Connecticut Community College System

IR Data Mart Meeting
January 27-28, 2009

Understanding the Analytics Needs of Institutional Research at Connecticut Community Colleges

The Connecticut Community College System (CCC) consists of the System Office and 12 colleges. There is a System-level Director of Institutional Research as well as 12 campus Directors of IR. Three of the colleges are active participants in “Achieving the Dream,” (a national initiative focused on helping more community college students succeed, especially students of color and students from low-income families.) The initiative, launched in 2004, involves a partnership of more than 20 organizations and includes 82 colleges in 15 states. It emphasizes the use of data to drive change that can improve student success at these colleges.

CCC has been a leader in this initiative since its inception in 2006. It is one of seven states to participate in crafting new intermediate and final measures of student progress, working with the other 6 states to design a more complete and accurate way of measuring student performance longitudinally and comparing outcomes to other participating states nationwide. The time has come and the need is pressing for CCC to acquire a BI (Business Intelligence) solution that will enable Directors of Institutional Reporting at the System and individual community college campuses to more easily analyze data and report on performance outcomes related to “Achieving the Dream” and at-risk students. This initiative is of paramount importance to Connecticut and one that is a culmination of three years of thought and design efforts to create a solution that will lead to greater insights into student success.

The custom data mart that CCC has designed incorporates internal data extracts from Banner Student Academics, including immigration data from SunGard’s Sevis Connector, Banner Finance, and Banner Financial Aid. HR/Payroll data extracts and tables are loaded from the State of Connecticut’s Oracle/PeopleSoft HR/Payroll application through the State’s data warehouse, PeopleSoft EPM, into Banner tables. External data extracts are pulled from NSC, the National Clearinghouse (CSV file format), CSSE (Excel format), Department of Labor (tracking graduates’ employment/salaries) and student survey data from Pell (Excel formats) and CSSE (Excel format). Unresolved is how and what data extracts will be available to CCC now that it has recently signed a multi-year contract with Blackboard to host its LMS, Vista. All these data must be current and readily accessible to support timely reporting and analyzing the needs of the targeted student population across the System of 12 colleges.

The purpose of the IRDB and BI Solution is for IR directors to use these tools to assess student performance with a longer term vision of extending the IRDB out to include a system-wide implementation. In future phases, it will focus on higher order analytics, as well as to allow end-users greater access to run standard and ad hoc reports with simple drill down capabilities on day-to-day operational issues like: How many employees do we have? What is the cost of instruction? How do CCC colleges compare to other community colleges/systems vis a vis CSSE and other national performance indicators?

The BI solution must be able to support three distinct types of end-users in a secure, user-friendly environment:

- **Power users** (those used to querying, manipulating, analyzing and reporting and using tools like Excel, SQL queries and SPSS for predictive modeling/report generation).
• **Functional end-users** (those who need to access standard reports, run ad hoc reports and simple queries/drift downs in a user-friendly, self-service environment – a portal or dashboard)

• **Casual Viewers** (the Chancellor, Deans and Academic Chairs) executives with limited time, but a high degree of need-to-know about current data/ institutional performance in a user-friendly portal or dashboard environment

Equally important to the IR Data Mart initiative is to acquire a BI solution that is low-cost, extensible, can be implemented incrementally, adding new data extracts/sources and new groups of end-users where there are demonstrated business needs.

**Architectural Design of the IR Data Mart**

CCC runs a single instance of Banner ERP that includes Banner Student, Finance, and Financial Aid. The CCC requires its 12 colleges to utilize its Oracle/Education, Perkins web survey data, Graduate Survey data, CCSSE, SENSE, and a hosted instance of Blackboard Vista for the LMS solution.

Figure 1 shows the initial dimensions of the proposed data mart that will allow IR directors and their staff to run standard and ad hoc reports, as well as conduct longitudinal analysis of data stored in Banner since 1998.

