistent
  - No central repository for reporting deliverables
  - No standard definitions of the data
- Lack of governance and governed data
- Campus knowledge base not broad enough

### Latency Fishbone



## Issue 2: Data Knowledge

This is a broad heading for different related problems, a lack of trust in reported—particularly summarized—data, and a lack of understanding of the data, even when its accuracy or completeness is not in doubt.

While mistakes can happen at many points during data extraction, transformation, and presentation, the basic issue is likely to be definitional: enrollment figures, for example, are generated multiple times a year for multiple audiences, and figures that may be accurate in one context or period may not be the right ones to show at a different time or to different consumers. When multiple data systems are in play, specifying where data should come from and how it should be treated becomes even more difficult.

We have observed elsewhere that when organizations move from dealing primarily with operational data to working with analytical data, there is a period of growing and learning.

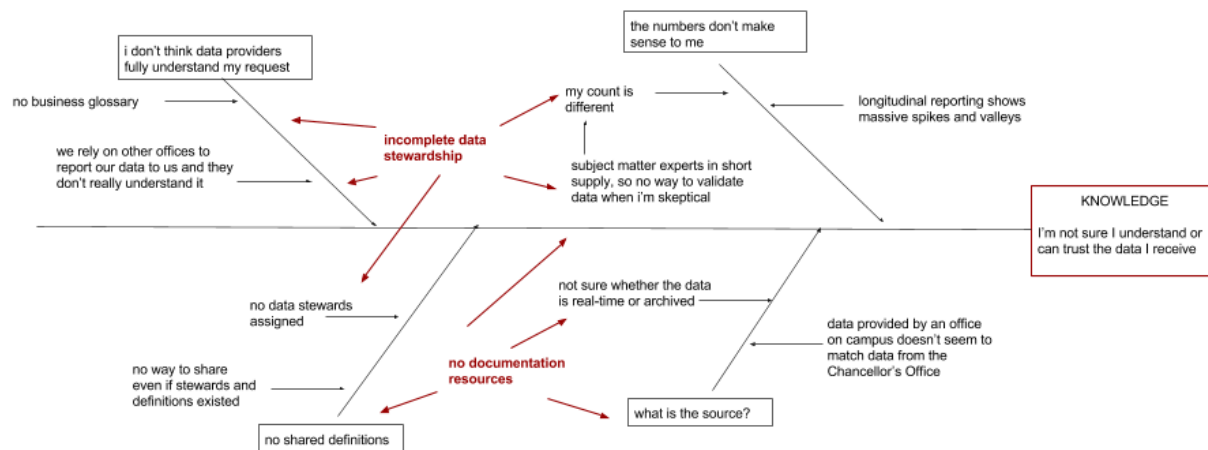
### Primary symptoms

- Numbers not consistent from year-to-year or period-to-period
  - Internal figures don't correspond to Chancellor's Office's figures
- Not clear where data comes from, or how it's calculated
  - Difficult for consumers and providers to be in sync
  - No process for asking for more information
- Consumers don't know enough about data definitions to make informed requests

### Root causes

- Weak data stewardship
  - No business glossary
  - General lack of subject matter expertise
  - No shared knowledge base
- Multiple data sources with limited documentation
  - No data asset inventory
  - No integrated data model

### *Knowledge Fishbone*



### Issue 3: Communication

In many cases this issue was described as another office making a decision about data capture or data entry and other offices not learning about it until later. Sometimes this can be a minor issue, but other times these “upstream” changes can have a significant impact on operations.

However, communications issues presented around nearly every discussion of reporting needs: data consumers can’t express their needs in terms that data providers understand; data providers can’t ask follow-up questions in language that’s meaningful to requesters; data analysts don’t have a forum to share their challenges (technical, procedural, whatever) with recipients of their analysis; data definitions are always in flux.

We also note that the Chancellor’s Office is a consumer of CSUCI data, which then provides data about the campus to a public audience. In the diagram below, the Chancellor’s Office is both source and symptom of difficulties communicating about and with data.

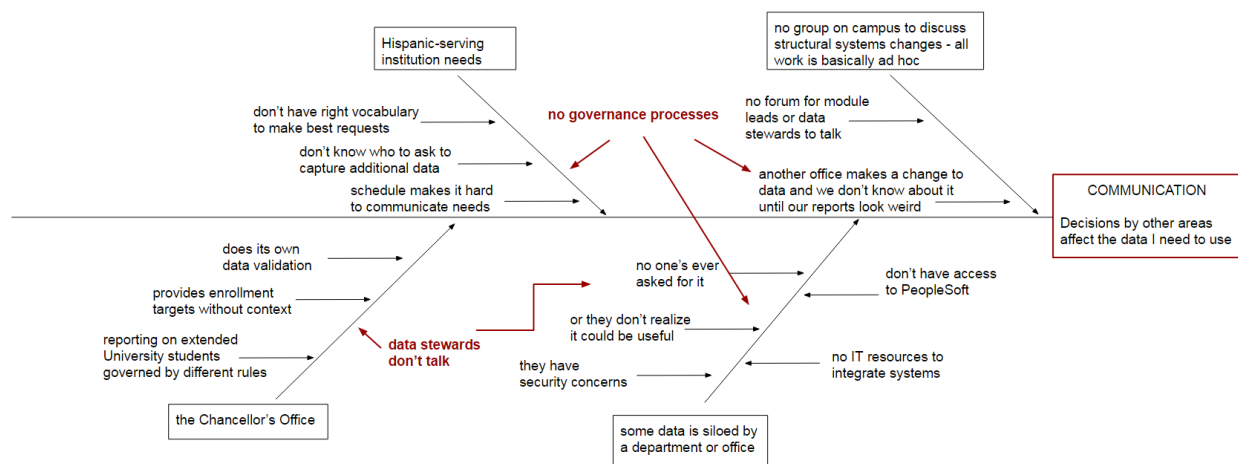
#### Primary symptoms:

