

Version 1.1

## Business Intelligence Strategy

Developed by the Business Intelligence Community of Practice

*Co-chairs:* Jay Eckles, Ed.D., and Dennis Hengstler, Ph.D.

*Tools subcommittee chair:* Ed Johnson

*Policy, Standards, and Definitions subcommittee chair:* Allen Dupont, Ph.D.

*Training and Support subcommittee chair:* Denise Haley

*Resources subcommittee chair:* Ron Loewen

*Members:*

Shawn Bryan, UT Health Science Center  
Mike Ebbs, UT Health Science Center  
Lisa Ford, UT Foundation, Inc.  
Kriss Gabourel, UT System Administration  
Denise Gardner, Ph.D., UT Knoxville  
Scott Gordy, UT Institute for Public Service  
Jeanne Hermann, Ph.D.,  
UT Health Science Center  
Janice Hodge, UT System Administration

Tom Hoover, UT Chattanooga  
Les Mathews, UT System Administration  
Desiree McCullough, Ph.D., UT Martin  
Cyndie Nichols, UT Institute of Agriculture  
Kristen Noblit, UT Chattanooga  
Steven Robertson, UT Martin  
Mark Savage, UT Knoxville  
John Toman, Ph.D.,  
UT Institute of Agriculture

*Submitted December 16, 2015 to the Statewide IT Committee*

Contents

- Executive Summary..... 4
- Estimated Resource Needs ..... 4
  - Low investment option details ..... 5
  - Medium investment option details ..... 6
  - High investment option details..... 6
- Goals ..... 7
- Summary of strategy..... 7
  - Procedural tactics ..... 8
    - Coordinate data access and stewardship ..... 8
    - Catalog metadata..... 8
    - Decentralize report authoring with coordinated verification ..... 8
    - Educate ..... 9
    - Training and Support..... 9
  - Technical tactics..... 10
    - Standardize tools ..... 10
    - Centrally consolidate data ..... 11
- Justification ..... 11
  - Accuracy..... 11
  - Functionality ..... 11
- Proposed BI Platform..... 13
  - Diagram..... 13
  - Components..... 13
  - Roles..... 14
  - Data Flow ..... 15
  - Access to each component of the BI technical platform ..... 15
  - Creation and maintenance procedures ..... 16
    - Data Standards creation and maintenance ..... 16
    - Metadata Catalog creation and maintenance ..... 16
    - Semantic Layer creation and maintenance ..... 16
    - ETL/Data Quality creation and maintenance ..... 16
- Proposed Data Access and Stewardship process..... 17
- Report Verification process ..... 17

- Outline of Implementation Plan ..... 17
  - Phase 1: Selling and Preparing..... 18
  - Phase 2: Initial Data Gathering ..... 18
  - Phase 3: Tool Implementations ..... 19
  - Phase 4: Initial Access Approvals ..... 19
  - Phase 5: Initial Training..... 19
  - Phase 6: Initial Report Verifications..... 19
  - Phase 7: Data Standards Maturation ..... 19
  - Phase 8: Data Quality Checks..... 20
  - Phase 9: BI Strategy Review ..... 20
- Glossary..... 21
- References ..... 23

## Executive Summary

*Business intelligence* (BI) is an umbrella term that refers to a variety of software applications used to analyze an organization's data. BI as a discipline is made up of several related activities, including querying, reporting, exploring data, predictive analytics, data mining, and data visualization.

- This document describes a strategy for a **university-wide business intelligence** (BI) system.
- This strategy will make it possible for campuses and UT System Administration (UTSA) to develop **reports that cross the boundaries of individual enterprise systems** (like HR, Finance, Student, Research, Outreach and Engagement, Advancement, and Facilities) **and cross the boundaries of individual campuses and institutes**.
- The BI strategy proposed herein is supported by three major components:
  - A UTSA-administered centralized data repository
  - A Report Verification Process
  - A Data Access and Stewardship Process
- The goals of the BI strategy are to improve the **consistency, accuracy, comprehensiveness, and timeliness** of information delivery to decision makers.
- This BI platform is intended to be a service that meets UTSA and campus/institute needs, but is **not intended to completely replace or preclude locally-developed solutions**.
- This strategy calls for a central data repository, but unlike past plans, establishing this repository does not necessarily require a large expenditure. This strategy **will accommodate low, medium, or high levels of investment**.
- The plan is to **start small and grow** rather than to start big, so as to require fewer resources and make resource planning easier.
- This strategy will take time to fully implement, but **small wins will come very quickly and will continue** throughout the implementation.
- The two most difficult aspects of implementing this strategy will be **negotiating common definitions** and coordinating system owners to provide **access to the data** housed in those systems.
- **All campuses and institutes will have access** to view data in this university-wide BI platform (controlled by security mechanisms, of course).
- This strategy differs from previous BI plans because it **can be accomplished using many resources already on hand**. The reduced resource needs are the result of a **simpler design** and because we are **not attempting to replace existing functionality** in current BI systems (like SAP BW), only **adding cross-function and cross-campus reporting capability**.

## Estimated Resource Needs

This strategy will accommodate low, medium, or high investment levels. Details about those three options are presented here.

It is worth noting that these investment options build upon one another. That is, choosing to start at the low investment option does not preclude moving to the medium and then to the high investment options in the future. Investments made at each level will be reused in the next higher level.

All of these options appear at this time to require two additional permanent employees (already on staff).

Here is a comparison of the options:

	Low investment	Medium investment	High investment
<b>First year cost</b>	\$70-75k+	\$600-750k+	\$1.5-2 million+
<b>Subsequent annual cost</b>	\$70-75k+	\$200-250k+	\$300-400k+
<b>Scale</b>	Small	Small	Large
<b>Growth</b>	Slow	Slow	Rapid
<b>Implementation Speed</b>	Quick	Quick	Slow
<b>Training</b>	Limited	Moderate	Extensive onsite
<b>Functionality</b>	Cross functional and cross campus reporting	New reporting, visualization, and predictive tools Enhanced cross functional and cross campus reporting	New reporting, visualization, and predictive tools Greatly enhanced cross functional and cross campus reporting Sophisticated ad-hoc reporting tools Consulting services for implementation and data gathering

Table 1: comparison of investment options

### Low investment option details

What it costs	What you get
<b>\$70,000-75,000+ annually</b>	A small and slow approach to building cross-campus and cross-functional reporting and analysis using software tools already in use at UT, as well as some on-site training.