**Figure 1: IR Data Mart Dimensions**

<table>
<thead>
<tr>
<th>Common Dimensions</th>
<th>Date</th>
<th>Term</th>
<th>Student</th>
<th>Program</th>
<th>Faculty</th>
<th>Course /Section</th>
<th>Transfer College</th>
<th>Placement Test</th>
<th>Student Grad</th>
<th>Fin Aid</th>
<th>Employee</th>
<th>Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Faculty Load</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Admissions</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduation</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer Data In</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Transfer Data Out</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Financial Aid</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Resources</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Surveys</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data Extracts**

Currently, IR runs a variety of extracts from a number of internal Banner applications and external data sources. Figure 2 shows where these extracts are stored – either inside Banner (CCC has both a single instance of Banner and a single instance of an Oracle database) or externally in various Access databases. The core data that supports IR reporting needs are frozen data that are run once or twice a semester. Over time, CCC would like to retire extracted data sets and move to accessing current data (near real time) as well as historical data.
**Student Academics Environment**

In Figure 3 that follows, features the Student Academics, Financial Aid, and HR IPEDS data extracts that comprise the heart of the IR Data Mart design. CCC has an experienced, seasoned IT department that has successfully modified Banner ERP modules over the last 11 years to meet the unique business/reporting needs of the CCC System and its campuses. Their decision to run a single instance of Banner on a single instance of an Oracle database has greatly enhanced their ability to normalize data, ensure data quality across the System and at the campus level, as well as to standardize business practices. This data/technology environment will greatly facilitate the successful implementation of the selected BI solution in that it will reduce the level of complexity while ensuring a ‘single source of truth’.

Still to be addressed is how CCC will access data extracts from Blackboard Vista being hosted by Blackboard. In the past, the previous hosting vendor ran reports for CCC against the database using SQL tools. It will be important for CCC to be able to extract relevant student data to incorporate into standard and ad hoc reports that analyze student success indicators.
**Financial Aid Environment**

At Risk students are likely to rely on federal financial aid to attend college. The Banner Financial Aid module has built-in algorithms to calculate student eligibility. Additionally, students are required to fill out a FAFSA application to qualify for federal (Pell), state (CAPP) and/or college foundation financial aid; this data is extracted from FAFSA and loaded into each student’s Financial Aid Banner record.

It is important for IR to be able to analyze the types of aid, the dates/amount awarded and to track these data against academic progress to better understand and design programs/plan CCC services that support at risk student success. Please refer to Figure 4 that follows.
Extending the IR Data Mart

The current IR Data Mart design/initiative does not include data dimensions or extracts from Banner Finance/AR applications. However, the longer term vision will extend the BI solution to include these data so that CCC IR directors can achieve a 360 degree view of all factors that impact students and institutional performance.

CCC’s Banner Finance/AR environment as portrayed in Figure 5 that follows has been supplemented to import data extracts from the State HR/Payroll application. The Custom System Integration solutions were created and run automatically and daily to limit manual processing, banking activity, collections and System and College AP requirements, which include purchasing card, PeopleSoft Finance AP and AR requirements from a number of state agencies, vendors and financial institutions. This will be an important extension to the IR Data Mart when CCC begins to analyze academic metrics such as cost of instruction.
The HR/Payroll Environment

The State of Connecticut requires that all its state, local and education agencies run Oracle/PeopleSoft HR/Payroll in support of employee-related activities. CCC created a module that integrates PeopleSoft Payroll with the Banner general ledger. CCC also relies on extracts to be delivered through the Oracle/PeopleSoft EPM (Enterprise Performance Management, a pre-packaged BI application and data warehouse for PeopleSoft applications load and store data in the Banner environment.). This data is currently used for reporting and analysis. The Oracle/PeopleSoft EMP is used to create custom reports and ad hoc queries at the colleges and System office.