- Code changes, batch processes, and other actions affect other offices’ work
  - Application modifications not shared beyond requesting office
  - Offices do not request critical information from the offices that know the data best
- Data from satellite systems not imported into/integrated with system of record
- Important requests and strategic initiatives hard to prioritize

#### Root causes:

- Data stewardship is fragmented and in some areas nonexistent
  - No forum to discuss impact of projected changes
  - Responsibilities in this area not clearly defined
- No formal data governance

#### Communication Fishbone



#### Issue 4: Consistency

This could also be known as the “one voice” problem. In many circumstances, the University needs to speak with one voice, whether that’s the President speaking to public audiences, the Development Office meeting with potential donors, the Enrollment Management office working with recruits, or the Hispanic Serving Initiatives (which cross many disciplinary boundaries) performing a range of functions.

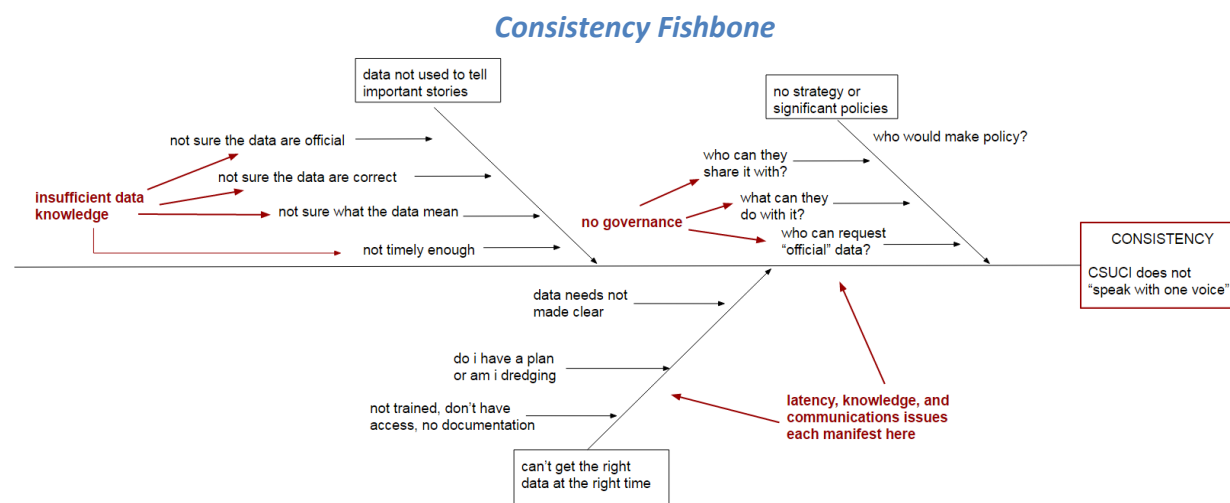
Having one voice is not necessarily the same as having one official data source, or set of numbers. Several offices on campus, most notably IRPE, are required to report figures to external agencies. These figures ought to be vetted, and they need to be understood as official within certain contexts.

Primary symptoms:

- Are these numbers official? Who decides? How long do they remain official?
- What is the difference between “accurate” and “accurate enough?”
- Not able to use data to tell stories (or not able to use consistent data to tell consistent stories)

Root causes:

- Not enough data knowledge/fluency on campus
- Lack of data governance
- Latency/accessibility issues
- No specific plan for leveraging data





### Summary of Findings

The overall findings from these interviews suggest the lack of a consistent data management structure, the relative absence of data stewardship, and an inability to be properly responsive to questions about data. Important controls and processes are missing or outdated and the reliability and accessibility of institutional data are thereby hampered.

Although we grouped our observations along four main dimensions, it's clear that these dimensions are related, and that the issues that cause one tend to help cause others. A lack of analysts, for example, leads to users seeking other avenues to get data, and making assumptions about what that data might represent without getting confirmation. The absence of a central repository of reports, and documentation of what they contain and how to obtain them, results in repeated inquiries and duplication of effort, and reflects both a dearth of knowledge and an inability to communicate effectively.

Note that what we are describing are process deficiencies, not personnel problems. What we observed was many, many employees going well beyond their job description to provide assistance, both to colleagues in their department or division as well as to coworkers across campus. We referred above to a number of good practices and admirable attempts to work with data, and undoubtedly saw only a tiny percentage of those efforts. But it's clear that they have occurred in an uncoordinated fashion, and in an environment that is not currently configured to build on them.

## Recommendations

### Overview

Using data to support decisions is a good goal, but it is not a strategy. How, exactly, are data going to be put to use, by whom, and in what ways? Who can provide data, and to whom? What level of analysis will be required before data can be brought to bear in decisions? Will enrollment management be the centerpiece, or financial management, or will there be a broad-based commitment to data quality? We have seen problems arise when there are competing sources of truth, when there is needless duplication of effort, when data requests take too long to fulfill. How will CSUCI get rid of distractions and focus on insights?

We saw a desire and/or need for the following (and this is only a partial list):

- Increased assurance and dependability of knowledge assets
- Improved information security
- More across-the-board accountability
- Maximization of asset potential
- Consistent data quality
- Improved communication and sharing

In order to meet these needs, CSUCI must, we believe, embark on a broad program of unified data management that spans the institution. CSUCI is already managing its data, of course, but its efforts aren't coordinated and its outcomes aren't consistent. What are the rules regarding data capture, storage, retention, and disposition? Who is responsible for identifying and/or remediating data quality issues? Where do data consumers go to learn more about appropriate usage of data, or even to find out more about what kinds of data are available to them? How and why do data marts and warehouses come into being, and who determines their content, and what is their purpose?