If there is a desire to minimize financial investment to execute this strategy, it is possible to work with mostly existing tools and licensing, though we will have limited ability to provide ad-hoc analysis, visual analysis, and automatically updating dashboards. Collaboration will likewise be limited and development will be slower. The minimum annual investment required to execute this strategy is approximately \$70,000 per year. Confidence in the accuracy of this number is high given the few items needed for this limited option and the clarity of pricing for products for which we have existing contracts and agreements. This low investment option is comprised of:

- Approximately \$30,000 per year for database servers and storage
- Approximately \$16,000 per year for extra licensing of currently contracted database and reporting tools
- Approximately \$25,000 for training

Medium investment option details

What it costs	What you get
<b>\$600,000-750,000+ in year 1, \$200,000-250,000+ annually thereafter</b>	A small and slow approach to cross-campus and cross-functional reporting and analysis using more functional software. More capable tools for loading data more quickly, and new reporting, visualization, and predictive analysis tools with more functionality than what UT currently uses. Some onsite training.

If additional resources are available beyond a minimum investment in business intelligence, we could invest those additional funds in more functional and faster tools to extract and transform data plus more functional reporting, visualization, dashboard, and analytic tools. This medium level investment is approximately \$600,000-700,000 in the first year, \$200,000-250,000 or more annually thereafter. Confidence in the accuracy of these figures is medium because tools will have to be bid out; educated guesses based on industry standards and published pricing lists are employed in this estimate. The medium investment option is comprised of:

- All items from the low investment option
- Approximately \$9,500 for licensing of a new metadata catalog tool
- Approximately an additional \$40,000 per year for application servers and storage
- Approximately \$100,000 annually for licensing of a new extract-transform-load (ETL) tool
- Approximately \$500,000 in the first year for licensing of new reporting tools, particularly new or expanded licensing for visualization tools as well as licensing for a statistical package

High investment option details

What it costs	What you get
<b>\$1.5-2 million+ in year 1, \$300,000-400,000+ annually thereafter</b>	A large and rapid approach using faster data loading software, more functional reporting, visualization, and predictive software, sophisticated ad-hoc reporting tools, and consulting services to quickly ramp up available data and reporting. Extensive onsite training.

If the university’s desire is to build a world-class business intelligence platform, expect to invest approximately \$1.5-2 million in the first year, \$300,000-400,000 or more annually thereafter. Confidence in the accuracy of these figures is very low because both tools and services will have to be bid out; educated guesses based on industry standards and published pricing lists are employed in this estimate, but levels of consulting needed will be unclear until a discovery phase is completed. The high investment option is comprised of:

- All items from the low and medium investment options
- Approximately \$500,000 in consulting services to
  - build out data quality services including automated data cleansing based on machine learning (approximately \$250,000)

- quickly acquire vast swaths of data in the consolidated data repository (approximately \$100,000)
- quickly build out the semantic layer (approximately \$100,000)
- Approximately \$250,000 for licensing of new online analytical processing software
  - Approximately \$150,000 for consulting services to implement the software
  - Approximately \$100,000 to build cubes based on the semantic layer
- Approximately \$100,000 for customized on-site vendor provided training delivery

## Goals

The goals of the BI strategy are to improve support for decision making by making improvements in each of the following areas:

- Consistency – providing a means of knowing when data in one report may be expected to match data in another report
- Accuracy – ensuring that data presented to decision makers is as accurate as possible
- Comprehensiveness – allow data from multiple enterprise systems (Finance, HR, Student, Research, Outreach and Engagement, Advancement, Facilities, etc.) and multiple institutes/campuses to be combined and queried together
- Timeliness – dramatically reducing manual labor required to produce reports that cross system or institutional boundaries
- Efficiency – less time cleansing data and preparing it for report and presentation
- User experience – improving accessibility to data -- including finding data – and improving clarity and effectiveness of use and presentation of data

## Summary of strategy

The basic strategy is two-fold: procedural and technical.

- The procedural portion of the strategy is the establishment of a data access and stewardship process and a report verification process.
- The technical portion of the strategy is adopting standard reporting and metadata tools, the central consolidation of institutional data and definitions, and creating automated data quality checks.

Centralization of data and definitions, decentralization of access approval, and decentralization of report authoring are the key aspects of this strategy.

Approval to access data will be accomplished according to roles and responsibilities in a Data Access and Stewardship Process that is to be developed.

There will be no central office solely responsible for authoring reports, queries, and other BI assets. UTSA's Institutional Research (IR) office will have a large role in defining reports and queries for UTSA executives, but campus departments and units will have the definitions, tools, access, and training to write reports that are useful on their campuses.

The tactics for achieving the goals of the strategy are as follows:

## Procedural tactics

### Coordinate data access and stewardship

A more formal and transparent data access and stewardship structure needs to be put in place. This structure would at a minimum include the following:

- Identification of roles, and the individual who should hold those roles, and responsibilities surrounding the management and stewardship of institutional data
- Procedures for requesting and granting access to data, including reports and systems
- Procedures for ensuring data quality, consistency, and definitions
- A mechanism for requesting and prioritizing BI work, such as new reports and queries and the inclusion of new data sources where necessary
- A means of addressing concerns about control, use, and access to data
- Appropriate levels of engagement for university leadership in the data stewardship process

### Catalog metadata

A metadata catalog, alternatively referred to as a data dictionary, will help resolve issues of inconsistency, poor and poorly enforced data standards, and address the operational variations among campuses and institutes.