Longer term, the BI solution that CCC selects will need to extract core data from within the State’s HR application to be able to analyze and report on queries such as cost of instruction as it relates to student access, affordability and success. Please refer to Figure 6 that follows.
Figure 6: HR/Payroll Environment
Types of BI Solutions in the Higher Education Market

One of the challenges in selecting an appropriate BI solution that meets both CCC’s immediate and longer term reporting and analytics needs is that there are a number of solutions in the market today. Solutions range from a simple reporting tool to robust toolkits to packaged applications that combine data warehouse and are offered by mega-vendors, by functional specialty companies, by professional services organizations, and by ‘next gen’ start-ups. Figure 7 provides a generic list of the BI/analytic solutions in the market today from vendors serving higher education, briefly describes nature of the offering, as well as the advantages and limitations of each. This is not meant to be a comprehensive list, nor does it attempt to evaluate the relevance to CCC’s IR Data Mart initiative. Rather, it is intended to describe the BI landscape in Higher Education today.
### Figure 7: Vendor Solutions in the BI/Higher Education Marketplace

<table>
<thead>
<tr>
<th>Types of Solution/Providers</th>
<th>Description of Offering</th>
<th>Advantages/Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analytic Reporting Tools</strong> (e.g., Business Objects, Hyperion, Cognos, Information Builders, MicroStrategy, SAS, SPSS)</td>
<td>These tools support standard and ad hoc, query, reporting functionality for ‘power-users’. They may include predictive modeling and/or data mining functionality.</td>
<td>As a standalone tool, they do not offer a complete analytics solution. Their effectiveness for casual users is largely dependent on the tool being integrated with data models optimized for analytics. This optimization usually requires significant effort and/or cost before major benefits can be realized.</td>
</tr>
<tr>
<td><strong>Analytic Tools – Desktop applications</strong> (e.g., Rapid Insights, Tableau)</td>
<td>These tools support standard and ad hoc reporting functionality and are user-friendly (drag/drop).</td>
<td>These analytics tools are a low-cost alternative, but are not web-based.</td>
</tr>
<tr>
<td><strong>Template-Based Tool Kits</strong> (e.g., SunGard, Campus Management, Datatel, Oracle/PeopleSoft, Jenzabar, SAP)</td>
<td>Many of the Higher Education ERP vendors provide template-based toolkits that extend their ERP stack to include analytic applications.</td>
<td>A single vendor solution with whom an institution has a long, established relationship is attractive to many institutions. While the template-based reports provide out-of-box usability, for most institutions the reports do not fully meet their unique reporting/analytics needs. These solutions require consulting services and the institution’s IT resources to configure/implement, taking months or even years to achieve a fully optimized analytics environment. Additional modifications and customizations are required to integrate external data sources outside of ERP modules.</td>
</tr>
<tr>
<td><strong>Custom-Developed BI/Analytics Solutions</strong> (Mega-vendors such as Accenture, Sapient, SilverTrain, Unisys; start-up companies leveraging the Microsoft BI/analytics platform, SharePoint server which now includes the reporting and analytics services from PerformancePoint such as Solver Solutions and Perfiniti)</td>
<td>Many institutions opt to Build vs. Buy a solution in order to develop unique analytic applications from simple data marts/data warehouses to comprehensive analytics applications. Custom-developed applications begin with the acquisition of a BI/analytics platform or tightly integrated suite of tools (e.g., Microsoft SharePoint; Oracle OBIEE)</td>
<td>You get what you want if you can afford the time, have the IT know-how and can afford the related expense to ‘build it your way’. In the past, a Build approach was a risky, costly undertaking. Today this option is viable as several of the BI platforms (i.e. Microsoft SharePoint) are low cost, user-friendly and can be undertaken by the institution’s IT staff and/or with a professional services company.</td>
</tr>
<tr>
<td><strong>Packaged Analytic Applications</strong></td>
<td>This is a ‘buy’ solution that offers pre-built, pre-packaged proven analytics applications that have been designed to be quickly implemented, easily adapted and extended to meet an institution’s business logic, rules, definitions, and BI/analytics requirements.</td>
<td>These analytics applications have user-friendly interfaces, support casual user exploration, ad hoc reporting needs and depending on the particular vendor’s product can easily map to external data sources, feral systems (departmental/desktop Excel Spreadsheets or Access Databases).</td>
</tr>
</tbody>
</table>
Functional Requirements for CCC’s BI Solution

CCC has crafted an extensive data mapping document describing the data sources and structures that it envisions would be housed in an Oracle database with star schemas defining data/dimensions for each component of the desired data mart. This will be useful in the RFP process. However, CCC staff will need to create a functional requirements document that describes the desired reporting, querying, drill down, predictive modeling and analytics capabilities needed as well as future migration requirements – all of which must be included in CCC’s RFP to interested BI/Analytics vendors so that they may respond in kind in their proposals.