We propose several specific, linked recommendations below, and offer a suggested timeline for implementation. Some of these recommendations have to do with establishing data governance roles and rules, and others are oriented toward improving data management practices. We believe that by doing so CSUCI will be well-positioned to further refine, enhance, and implement its data strategy.

In the analysis in the preceding section, we identified several root causes that repeated from one diagram to another. Common root causes include a lack of resources, gaps in knowledge, confusion about process, and a general disconnect between information goals and data practices.

These root causes are bound together by a lack of data governance and by inconsistent data management practices, which we find contributes greatly to any organization's inability to turn data into actionable knowledge. First and foremost, then, is the establishment of data governance structures within the institution, in concert with steps to increase access to and

knowledge about institutional data. These are summarized below, and then followed by detailed recommendations.

### **Data Governance**

A formal data governance structure is recommended in order to address the wide variety of data issues on campus. Without this as a foundation, efforts to improve perceived reporting issues are less likely to succeed. We caution that data governance has to be more than a collection of *ad hoc* data improvement projects; rather, a robust data governance framework will provide the structure and institutional oversight necessary to establish a culture of data fluency across the institution.

### **Data Accessibility**

Many of the information needs on campus are not extraordinarily complex, yet the processes to extract and share data are cumbersome and inefficient. Upgrades to resources, procedures, and the reporting infrastructure are recommended. User roles may need to be reconfigured or created new, and security protocols should be reviewed and quite possibly updated.

### **Data Knowledge**

Have key business data elements been identified? How are they chosen? How are they defined? Who has done this work? What are the creation and usage rules around them? Answering these questions is core to the work of data stewardship, and will assist in moving knowledge from scattered pockets to a centrally managed and accessible resource.

### Recommendations for Data Governance

Data governance as a concept was discussed above in some detail. Briefly, data governance refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. A sound data governance program includes a governing body or council, a defined set of procedures, and a plan to execute those procedures. A working data governance structure should ensure that authority is delegated from senior leadership to appropriate parties and that these parties be held accountable.

Creating a data governance program involves identifying the stewards of the data assets in the enterprise, as well as representatives from across the organization who participate in decisions regarding the use and application of data assets, such as academic deans, executives, and information technology professionals. These people usually form the governing council.

Policies must be developed that specify who is accountable for various aspects of data, including its accuracy, accessibility, consistency, completeness, and updating. Processes must be defined concerning how the data is to be stored, archived, backed up, and secured. Standards and procedures must be developed that define how data is to be used, by whom, and for what purposes. Finally, a set of controls and audit procedures must be put into place that ensures ongoing compliance with governance policies.

#### Establish a Formal Data Governance Framework

Governance leadership must be identified: an executive sponsor, who will provide ongoing leadership and retain overall authority for monitoring the Data Governance Council and its activities; a project leader or leaders, who will head up the initial projects of the Governance Council; and other roles (see a sample list in Appendix D).

We are recommending starting with one governing body for data governance, although a mature data governance framework would eventually include multiple levels of involvement, for example a bicameral structure. We emphasize that data governance is a business need, and thus that the most successful governance efforts are characterized by support and sponsorship from the highest levels.

The Data Governance Council ought to be composed of representatives from functional areas both academic and administrative. It will include data stewards but also people with an interest in and ability to promote governance. The Council's responsibilities will include:

- Creating a data governance charter
- Establishing policies related to data (usage, access, quality, etc.)
- Making business decisions on issues or changes that have an impact on data, such as:

- Adding data, needing a new field, or changing data usage that will impact more than one group
  - Reviewing decisions of a new use or change of use of institutional data
  - Managing new/missing data, deciding how it will integrate with institutional data
- Establishing standards for definitions and reports
- Starting additional projects to correct or address data issues that have not been identified or prioritized
- Growing, evolving, or even shrinking over time as the situation warrants

#### Identify Data Stewards

Within each area that manages data across the institution, we would want a Data Steward identified who would represent that area on the Governance Council. Their role would be to manage knowledge and definitions for data elements within their own realm, to serve as a point of contact for people with questions, and to escalate issues to the Governance Council:

- Subset of members of the Data Governance Council
- Identify representatives from functional areas
- Establish roles and responsibilities to include:
  - Review and approval of definitions in their own areas
  - Establishing policy within their own areas
  - Review and decisions on data quality issues
  - Responding to requests for access & clarification
  - Escalating issues to the Data Governance Council

#### Review and Update Security around Data Sources

Make sure users can run reports (or view individual transactions) against the data they need. Make sure Data Stewards know which users and roles have access to data in their area, and institute a process by which users may request access and have that request reviewed and addressed quickly.

### Recommendations for Improved Access to Data

The exact form of more distributed access can take many shapes, and we are not in a position to make an exhaustive list of recommendations here. The Blackboard Analytics project can be a point of entry here, but perhaps additional data marts or even simply custom user-facing views are a good path.

#### Fill Open Positions

We understand there are several open analyst positions across campus. Since it is critical to increase data knowledge and access, these positions should be filled. However we also recommend that CSUCI carefully consider how to align their work with existing needs, how to allocate these resources, and then how to provide the training they're going to need.

#### Increase Expertise among Current Staff

As part of the new report request processes, identify staff in each area who have technical abilities (and/or aspirations), and develop and provide training to help them handle data needs more independently. We have been using “functional representative” as the title for this role, but CSUCI may decide on a different name.

PeopleSoft contains a powerful query tool, albeit one with a steep learning curve and an inflexible interface, and further training in this tool would be beneficial, because so much institutional reporting comes from this data source using this tool. Ultimately, there may be additional tools provided to make reporting easier (see later recommendations), but this will alleviate the burden on IT while allowing offices to have quicker internal response to their simpler, routine inquiries.