Having a metadata catalog that is maintained and used by report authors would ensure that report authors utilize common definitions when writing reports. Theoretically that means a report on enrollment from UT Martin could contain comparable data to an enrollment report from UT Chattanooga. It also means that two different report authors at UT Knoxville who construct two research expenditure reports would be using the same definitions for terms like *proposal*, *award*, etc.

In practice, definitions for terms are likely to be different at different campuses. For example, the definition of *student* might differ at UT Chattanooga versus UT Health Science Center because of the different levels and types of education offered at each campus. Reliance on a metadata catalog would require the report authors at each campus to define what *student* means in their contexts and label those different definitions. Forcing those different labels where meaningful differences truly exist will help readers of reports know that each campus truly means something different in otherwise similar-looking reports.

### Decentralize report authoring with coordinated verification

Finding and implementing software that allows report and query development at the campus level will be important. Doing so will allow the campuses to take advantage of the consolidated data repository in a way that most closely aligns with their individual needs. This will significantly improve campus and institutes' influence over BI prioritization.

If a process is put in place to verify reports and queries built from the centralized repository, then those reports and queries could be shared with other campuses and institutes, reducing everyone's load. This verification process would need the following components:

- Education of report and query consumers, teaching them what verification means
- Education of report and query authors, teaching them the value of verification and how to go through the process.



- A means of negotiating definitions, calculations, criteria, etc. among all campuses and institutes at the UT System level or among all stakeholders at the campus or institute level, including recognizing legitimate needs for multiple definitions. This must include subject matter experts.
- A means of documenting negotiated definitions, calculations, criteria, etc. as well as the semantic meaning of a given report or query.
- Testing of the report or query to ensure it operates correctly.
- Validation of the report or query to ensure that it contains correct and accurate data.
- A visual indicator that a given report or query is, in fact, verified.

Once verified, reports and queries need a means by which they can be effectively shared.

This process of verification, particularly the documentation of negotiated definitions, calculations, criteria, semantic meaning, etc. should dramatically improve accuracy and consistency of reports and queries. It will also make possible the construction of menu-driven reports and queries that can be used by non-power-users and can be trusted by decision makers.

The process for negotiating definitions will be detailed in the Data Access and Stewardship Process.

#### Educate

We need to build a communication plan that emphasizes the value of data as an asset to the institution. Such a communication plan would need to identify stakeholders in data and their interest in data. The communications emerging out of this plan would seek to persuade stakeholders that the following assertions are true:

1. That institutional data belongs to the university as a whole, not to any individual, department, or campus;
2. That institutional data is an asset – like human resources, money, or physical plant – that can and should be used to advance the University of Tennessee’s mission;
3. That the value of data increases with expanded access to it;
4. That because of its value, data should be managed and safeguarded like other resources.

Similarly, we need to build a communication plan that explains the intended uses of data, particularly data that is centralized at the UT System level. In this case, the stakeholders of particular import are those who enter the data and those who have physical or administrative custody of the data. It is appropriate for these individuals to safeguard privacy of data entrusted to them because it is an institutional asset. Accordingly, they should be assured that sharing data with others will not result in misuse of that data.

In particular, it will be important to assure campus-level stakeholders that their campus will have equal or greater access to their own data than ever before. (It is even more critical to ensure that this improved access is made a reality). Communications should affirm the right of campuses to build supplementary business intelligence resources to address institutional or departmental decision support needs that are not met adequately by any centralized repository or tool set.

#### Training and Support

In order to facilitate decentralized report and query authoring, a training and support process must be put in place.

Both to accommodate participant schedules and because research suggests it is more effective, initial training will take place using a format known as *microlearning*. This format utilizes hands-on classes for users once per week in one or two hour blocks for three to four weeks. This training would encompass both the use of the tools and understanding of data. The training would be conducted on site at all campuses.

To facilitate continued learning and to provide reference material, a website will be established with links to the metadata catalog, downloadable step-by-step instructions, short easy-to-read written help documents, and short (less than 5 minute) videos covering a wide range of topics.

For technical support, we will establish a number of options. For connection and password issues, we will enlist the help of local campus help desks; users already take advantage of campus help desks for these issues for other systems. For more complex issues, we will have a single university-wide point of contact for BI help. That single point of contact will distribute help requests to a wide range of people including the UTSA Information Technology Services (ITS) BI team and trained tool experts at each campus. This single point of contact will be accessible by phone, and issues will be tracked and managed via a ticketing system.

The BI website will, in addition to the training material already listed, contain links to reports that have been shared with the general community, FAQs, and an easy search feature.

## Technical tactics

### Standardize tools

A standard set of presentation-layer tools will be adopted. Presentation layer tools are those software applications that allow one to create and view reports, queries, visualizations, dashboards, and analyses.

Standardizing tools will create the ability for Report Consumers and Report Authors to support one another across campuses and institutes. It will greatly improve users' ability to share reports and queries, potentially easing the load on each individual institution when one campus's solution may be applied to another's similar problem.

The UTSA-ITS and IR offices will be able to provide greater shared support for the tools because they are the same tools used by system administration and by every institution.

New standardized tools also offer the opportunity to bring in new desired features and functionality. There is a lot of interest in drill-down capabilities, slice-and-dice capabilities, and attractive presentation tools that make data ready for presentation to executives. Likewise, mobile access to reports can be prioritized.

Purchasing standardized tools can result in financial savings. However, fiscal rewards will be realized only if other individually purchased tools are eventually retired. Without retirement of existing systems, investments into this strategy will be primarily for expanded BI capacity and functionality rather than for cost reductions.

### Centrally consolidate data

Data originating in multiple enterprise systems will be consolidated in a single UTSA-administered repository. Initially this consolidation will be a select amount of data, with a long term goal to have comprehensive coverage.

Without a central consolidated repository of data, the goal of creating reports or queries that join together data from multiple enterprise systems or multiple campuses is technically infeasible. This fact is doubly true for reports and queries that join together data from multiple enterprise systems and multiple campuses simultaneously.

Centrally consolidating data also provides the opportunity to apply quality checks to data, improving the accuracy of what is presented to decision makers.