There are several standards organizations (including the IEEE) that have identified nine topics that must be addressed when designing and writing a Software Functional Requirements which can be found at: [http://www.techwr-l.com/techwhirl/magazine/writing/softwarerequirementsspecs.html], some of which CCC has already addressed.

1. Interfaces
2. Functional Capabilities
3. Performance Levels
4. Data Structures/Elements
5. Safety
6. Reliability
7. Security/Privacy
8. Quality
9. Constraints and Limitations

Equally helpful to successfully communicating functional requirements within an RFP is the inclusion of use cases – narrative descriptions of who the end-users are, what skills levels must be accommodated (power user, functional area staff, casual viewer), and what types of BI/analytics activities each user type will most likely employ (standard, ad hoc, predictive modeling, longitudinal analyses, executive dashboard/viewing).

Figures Summarizing the BI Challenge

The following figures portray the BI Needs of the Connecticut Community Colleges, both for the Current RFP and in the Longer-Term Future:

Figure 8: Desired BI Solution for Current RFP

Figure 9: Migration Paths – Cultural/Problem-Solving and Technical/Analytic Environments

Figure 10: What Are the Types of Reporting/Analytics?

Figure 11: What BI Functionalities Will Be Deployed? Today, Current RFP, Future BI

Figure 12: Who Uses BI? Today, Current RFP, Future BI
Figure 8

**Desired BI Solution (RFP)**

**Connecticut Community Colleges**

**50+ Users**

- **Presentation Layer**
  - GUI and User-Friendly Interface
  - Authenticated/Authorized

**Data Mart**
- Data Staging
- Data Transformation (Dimensional Tables)
- Dimensional Model

**BI Tool/Application**
- Interactive Reporting
- Ad Hoc Query
- Dashboards
- Dimensional Models/Star Schema
- OLAP
- Supports Predictive Modeling

**Extract/Data Sources**
- Frozen Extracts (14 Tables)
- Grad Extract
- NSC Transfer Data
- Student Surveys
  - CSSE
  - Non Student Survey
  - Porkas Web Survey
  - SENSE
  - Grad Survey
- Fin Aid Extract
- Admissions
- Registration Headcount
- Faculty Load Assignment
- Dept. of Labor Workforce Data

**Operational Data Sources**
- SIS
- Fin Aid
- FIN
- HR
- Advancement
- LMS Blackboard
- Surveys
- External Data Sources

**Strategic Initiatives 2009**
Migration Paths

Connecticut Community Colleges

Cultural and Problem-Solving Environment

- Culture of Evidence
- Standard Push Button reports for General Users (90 users)
- Statistical and Predictive Modeling and Analysis
- Power Users in BI Direct More Energy into Analytics and away from Reporting/Querying
- Achieving the Dream Initial Focus
- Limited Cross-sector Analysis

- Culture of Performance Management
- Push Button Reports for Busy Users, Easy Query for the Masses (1000+)
- Sophisticated Statistical and Predictive Modeling, Forecasting
- Power Users in BI focus on Higher Order Analytics; Identifying Data Sources, constructing Relevant Reports
- Focus on enhancing Student Access, Affordability and Success, all campuses
- Extensive Cross-campus, Cross-sector Analysis, PK-20 Partnerships

Technology and Analytical Environment

- ERP
- Extract-based Analytics, Complex and Time Consuming
- Basic ERP Data and Limited Additional Sources
- No OLAP
- SPSS and SQR Tools for Power Users
- No Dashboards

- ERP + Phase I BI/Analytics
- Data Mart-based Analytics, For Less Complex and Time Consuming
- Basic ERP Data and Limited Additional Sources
- First Gen OLAP
- Substantial Longitudinal Analysis
- First Gen KPIs/Dashboards