#### New Report Request Process

Train the functional representatives in each area to initiate, review, and follow up on report requests. Identify queue managers. Establish metrics and analytics to monitor the request process and handle escalations. IT & IR must decide how to share work.

- Centralized management of requests
- No more direct requests to Ana from offices other than her own
- Shared effort between Application Services and IRPE
  - Nature of request will determine which office responds to it
  - Keep departments aligned in best practices for reporting and analytic processes
- Updates central repository
  - Specifications on Reports and Analytics
  - Data Systems utilized

*Between these last two access recommendations, the reporting model for campus becomes partially de-centralized. The demand for data is already too great for a central resource (or two) to keep up, and that demand should be expected to grow. At the same time, the combination of Application Services and IRPE can function like the*

*hub in a hub-and-spoke model, evaluating and parceling out requests prior to their being fulfilled.*

### **Complete the Blackboard Analytics Data Warehouse**

While this doesn't solve all the issues surrounding official data, this can become the primary source for compliance and other traditional institutional research reporting, and we expect that in most cases IRPE will be the source for official figures. Because this tool is potentially so useful, developing strategies to make it available more widely ought to begin now.

### **Business Intelligence Audit**

Neither the PeopleSoft transactional model nor its embedded query tool is especially conducive to end-user reporting. Still, better use can probably be made of the existing tools and architecture. In the long term, however, it is worth investigating ways to upgrade this architecture to accommodate data discovery and enhanced analytics. This may include an expanded implementation of Bb Analytics as well.

### **Gather Reporting and Analytics Requirements from Each Area**

- Execute using the new report request process
- Discussions with each area on their reporting and analytic needs
  - self-service reporting and analysis
  - establishing definitions of metrics
- Identifying needs for accessibility
  - Interaction with the data
  - Presentation of the data

### **Gather Architecture Requirements**

- To provide an accessible reporting and analytics architecture to all users
- Combining requirements from IT and IE as providers of reporting and analytics
- Assess current tools
- Identify additional technology needed to meet the requirements

### **Provide Architecture**

- Establish a planned reporting architecture (snapshots, ODS, warehouse, how do they fit and why use one over the other)
- Procure tool(s), if necessary
- Implement tool(s)
  - Provide access
  - Provide training

### **Provide Reporting and Analytics**

- Based on requirements gathered earlier
- Utilizing new architecture for delivery

- Supported by definitions and specifications from the central repository



### Recommendations for Data Knowledge

A knowledge management program that brings together multiple functional offices, collects and shares information about policies, procedures, and resources, and offers functional and technical training will contribute to better data requests, timelier and more accurate reporting, and improved data capture and management. In conjunction with improved end-user access to data, analytical support for decision-making will be enabled.

There's a role for nearly everyone on campus in this work. IRPE needs to make clear (on its web site and elsewhere?) what kinds of data it reports out, to whom, and for what purposes. Application Services, and Computing & Technology generally, can document, or better share existing documentation of, logical and physical data models. Functional offices can document data flow as part of their business process—see the training manuals authored by Financial Aid as examples. Data consumers, particularly in Academic Affairs, can continue to share their program review and grant application needs. And the Data Cookbook, discussed below, will support this process.

### Business Glossary (Data Cookbook)

Implementing the Data Cookbook would go a long way towards resolving many of the problems identified. The Data Cookbook is not a panacea, but it is a central knowledge resource that assumes collaboration as its business model.

### Central Repository for Information about Data and Reports (Data Cookbook)

Document existing reports in the Data Cookbook – not all reports, necessarily, but critical ones and new requests. Adjust the applications team practice to include the Data Cookbook in their requirements-gathering phase. For those offices that have a library of PSQueries, import those libraries into the Data Cookbook.

- Specifications for critical reports and reporting deliverables
  - PSQueries
  - Chancellor's Office submissions
  - Existing data marts
  - BbA warehouse and customization
- Definitions and workflow to approve them
- Visibility and availability to all data consumers