With a central consolidated data repository, users from multiple departments and campuses/institutes and system administration can all build their reports and queries from the same source. This will dramatically improve consistency of data in multiple reports.

## Justification

### Accuracy

Data inaccuracy can have significant financial, political, and public relations impacts on every UT campus and institute. State funding for the University is determined in part by data we submit to the Tennessee Higher Education Commission (THEC) – these data are inputs into the funding formula. Under the current THEC funding formula, failing to count a single student costs as much as \$3,000. Failing to count a single degree costs as much as \$65,000.

We know that our data are inaccurate. An initial draft of data for submission to THEC from campuses contained an extremely large number of errors. In the transition to our new research data system, we discovered many proposals that do not have sponsors. In preparing an interactive display of outreach and engagement impacts and economic outcomes, we found multiple versions of the word *Knoxville*, which led to finding many more errors impacting the geographic allocation of outcomes.

Creating a centralized data repository with data quality checks provides a unique – and proven – opportunity to identify errors in data, report those errors to the individuals responsible for the data, and fix those errors in the enterprise systems before they are reported and used for decision making.

### Functionality

The proposed strategy would provide to departments, campuses and institutes, and system administration the ability – where none exists today – to generate reports and queries that join together data from different systems. Examples of specific requests for such reports as gathered from a recent stakeholder survey include:

- Enrollment modeling for budget projections that combines student and finance data
- Gift account balances by area with drill down to transaction detail that combines finance and advancement data
- Cost/benefit analysis for instructional processes at the program or course level that combines student and finance data

- Analysis of successful alumni with admission criteria that combines advancement and student data
- Faculty paid by grants and contracts that combines HR, payroll, and research data
- Class scheduling reports that combines student and facilities data

The only way to create any such reports today is to extract data from each system and manually tie the data together in productivity tools such as Microsoft Access or Excel.

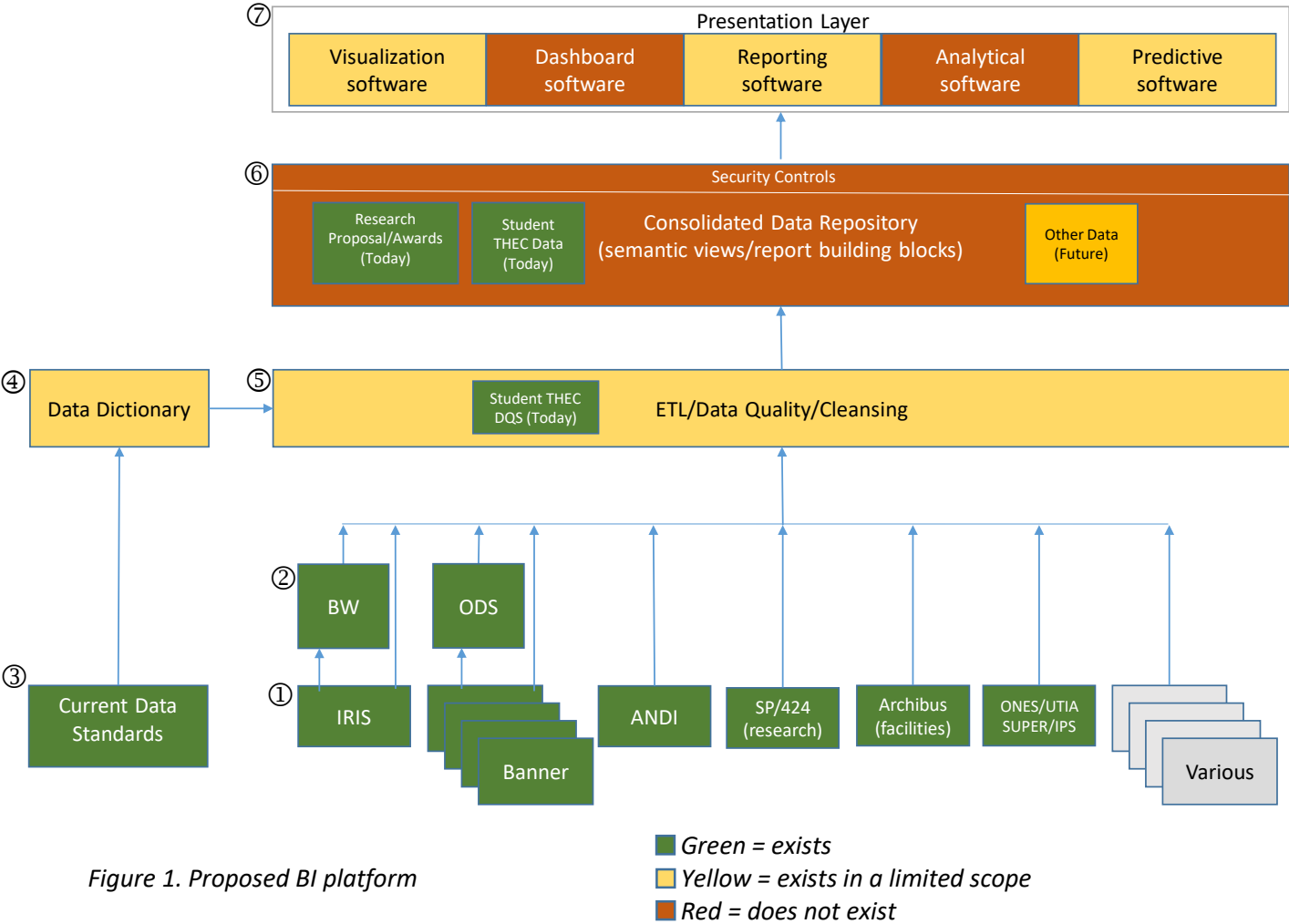
The ability to automate data quality checking will vastly improve the quality of data in the system. Reports generated from the results of these quality checks can also be distributed to data entry and processing personnel and to data stewards, all of whom have a stake in the quality of the data in the source systems.

Finally, new software features may be acquired as a part of tool standardization. Mobile access to business intelligence is an oft-requested feature. Other new features include improved visualization and dashboard capabilities, more attractive reports for presentation to executives, and self-service ad-hoc exploratory data analysis.