- ERP + Fully Functional Enterprise BI/Analytics
- Data Warehouse-based Analytics, Even Faster and Easier
- All ERP Buckets, LMS, Assessment, Advancement, External Data Sources
- More Advanced OLAP
- More Advanced Longitudinal Analysis
- More Advanced KPIs, Dashboards, Bloomberg-style ticker tape Presentation of Measures

<table>
<thead>
<tr>
<th>Current</th>
<th>RFP</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Strategic Initiatives 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3/27/2009</td>
</tr>
</tbody>
</table>
Note: Typically, BI/Analytics solutions providers have embedded enterprise level security and user authentication functionality that will leverage the organization’s policies and role-based permissions hierarchies to ensure that only selected, authorized end-users have access to sensitive data (i.e. student contact information, grades, grievances). Most BI solutions provide the institution the capacity to define and limit end-user access to data, including individuals’ ability to drill down to more granular levels of detail. Some solutions can track/report on the back end the details of end-users’ activities in the BI/analytics environment such as who accessed the data, who created reports, and with whom were data/reports shared.

Additionally the BI/analytics vendors serving the Higher Education market are aware of the new compliance/privacy regulations that have taken effect and are ensuring that their security/compliance functionalities address these new requirements.
## Initial Focus: Student Success Over Time

### What BI Functionalities Will Be Deployed?

<table>
<thead>
<tr>
<th>Types of Analytics and Reporting</th>
<th>CT CC Today</th>
<th>BI Tomorrow (First-Stage RFP for BI Tools)</th>
<th>BI Future (Advanced BI Tools &amp; Applications)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>• Optimization is in the conceptual phase today</td>
<td>• IR will access internal and external data sources and move toward optimization</td>
<td>• IR will expand the range of internal and external data sources, and move toward optimization across the System</td>
</tr>
<tr>
<td>Predictive Modeling</td>
<td>• No bandwidth/limited resources to do this kind of analysis</td>
<td>• Analysis of target population of students – Achieving the Dream</td>
<td>• Cohort coding; sub-group interventions</td>
</tr>
<tr>
<td>Forecasting and Extrapolation</td>
<td>• Excel-based tools; mostly focused on finance</td>
<td>• Analyze IR-related course enrollments, staffing; plan accordingly</td>
<td>• Forecasting spreads in all data areas</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>• Snapshots/extracts available for statistical analysis</td>
<td>• Able to conduct IR longitudinal analyses with ease using historical and internal/external data</td>
<td>• Statistical analysis will become progressively easier</td>
</tr>
<tr>
<td>Alerts</td>
<td>• Current alerts are limited to Achieving the Dream: At Risk students; performed manually</td>
<td>• Automatic/Dashboarding</td>
<td>• Enterprise-wide alerts potential</td>
</tr>
<tr>
<td>Query/Drill Down</td>
<td>• Very limited query and drill down</td>
<td>• Highly secure environment for simple queries through a CT portal</td>
<td>• Expand the universe of data</td>
</tr>
<tr>
<td>Ad Hoc Reports</td>
<td>• Ad hoc reports through telephane requests to IT</td>
<td>• Automatically generated and available in portal/dashboard</td>
<td>• Expand the universe of access to internal/external data sources, including all ERP modules; phase out extracts</td>
</tr>
<tr>
<td></td>
<td>• Feral systems abound – Access DBs and in Access and Excel formats</td>
<td>• Alerts/Interventions supported by workflow and are role-based/secure</td>
<td></td>
</tr>
<tr>
<td>Standard Reports</td>
<td>• 530 custom reports with many changeable parameters. Additionally, there are 270+ standard reports delivered in Banner that the college/System Office use</td>
<td>• BI Solution will build on existing standard reports; focused on extracts from Banner SIS, FIN, PeopleSoft HR, and external data sources</td>
<td>• Future BI functionality will continue to incorporate new data sources and phase out the use of extracts</td>
</tr>
<tr>
<td></td>
<td>• Limited KPIs; no dashboards</td>
<td>• Established KPIs and dashboards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Job submission function in Banner wherein users can pick report and/or extract to run with many parameters. Customizations allow users to view reports or extract data online</td>
<td>• BI solution will run against the IR-designed data mart and will serve as a single source of trust for all 12 Connecticut community colleges</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No ability to support longitudinal analyses</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Adapted from Davenport/Harris, 2006*
### Initial Focus: Student Success Over Time