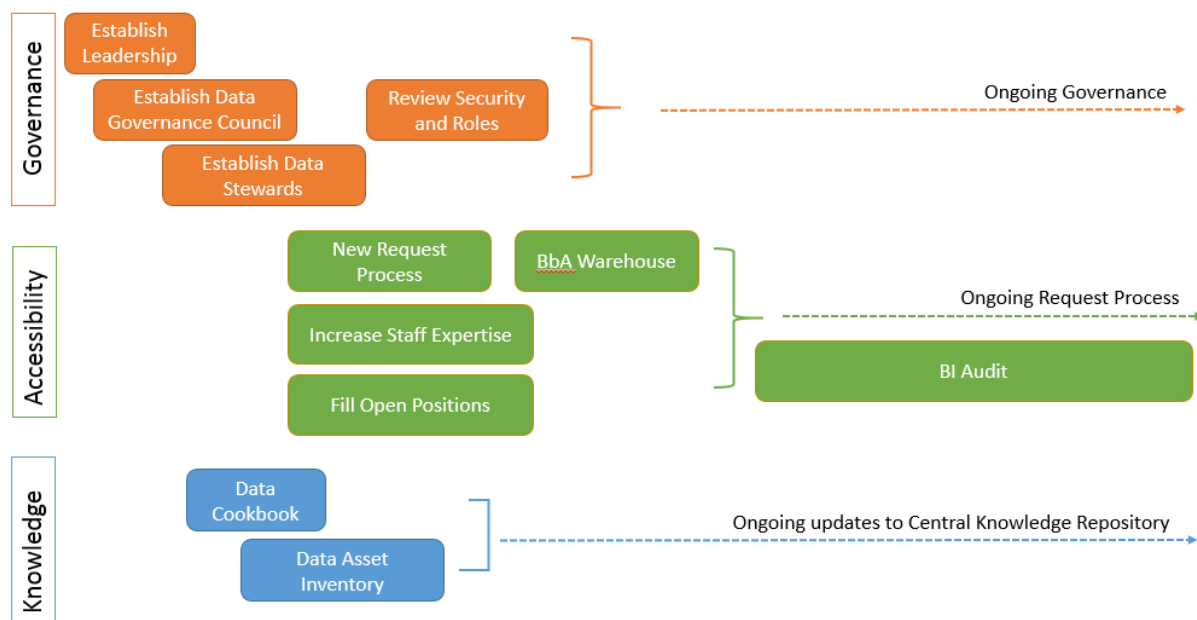
### Data Asset Inventory

Communications & Technology has already identified nearly 250 systems on campus, many of which transmit data back and forth. However the extent to which these systems are integrated, and the data housed within each, does not seem fully known or documented. This will likely be an ongoing effort as new tools are purchased and as new reporting and analytics needs emerge. Issues identified during this inventory, such as shadow systems or data siloes, will be escalated for review by the Data Governance Council.

At a minimum this inventory should include data systems, reporting tools, and data integrations. A template is provided in Appendix C.

## Recommendation Roadmap

Timeline and decisions on which to execute first will be made after presenting these recommendations.

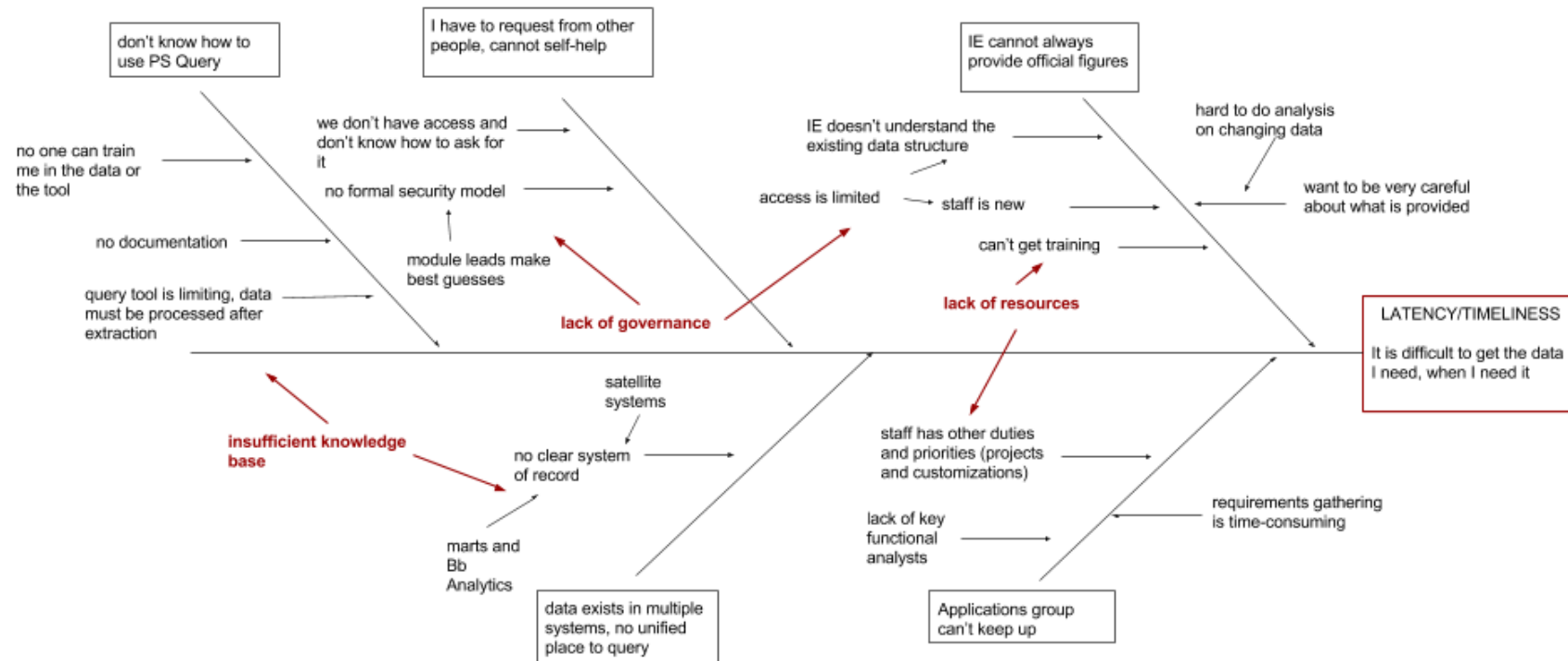


- 1) Governance work drives all the other recommendations. Establish executive sponsor for Data Governance, set up Council/Committee, identify and begin to train data stewards.
- 2) Data stewards and report developers begin to establish central repository of reports using the Data Cookbook, and the business glossary that goes with them.
- 3) The data asset inventory and review of user roles & permissions runs in concert with but begins after the central repository. Cookbook work isn't IT; data asset & roles review is mostly IT.
- 4) Fill open positions and embark on training & knowledge program for current staff (especially data stewards and specification requesters).
- 5) Establish/reconfigure central report request process. IRPE and Application Services divvy up reporting responsibilities and make those responsibilities public. Work to make existing queries easier to access & execute.
- 6) Push to complete the implementation of the warehouse, and use it as "official" source where feasible. Use frozen data sets for longitudinal research and not operational reporting, not even operational analytics.