### Proposed BI Platform

This section describes the technical system that implements the strategy and tactics proposed in this document. In the diagram below depicting the BI platform, green boxes indicate that the component exists today. Yellow boxes indicate components that exist in some partial state, which is to say that they exist only at some subset of the university, like a department or campus, or that they exist in some limited scope, such as for data in a single enterprise systems (like Banner or IRIS).

### Diagram



### Components

1. Enterprise systems – IRIS, Banner, ANDI, etc., each representing a domain of knowledge (Finance, HR, Student, Advancement, etc.). *Various* represents other systems or future systems. These are the blood of the BI platform.
2. Existing data repositories – These siloed reporting data marts already exist or are in construction and represent legacy business intelligence solutions.

3. Data Standards – Rules establishing what constitutes quality data. Data Standards are a primary source of definitions documented in the Metadata Catalog. It is conceivable that data standards are themselves maintained in the Metadata Catalog tool. These are the morals of the BI platform.
4. Data Dictionary (metadata catalog) –A system that will document definitions of individual data elements (from the Consolidated Data Repository/Semantic Layer), reports, queries, and other BI assets. Reports are constructed using definitions drawn from or added to the Metadata Catalog. Report Consumers may refer to the Metadata Catalog to understand items in reports or other presentations. This is the soul of the BI platform.
5. ETL/Data Quality/Cleansing – A component that extracts data from source systems to load it into the consolidated data repository, potentially running data quality checks along the way for correction in source systems later. This is the heart of the BI platform.
6. Consolidated Data Repository –The UTSA Data Warehouse. Semantic SQL views based on the tables pulled in from source systems via the ETL/Data Quality component. The views join the raw tables in such a way as to intelligently tie together multiple enterprise systems and multiple campuses/institutes together in useful ways. This is the brain of the BI platform.
7. Presentation layer software – the sets of features that constitute the presentation layer may be as few as one software package or as many as five separate software packages. These tools get their data from the Semantic Layer. These are the hands and feet of the BI platform.
  - Visualization software – features that create visual aids for interpreting data
  - Dashboard software – features that create visual indicators suitable for executive and public presentation
  - Reporting software – features that create standard tabular reports that are easily scheduled, distributed, and exported to productivity software like Excel. Provides for drill down capabilities from aggregate data to detail data.
  - Analysis software – features that allow a report consumer to “slice and dice” and explore data.
  - Predictive software – features that allow for the creation of models that predict future outcomes based on past data.

## Roles

- a. Data Stewards – individuals responsible for governing the use of data, including establishing data standards and approving requests to access data.
- b. Report Authors – anyone who uses presentation layer software to create BI assets (reports, queries, etc.) based on Semantic Layer views.
- c. Report Consumers – anyone who uses the output of presentation layer software for decision making or decision support purposes.