#### Who Uses BI?

<table>
<thead>
<tr>
<th>Types of Analytics and Reporting</th>
<th>CT CC Today</th>
<th>BI Tomorrow (First-Stage RFP for BI Tools)</th>
<th>BI Future (Advanced BI Tools &amp; Applications)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>• Optimization is in the design phase</td>
<td>• IR internal/external data sources optimized</td>
<td>• All internal/external data sources fully optimized</td>
</tr>
</tbody>
</table>
| Predictive Modeling              | • No time to do this kind of analysis | • Power users conduct analysis of target population of students – Achieving the Dream/ at Risk | • Power users  
  • Spend increasing proportion of time on predictive analytics |
| Forecasting and Extrapolation    | • Power users | • Power users; selected C-level and functional administrators/staff (#550)  
  • IR administrators/staff focus on student/academic forecasting | • Power users  
  • Spend more and more time on advanced forecasting |
| Statistical Analysis             | • System and campus IR Officers perform statistical analysis, but it is very laborious | • Analysis is easier  
  • IR officers dedicate more time on statistical analysis | • Statistical analysis gets progressively easier and IR officers spend more time on it |
| Alerts                           | • Manual alerts, limited to a few supported by power users | • Power users create/support implementation of alerts/interventions  
  • Alerts by faculty, counselors, advisors made progressively easier and more effective |
| Query/Drill Down                 | • Limited to a few Power Users | • Essential to vision: easy to execute, highly secure environment for simple queries through a CT portal | • Expand the universe of end-users and access to more complete data |
| Ad Hoc Reports                   | • Ad hoc reports run by IT and IR officers  
  • Feral systems run by many end users | • End users will include IT, Payroll, CXOs, IR and key administrators and selected functional staff/administrators | • Expand the universe of end-users and access to internal/external data sources, including all ERP modules; phase out extracts |
| Standard Reports                 | • 270 standard reports and 530 custom reports with many changeable parameters that are run by end users at the 12 colleges and System Office  
  • 800 reports can be run by authorized groups/end-users; role-based access/security  
  • Personalized dashboard capability for authorized end-users  
  • Longitudinal reporting can be accomplished routinely by System and campus IR offices | | • Move out to a broader audience of end-users (i.e. 1,000s) |

Adapted from Davenport/Harris, 2006
Appendix A

Glossary of Terms: Data, Information, Reporting, and Analytics

As part of the process of raising the “analytics IQ” of users at Connecticut Community College System, we suggest the following collection of terms and definitions.

**Advanced Data Visualization.** The formation of mental visual images, the act or process of interpreting in visual terms or of putting into visual form. A tool or method for interpreting image data fed into a computer and for generating images from complex multi-dimensional data sets. Advanced data visualization functions allow the viewer/end-user to build mental models (graphical representations), the visual attributes of which represent data attributes in a definable manner.

**Analysis.** A combination of activities that leverages the institution’s data and information resources to build on and look beyond query and reporting to answer more complex and future-oriented issues. These include statistical analysis (Why is this happening?), forecasting (What if these trends continue?), predictive modeling (What will happen next? Who will be most affected?), and optimization (What is the best that can happen? How can the pieces best fit together?).

**Analytic Application.** A software application designed to fulfill all the analytics requirements of a business process (e.g., student analysis, financial, human resources). A pre-packaged analytic application is a baseline application that has been formally reviewed and agreed upon by customers/users in other settings and serves as the foundation for further developments in the baseline functionality. An analytic application may combine ETL, OLAP, DW, and BI tools, and is typically presented in a user-friendly wrapper – one that allows end-users to easily access standard reports, run ad hoc reports and perform simple queries based on templates and packaged data sets, rather than having to learn the arcane language required of complex query tools familiar only to power users.

**Analytics for the Masses.** Analytics that are “user friendly” or even “user obvious” can be utilized by the average end user without having to understand complex rules and procedures or to engage in lengthy training sessions. These tools enable institutions to deploy enterprise-wide analytics – “analytics for the masses.”