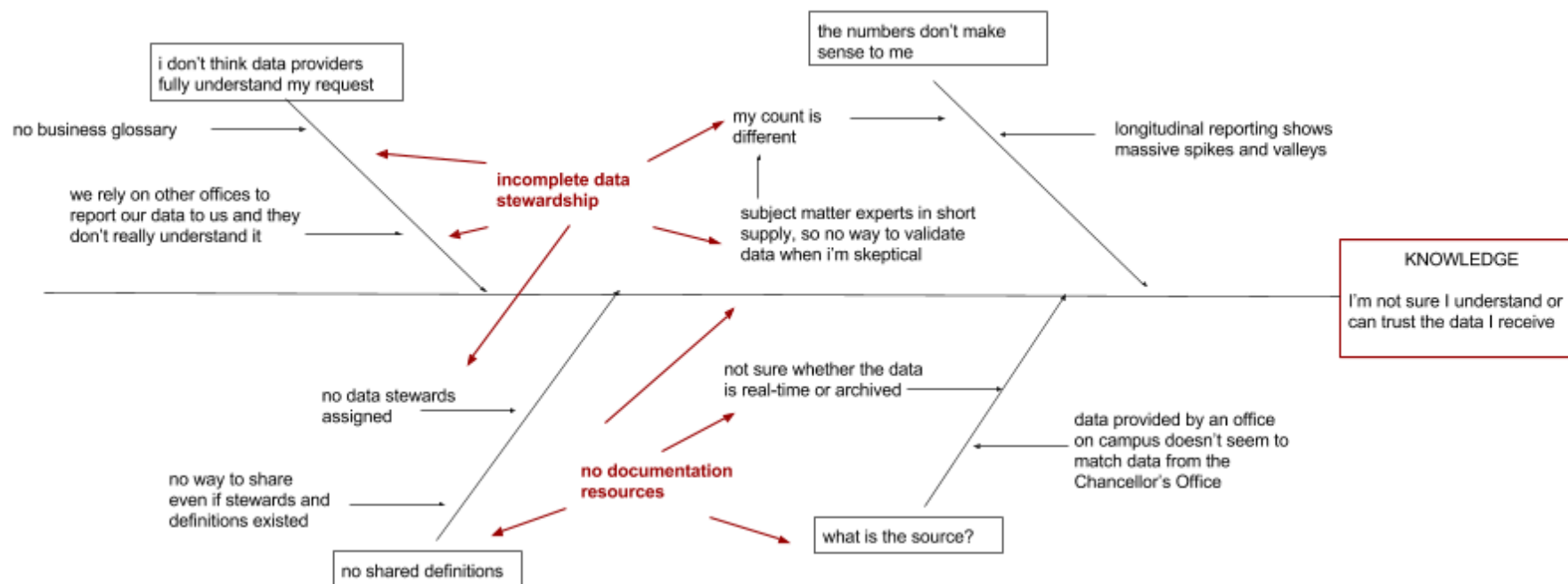
- 7) BI audit - Define reporting & analytics needs as part of “one voice” strategy; begin to define additional needs by department/division.