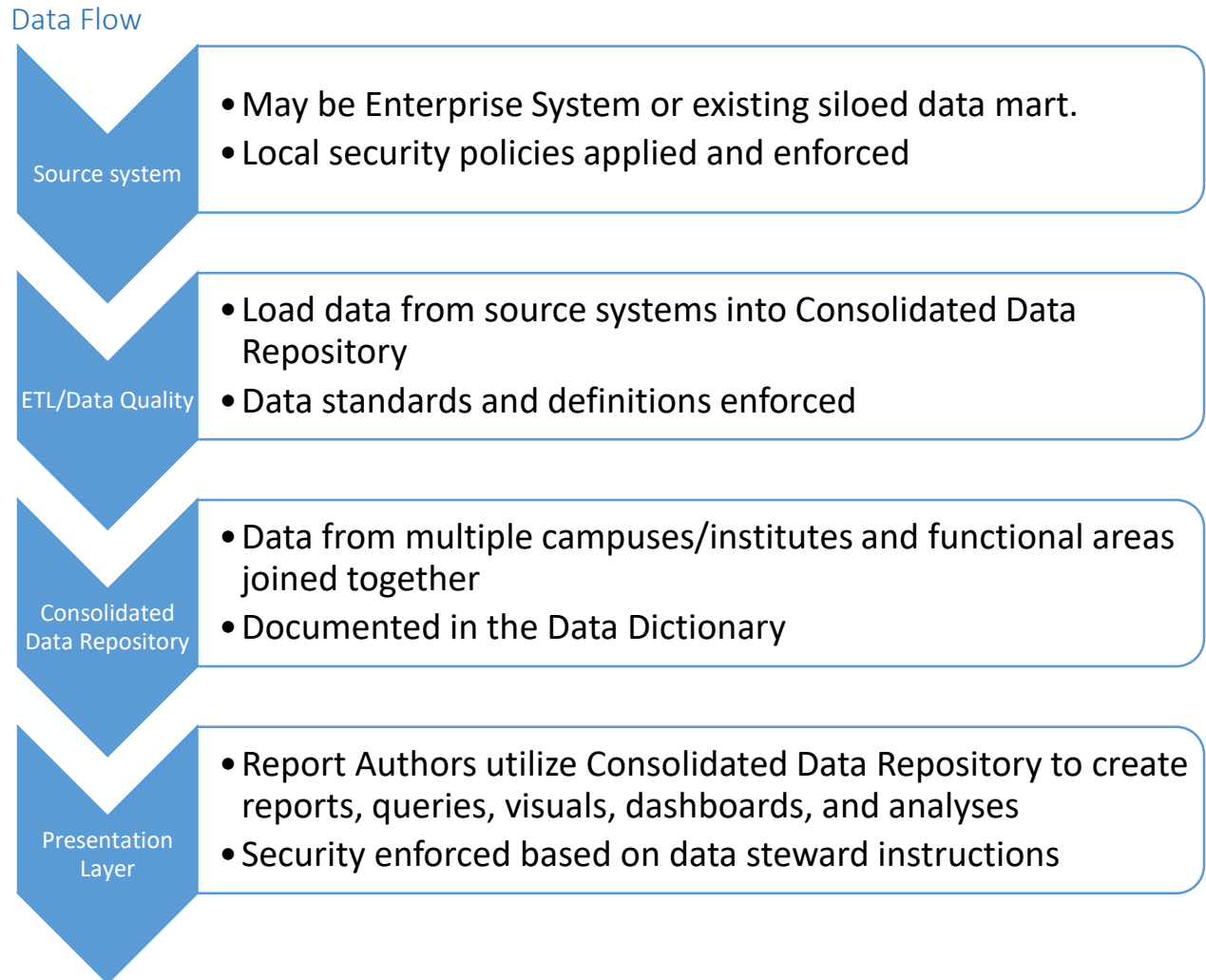


Figure 2. Data flow diagram

#### Access to each component of the BI technical platform

- Source system and siloed data marts – no changes to status quo.
- ETL/Data Quality – UTSA-ITS will build and maintain the ETL and Data Quality tools, Data Stewards will build and maintain the rules checked by the Data Quality tools. Data Quality reports will go to Data Stewards for distribution as appropriate.
- Consolidated Data Repository – Data stewards will dictate access restrictions; UTSA-ITS will have access for development and maintenance and to implement security controls; Report authors (across the university) will have access based on access restrictions.
- Presentation Layer – Report authors (across the university) will be able to build reports, queries, analyses, visualizations, etc.; Report Consumers (across the university) will be able to view the objects created by Report Authors.
- Data dictionary – everyone will have access to view all definitions.
- Data Standards – No change to status quo.

## Creation and maintenance procedures

### Data Standards creation and maintenance

- A data access and stewardship process needs to be created to define this process.
- Data stewards will be responsible for determining data standards for enterprise information systems.

### Metadata Catalog creation and maintenance

- A data access and stewardship process needs to be created to define this process.
- University-wide negotiation will be required to create and modify definitions.
- Data standards will serve as a primary source of definitions.
- Report authors will identify areas where definitions are needed or need to be modified.

### Semantic Layer creation and maintenance

- Creation
  - The initial set of Semantic Layer views will be determined by the BI Community of Practice based on requests from university-wide stakeholders.
  - The initial set will be based on
    - Most easily accessible data
    - Most easily negotiated definitions
    - Most impact on the university
    - Greatest relationship to existing Semantic Layer views (i.e., the Semantic Layer grows outwards from a seed).
  - UTSA-ITS, perhaps with assistance from other IT or IR offices, will implement the initial set of views.
  - Created views will be documented in the Metadata Catalog.
- Maintenance
  - After the initial set of Semantic Layer views are created, requests for new views will be made in the form of requests for definition creation or modification.
    - These requests will be entered into the Metadata Catalog by data stewards or power users.
  - The data access and stewardship process determines how the request is negotiated and disposed. The BI Community of Practice is available to resolve priority conflicts.
  - UTSA-ITS receives requests that pass through the data access and stewardship process, implementing the appropriate changes in the Semantic Layer.
  - The requested change is documented in the Metadata Catalog.

### ETL/Data Quality creation and maintenance

- UTSA-ITS will partner with campuses to create ETL processes according to Semantic Layer view needs.
  - During periods of low workload, UTSA-ITS will proactively work to build additional ETL processes that bring in data not yet used by the Semantic Layer but which may be used in the future.
  - ETL processes will refresh data at a frequency that is appropriate for the need (daily, weekly, monthly, semesterly, etc.)



- UTSA-ITS, perhaps with assistance from other IT or IR offices, will create Data Quality checking processes as needed according to Semantic Layer view needs and according to available Data Standards.
  - Data stewards or their designees will create the actual rules that will be used in the Data Quality checking processes.
  - Data Stewards or their designees will document the rules, either through Data Standards or in the Metadata Catalog (or both).
  - Data Stewards or their designees will validate the results of Data Quality checking processes before UTSA-ITS puts those processes into production.

### Proposed Data Access and Stewardship process

The strategy described in this document relies upon an established data access and stewardship process. This process is currently under development.

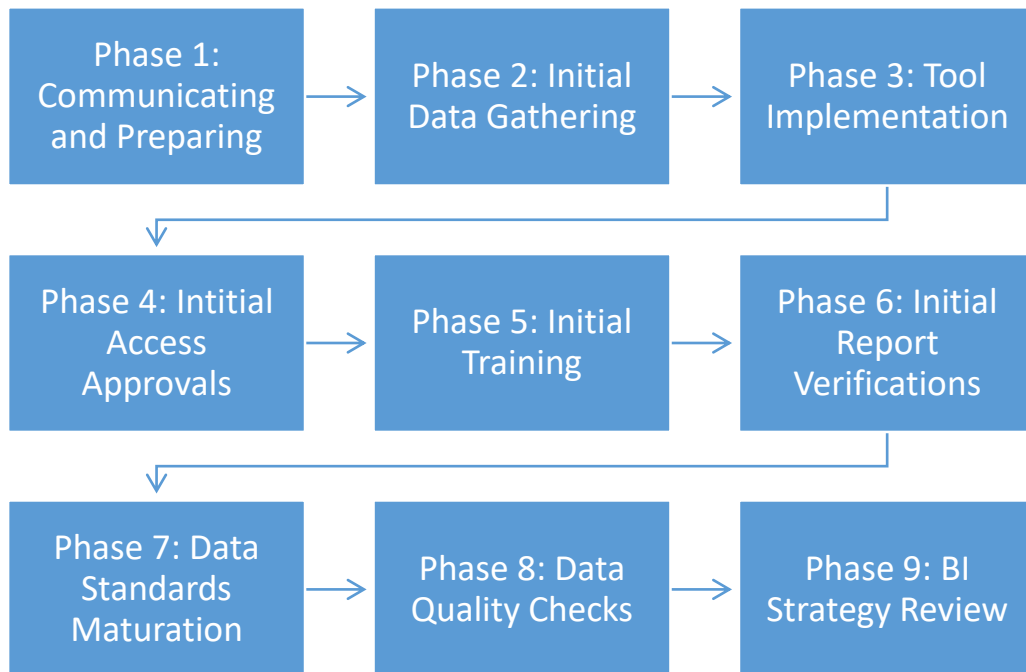
### Report Verification process

As a means of providing confidence in reports and analyses generated out of the BI platform described in this strategy, a report verification process has been developed and approved by the BI Community of Practice. For details, please see the [Report Verification Process](#).