## Appendix A: Fishbone Diagrams

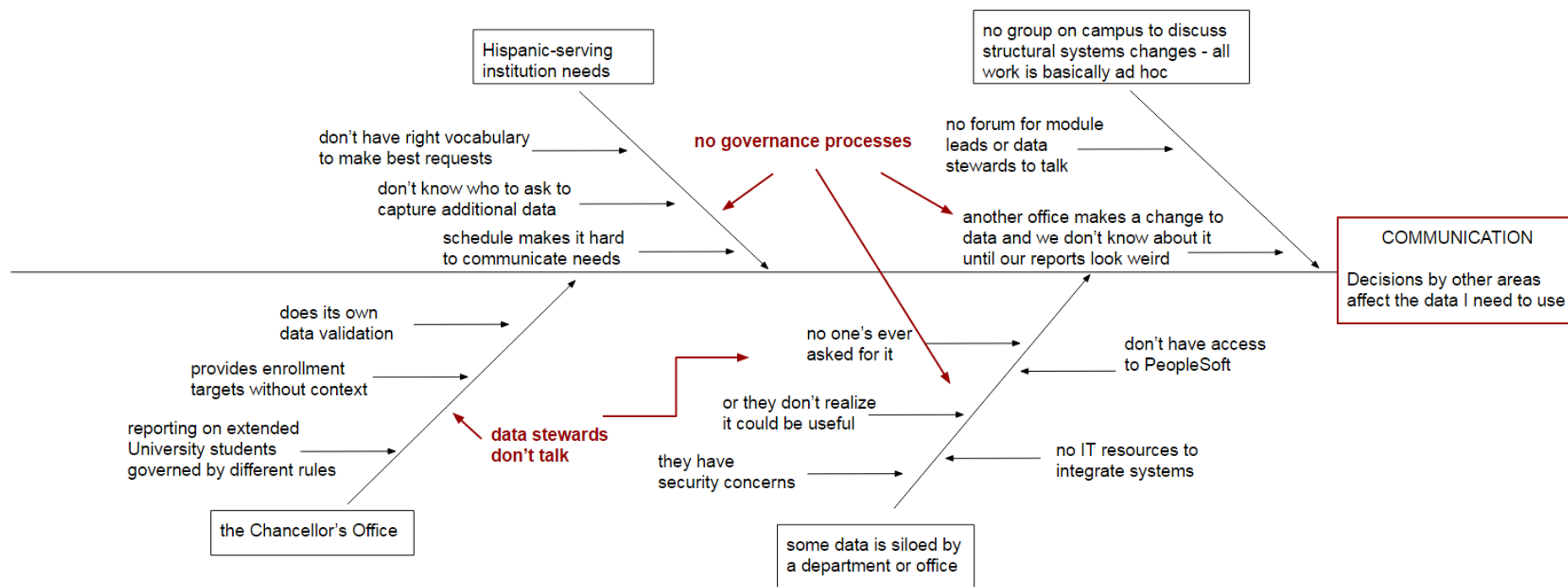
### LATENCY



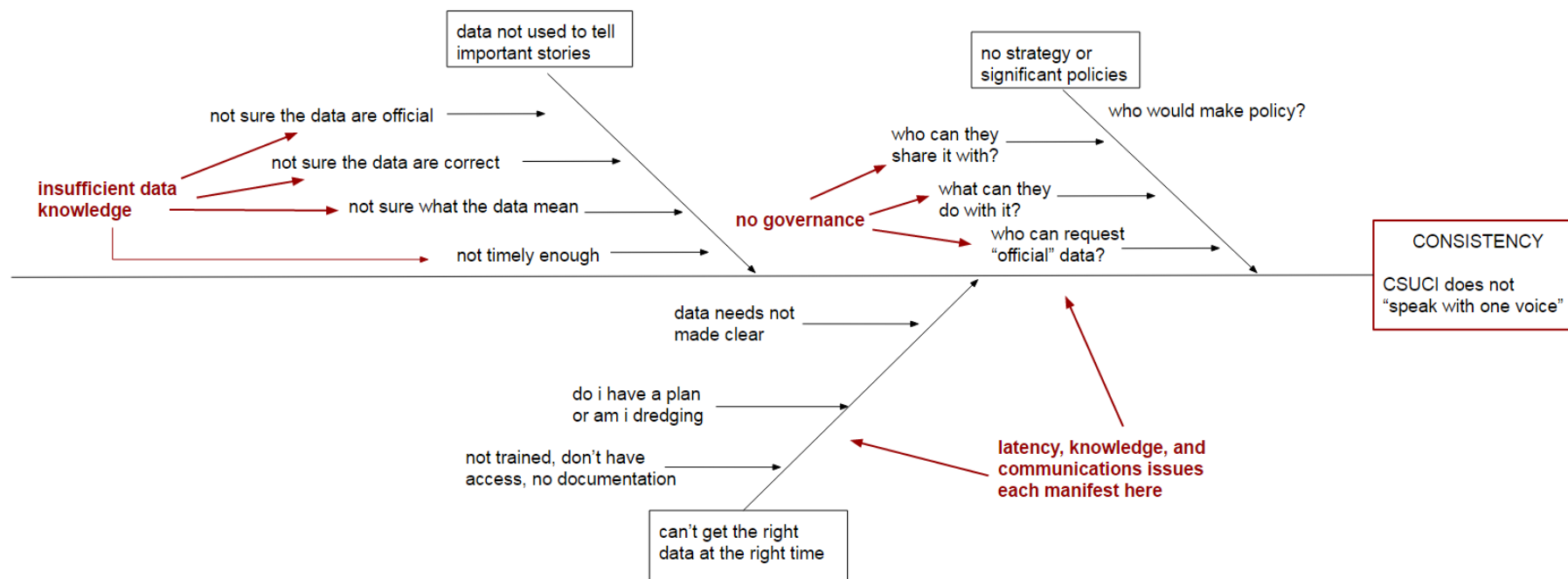
## KNOWLEDGE



## COMMUNICATION



## CONSISTENCY



## Appendix B: Interview Schedule

The following interviews were held during the on-site visit by Scott Flory and Aaron Walker (IData) in September 2015.

Name	Title	Date/Time
<b>Ginger Reyes</b>	Assistant Vice President for Enrollment Management and Director of Admissions and Recruiting	9/15, 10:30am
<b>Hung Dang</b>	Associate Vice President for Enrollment Management	
<b>Ana Rosa Duran</b>	Director of Student Systems	
<b>Amanda Quintero</b>	Director, Hispanic Serving Institution Initiatives	9/15, 11:30am
<b>Phil Hampton</b>	Professor of Chemistry	
<b>Marie Francois</b>	Director of University Experience	
<b>Cindy Wyels</b>	Professor of Mathematics	
<b>Harley Baker</b>	Professor of Psychology	
<b>Dan Wakelee</b>	Associate Provost	9/15, 12:30pm
<b>Andrea Skinner</b>	Academic Scheduler	
<b>Sunshine Garcia</b>	Director of Financial Aid and Scholarships	9/15, 1:30pm
<b>Susan Allison Arias</b>	Associate Director of Financial Aid and Scholarships	
<b>Nick Fuentes</b>	Director of Operations, Extended University	
<b>David Carlson</b>	Associate Architect	9/15, 2:30pm
<b>Jill Leafstedt</b>	Executive Director, Teaching and Learning	9/15, 4:00pm
<b>Nancy Gill</b>	Director of Communications and Marketing	
<b>Herb Aquino</b>	Manager of IT Infrastructure	
<b>Michael Berman</b>	Vice President, Technology and Communication	
<b>Neal Fisch</b>	Director Enterprise Services and Security	
<b>Carlos Miranda</b>	Manager of User Services	
<b>Peter Mosinskis</b>	Director of IT Strategy	
<b>Peter Scamardo</b>	Operating Systems Analyst	
<b>Nasser Mansour</b>	Software Support Analyst	
<b>Missy Jarnagin</b>	Associate Vice President, Financial Services	9/16, 9:30am
<b>Lacey Lovejoy</b>	Director of Financial Systems	



<b>Rachel Linares</b>	Manager, Student Business Services	
<b>Anna Pavin</b>	Associate Vice President, Human Resources	9/16, 10:30am
<b>Diana Enos</b>	Human Resources Manager	
<b>Sherie Frame</b>	HRIS Coordinator	
<b>Neal Fisch</b>	Director Enterprise Services and Security	9/16, 1:30pm
<b>Hung Dang</b>	Associate Vice President for Enrollment Management	9/16, 2:30pm
<b>Ana Rosa Duran</b>	Director of Student Systems	
<b>Michael Bourgeois</b>	Director of Institutional Effectiveness	
<b>Genevieve Evans-Taylor</b>	Chief of Staff	9/16, 3:30pm
<b>Melissa Remotti</b>	Director of Special Projects and Operations	
<b>John Reid</b>	Director of Public Safety and Chief of Police	
<b>Michael Morris</b>	Lieutenant	
<b>Stephen Stratton</b>	Head of Collection Development and Technical Services	9/17, 9:00am
<b>Amy Wallace</b>	Associate Vice President, Library	
<b>Jason Miller</b>	Senior Research Office	
<b>Jeanne Grier</b>	Chair, Academic Senate	
<b>Tina Knight</b>	Interim Director, Sponsored Programs	
<b>Karen Carey</b>	Associate Vice President, Arts & Sciences	9/18, 10:00am
<b>Michael Bourgeois</b>	Director of Institutional Effectiveness	9/19, 11:00am
<b>Bob Ngo</b>	Senior Analyst	
<b>Judy King</b>	Data Analyst	
<b>Guy Timpanaro</b>	Expert Analyst/Programmer	
<b>Jennifer Miller</b>	Director of Student Affairs Assessment	9/19, 1:00pm
<b>Cindy D'Errico</b>	Executive Director, Housing and Residential Education	
<b>Kirsten Moss</b>	Director for Multi Access Programs	
<b>Kelly Macias</b>	Manager, Application Services	9/19, 2:30pm
<b>Joseph Dobzynski</b>	Expert Analyst/Programmer	
<b>Nasser Mansour</b>	Software Systems Specialist	
<b>Shawn Bochat</b>	Expert Analyst/Programmer	

<b>Wayne Goodwin</b>	Expert Analyst/Programmer	
<b>Angela Stockmon</b>	Analyst/Programmer	
<b>Tom Froelich</b>	Advancement Database Specialist	9/19, 3:30pm
<b>Eva Gomez</b>	Director of Annual Giving & Special Gifts	
<b>Christopher Abe</b>	Director, Advancement Operations	
<b>Tania Garcia</b>	Director of Development, Alumni Relations	
<b>Nichole Ipach</b>	Vice President for University Advancement	

## Appendix C: Data Systems, Reporting Tools, Integrations

The following tables can serve as a template for an inventory of data systems, reporting tools and integrations.

### Data Systems

Data Systems	Description

### Reporting Tools

Reporting Tools	Description/Use

### Integrations

Source System	Output System	Notes

## Appendix D: Data Governance Roles and Responsibilities

### Executive Sponsor

The Executive Sponsor is responsible for the success of the project and ongoing data governance processes that follow. The person in this role must have a good understanding of the importance of data and the impact it has on the success of the organization. This role does not require subject matter expertise for any particular functional area(s), but a high-level understanding of operations is important.

### Level of Effort

- Level of effort for this role may be 2-4 hours per week, but while implementing recommendations, this may be more. The actual hours will be based on the project durations and data governance processes established during the project.

### Responsibilities

- Chair the Data Governance Council
- Make sure the project goals, data governance processes, and institutional strategies are in alignment. Focus the team's work so that it has a positive impact on the institution.
- Help secure funding, support and political backing
- Help to determine measurable success indicators
- Connect with a broad network of relationships
- Help identify and overcome obstacles and resistance within the institution
- Support project team(s) with communication and visibility

### Qualities

- Passion for the project
- Ability and profile to champion data governance and governance project teams
- Strong communication skills
- Able to see the big picture but can also identify specific challenges
- Motivational to the team
- Accountable
- Can manage risk

## Project Leader

The Project Leader provides guidance, instruction, direction and leadership to the project team(s). They are to see that tasks are completed by using all of the assigned resources. This role has to have a good understanding of the importance of data and the impact it has on the success of the College. Like the Executive Sponsor, particular subject matter expertise is not a requirement.

### Level of Effort

- Level of effort for this role may be 4-6 hours per week, but while implementing recommendations, this may be more. The actual hours will be based on the project durations and timelines.

### Responsibilities

- Manage the resources on the project team(s). Negotiate the tasks of the team members in conjunction with their managers. Make sure the project is a priority
- Facilitate team interactions
- Monitor team participation and create reports to update the institution on progress.

### Qualities

- Communication skills
- Collaboration skills
- Facilitation skills
- Organizational skills
- Must be influential and able to wield influence
- Delegation skills
- Negotiation skills
- Respected and fair
- Project Management skills

## Appendix E: Glossary

**Analytics:** the discovery and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance. Analytics often favors data visualization to communicate insight.

**Business Intelligence (BI):** tools and techniques for transforming raw data into actionable information. Occasionally one will hear reference to a “BI stack,” which comprises transactional databases, integrations, one or more reporting databases, and analysis tools.

**Data Fluency:** a term we have borrowed from Juice Analytics and find preferable to “data literacy” (although that term may appear in this document and will be used regularly by IData staff). Data fluency refers both to the ability for individuals to communicate with and about data and to an organizational culture that recognizes and acts on data insights.

**Data Governance:** a set of policies, processes, and procedures that govern or administer data. Formal data governance would be characterized by written documentation, managerial controls, assigned responsibilities and accountabilities, and so on. Governance is typically thought of as a subset of data management, but it is a critical component of data management practices.

**Data Management:** the Data Management Association (DAMA) defines this as “the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.” A catchy phrase for what data management covers is “from acquisition to disposition.”

**Data Warehouse:** a simplified (“flattened” or “denormalized”) database optimized for improved performance and ease of use in reporting, data extraction.

**Decision Support:** generic term for the process of providing information to decision makers. More recent terminology for this process includes “evidence-based decision making” or “data-driven decision making.” In theory, data could be internal and/or external to the organization, and could be quantitative as well as qualitative. “Evidence-based” implies a greater level of empirical and/or statistical rigor than “data-driven,” but this varies by use.