### Outline of Implementation Plan

The implementation of this strategy will use a concurrent and iterative approach.

Concurrent means that several activities will be taking place simultaneously. Iterative means that small pieces of functionality will be developed and deployed on a frequent basis rather than waiting until a large cohesive system is complete.



*Figure 3. Phases of Implementation Plan*

### Phase 1: Communicating and Preparing

The Data Access and Stewardship Process and the Report Verification Process are both key pieces of this strategy and must be established early. In particular, data stewards must be identified so that we have people with whom we can negotiate issues.

Data stewards and owners of information systems will be shown the benefits of sharing data with UTSA for inclusion in the centralized data repository.

While this communication is taking place, the UTSA-ITS BI team must be preparing foundational systems. Among the systems required by this strategy are:

- Metadata Catalog
  - Definition request workflow
  - Report verification request workflow
- Access Request System
  - Data Steward Registry
- Centralized Data Repository database

The metadata catalog will be created first, followed by the access request system. The technical architecture for the centralized data repository database will be established, and then the database will be created.

### Phase 2: Initial Data Gathering

Because of the BI work that has come before the present effort, some requirements have already been gathered and a centralized data repository database can be quickly established. Once that database is created, initial data gathering will begin. Beginning in this phase and continuing throughout the life of

the BI platform, UTSA-ITS BI team will repeatedly execute data gathering iterations. Each iteration will consist of the following:

1. Define a question to be answered, report to be written, or analysis to be performed.
2. Identify data necessary for item 1 that is not currently in the centralized data repository.
3. Negotiate definitions for data identified in item 2 per Data Access and Stewardship Process.
4. Partner with campuses to determine best way to acquire for the centralized data repository data identified in item 2. Acquire the data.
5. Build semantic layer views for the data identified in item 2 per the definitions negotiated in item 3.
6. Document semantic layer views built in item 5.
7. Test semantic layer views built in item 5.
8. Make publicly available semantic layer views built in item 5.

### Phase 3: Tool Implementations

While the first data gathering iteration is underway, UTSA-ITS BI team will work on implementing the selected centralized reporting and analysis tools. Implementation will be done one tool at a time, and implementations will continue concurrent with future phases until all selected tools are implemented. Maintenance of these tools will likewise take place concurrent with future phases.

### Phase 4: Initial Access Approvals

Concurrent with the initial data gathering iteration and the centralized tool implementation, we will ask data stewards to approve initial requests for access to data and tools.

### Phase 5: Initial Training

Once at least one data gathering iteration is completed and at least one reporting and analysis tool is implemented, training will begin according to the plan that will be laid out by the BI CoP Training and Support subcommittee. Initially, this training will be focused on the metadata catalog and using the semantic layer with centralized reporting and analysis tools.

This initial training should be aimed at individuals on campuses who desire to build reports that cross information system boundaries.

### Phase 6: Initial Report Verifications

Once training has been started and at least one report has been created using the semantic layer and centralized reporting and analysis tools, we will encourage a report author to undertake report verification per the Report Verification Process.

### Phase 7: Data Standards Maturation

Once the BI platform is in use, we will encourage the maturation of data standards. We will ask data stewards to identify the data standards documents for data the stewards are responsible for. For those data where standards exist, we will encourage stewards to work towards making sure the metadata catalog is consistent with those standards and that the standards are known to data entry and processing personnel. For those data where standards do not exist, we will encourage their creation.

### Phase 8: Data Quality Checks

As data standards documents are identified and their contents reconciled with the metadata catalog, we will market data quality checking services to data stewards. We will solicit requests for automated data quality checks and move on the first project that has reasonable scope and the available functional resources for rule creation and maintenance.

Beginning with this phase, UTSA-ITS BI team will repeatedly execute data quality service iterations. Once these iterations are reliably repeatable, we will offer this service as part-and-parcel with the data gathering iterations. Each iteration will consist of the following:

1. Identification of data to be quality checked and the relevant standards for those data.
2. Determination of what portion of the standards should be checked in this iteration.
3. Creation of the required technical infrastructure for the data quality checking.
4. Support of functional resources in building rule sets for the quality checking.
5. Testing of the infrastructure and rules.
6. Elevation of infrastructure, rules, and processes to production, including delivery of quality reports to recipients identified by data stewards.

### Phase 9: BI Strategy Review

At this point, the following activities should be a part of the normal course of business at the University of Tennessee:

- Data are regularly gathered for inclusion in the centralized data repository.
- Data quality checks are routinely established, often in concert with data gathering.
- Data standards exist or are being created for all major institutional data.
- Data stewards regularly receive and approve requests for access to data and tools. UTSA-ITS BI regularly provisions access based on these approvals.
- Users of the BI platform are training on the use of the metadata catalog and reporting/analysis tools.
- Users across the university use centralized reporting and analysis tools to access data in the semantic layer of the BI platform. They regularly build reports and analyses that cross information system boundaries and campus/institute boundaries.
- Definitions are created and maintained in the metadata catalog. The definitions are used in the creation of reports and analyses.
- Reports are sometimes submitted for verification.
- Executives trust verified reports over all other reports and analyses.

An assessment of the BI program will seek to determine the degree to which the strategy has succeeded in meeting its goals and what changes need to be made going forward.

## Glossary

**ANDI:** UT acronym for "Alumni and Development Information"; the main information system used by the advancement group.

**API:** Application Programming Interface. Say software package A has an API. A computer programmer creating software package B can use A's API to connect his package B to existing package A. Whenever you visit a retail chain's website and use the store locator feature, if the results look like they're using Google Maps, it's because the site used the Google Maps API.

**Business Intelligence:** An umbrella term that refers to a variety of software applications used to analyze an organization's data. BI as a discipline is made up of several related activities, including querying, reporting, exploring data, predictive analytics, data mining, and data visualization.

**BW:** An SAP product that provides business intelligence functionality for IRIS.

**CBER:** UTK's Center for Economics and Business Research. Home to a business intelligence project called TLDS (Tennessee Longitudinal Data System).

**CIO:** Chief Information Officer, the senior most IT officer in a division.

**COEUS:** Sponsored programs data base systems (proposal and award tracking system). This is the system that preceded TERA/PAMS.

**CoP:** Community of Practice. The term used for the governing committees for information technology at the state-wide level for UT.

**Data Mining:** The practice of applying statistical algorithms to large sets of data for the purpose of either describing the data (such as discovering natural groupings) or extrapolating trends (such as enrollment or retention rates).

**Data Visualization:** Presenting data in a visual and non-tabular format (such as graphs and charts) to improve users' ability to understand and interpret the data.

**Data Warehouse:** A large store of data accumulated from a wide range of sources within an organization and used to guide management decisions.

**Database:** A place where data are stored in a structured manner. By comparison, a data warehouse may be one or several very large databases.

**Database Management System (DBMS):** A piece of software that provides management of and access to a database. Common examples are Oracle and Microsoft SQL Server.

**ETL:** An acronym for "Extract, Transform, and Load". It is the set of steps in data warehouse maintenance where data are extracted from their source system (e.g., IRIS), transformed into more useful formats (e.g., codes translated to match a system standard), and loaded into the data warehouse for later retrieval.

**eVisions:** The vendor of the new research management system being implemented at UT in 2015.

**Integrated R3 Information System:** aka IRIS, the finance and HR SAP application suite.

**IPS:** Institute for Public Service.

**IRIS:** aka SAP, stands for Integrated R3 Information System. Provides centralized HR and Financial management for the UT system.

**IR:** Institutional Research.

**Metadata:** Literally, data about the data. Metadata for a field in a report might include the database table and field from which the data is drawn, the type of data (i.e., numeric, date, text, etc.), or the semantic meaning (a narrative description of the field's content).

**Microlearning:** A training format that utilizes hands-on classes for users once per week in one or two hour blocks for three to four weeks .

**Presentation layer or tool:** When talking about a presentation tool, most often one is describing a tool that will provide a visual representation of aggregate data to a high-level audience (i.e., a dashboard). In the discipline of business intelligence overall, the presentation layer describes all tools used to deliver data to end-users and would include query and reporting tools.

**RFI:** Request for Information, a formal voluntary step in the procurement process used to get general information (but not bids or pricing information) from vendors about possible solutions.

**RFP:** Request for Proposal, a formal mandatory step in the procurement process used to get competitive bids from vendors for products and services.

**SAP:** A Germany-based corporation whose main product is the eponymous enterprise resource planning system.

**SAS:** Vendor providing a number of statistical and business intelligence software packages.

**SP/414:** The product name of UT's new research system, produced by eVisions. This is the successor to TERA-PAMS.

**SPSS:** IBM-owned statistical software system. SPSS formerly stood for "Statistics Program for the Social Sciences".

**Strategy:** An answer to the question of how to implement the organization's "business" strategy. In the case of UT, that is the *University of Tennessee System Administration Strategic Plan*. The BI Strategy must address policy, standards, and definitions; tool recommendations; training and support recommendations; resource need recommendations.

**TERA-PAMS:** Tennessee Electronic Research Administration and Proposal and Award Management System. This is the legacy research management system at UT.

**THEC:** Tennessee Higher Education Commission.

**TLDS:** Tennessee Longitudinal Data System, a program managed by CBER and hosted in the Haslam School of Business at UTK.

**UTSA-ITS:** University of Tennessee System Administration Information Technology Services.

## References

The following sources of information were used in developing this strategy.

- Chapple, M. (2013). Speaking the same language: Building a data governance program for institutional impact. *EDUCAUSE Review*, November/December, 14-26.
- Dell, T., and Fullerton, L. (2015). Empower your report builders with an extensive university data dictionary. *Ellucian Live 2015*. Anaheim, CA. April 15, 2015.
- Hengstler, D., and Eckles, J. (2015). BI Stakeholder Survey. University of Tennessee. Knoxville, TN.
- Hengstler, D., High, K., Mathews, L., Perry, J., and Sauceman, J. (2012). Strategic Plan for Improving Data and Enterprise Reporting: Data Listening Tour Report. University of Tennessee.
- Hug, T. (2007). *Didactics of Microlearning*. Waxman Verlag GmbH: Muenster, Germany.
- Kelly, M.C. (2015). The Chief Data Officer in Higher Education. *EDUCAUSE Review Online*. June 8.
- Koger, D.L. (2012). Business Intelligence: Proposed methodology and technology clarified. University of Tennessee. Knoxville, TN.
- Oracle. (2011). An Oracle White Paper on enterprise Architecture -- Enterprise Information Management: Best Practices in Data Governance. Redwood Shores, CA.
- Reynolds, I. (2015). Draft requirements for Request for Proposals for a BI System. Tennessee Board of Regents.
- University of Tennessee. (2012). IT Governance – University of Tennessee. Knoxville, TN.
- University of Tennessee. (2013). Business Intelligence: BI/EDW Proof of Concept (Version 1.0). Knoxville, TN.
- University of Tennessee. (2013). Business Intelligence: Executive Briefing on BI/EDW Strategy (Version 2.0). Knoxville, TN.
- University of Tennessee. (2014). Business Intelligence: BI/EDW Proof of Concept Results (Version 1.0). Knoxville, TN.
- University of Tennessee. (2015). Defining the Future: The UT System Strategic Plan, 2012-2017. Retrieved June 2, 2015, from <http://president.tennessee.edu/strategicplan/plan.html>.
- University System of Georgia. (2014). *Information Technology Handbook*. Section9.0: Data Governance and Management Structure. Athens, GA.