**Assessment Data.** Community colleges generate extensive assessment from student/course evaluations, enterprise surveys like the Community College Survey of Student Engagement (CCSSE), and pull data from external sources such as the U.S Department of Labor.

**Application Development Framework.** A solution that allows users to create their own applets for their internal business processes.

**BI (Business Intelligence) Tools.** A broad category of technologies that allows for gathering, storing, accessing, and analyzing data to help campus users make better decisions. In its broadest sense, BI encompasses all of the tools and systems that play a key role in the data-shaped decision making and strategic planning processes of institutions. In its narrower sense, the term “BI Tool” usually refers to a tool such as ProClarity or Hyperion that allows reporting and analysis to be carried out via a graphical user interface by end users who do not have extensive knowledge of the technical details of the underlying data structures. Even so, many BI tools require users to master protocols and operating details that are sufficiently complex to discourage non-power users (See definition of “power users”).

CCC IR directors have relied on IT to develop standard reports (300-500 to date) and use SPSS as a primary statistical tool.

**BI 2.0 Stack.** Traditional BI solutions are complex, difficult to configure/implement and designed for power users. A full BI Stack BI 2.0 architecture allows functional staff and casual viewers access to user-friendly tools that enable data visualization so that users can create their own models to see the relationships among data, interactivity that lets users manipulate the data and an intuitive manner of working that suits the way business users think, for example, in asking new questions as they arise. BI 2.0 tools are more intuitive for business users than traditional business intelligence tools, specialty software and spreadsheets. They tend to be more adaptable and run faster and are organized around the user’s experience, not the structure of the data.
**Assessing the BI/Analytics Landscape**

**APPENDIX A – Business Intelligence and Analytical Solution RFP**

This is an important consideration for CCC because of the longer term goal of implementing BI/analytics initially to IR directors to include administrators, staff and faculty across the System.

**Dashboards.** A reporting tool that consolidates, aggregates, and arranges measurements, metrics (measurements compared to a goal), and sometimes scorecards on a single screen so information can be monitored at a glance through graphical representations of the data.

**Data.** The basic elements contained in institutional data bases. Data assume meaning when they are strung together, combined with other data elements and expressed as information. Information becomes knowledge when it is understood in particular contexts.

**Data Dictionary.** This term encompasses a fairly wide spectrum of meanings, all of them related to understanding the structure and meaning of the data elements within a database. On the technical end of the spectrum, it refers to an actual database (or subset of a database) about data and database structures; a catalog of all data elements – their names, structures, and information about their usage; a central location for metadata. On the user-oriented end of the spectrum, it refers to documentation (often a text or hypertext document rather than a database per se) that describes the meaning of the data elements and their relationships in a less formal but still precise manner.

**Data Extension Architecture.** The ability for end-users, typically departmental or divisional users, to combine their own “What if” data with the centralized ERP data. The solution prevents modification of the centralized data and allows the users to merge their department specific detail in a way that eliminates the need for shadow systems.

**Data Governance.** This is the process for organizing data definitions, setting and monitoring responsibilities for data stewardship, and adjudicating differences of perspective on data issues. Also setting and memorializing standards, processes, and protocols for data governance and stewardship.

**Data Mapping.** The process of identifying the relationships between data elements in a source environment (e.g., a transactional ERP system such as Banner) and related data elements in the target environment (e.g., a data mart or data warehouse). These relationships can be simple, one to one mappings, or considerably more complex.

**Data Mart.** An analytical data store designed to focus on a specific business function for a specific community within the institutions (i.e. the IR Data Mart focusing on Achieving the Dream/At Risk Students). The data mart can be either a subset of an organizational data store (ODS) oriented to a specific purpose or major data subject, be derived from subsets of data in a data warehouse, or be linked to other data marts to create a data warehouse.

**Data Mining.** A process of analyzing enterprise data (often stored in data warehouses) to uncover hidden trends and patterns, and establish relationships. Data mining is normally performed by power users using specialized software tools.

**DW (Data Warehouse).** A data warehouse is a database geared toward the business intelligence requirements of an institution. The data warehouse integrates data from the various operational systems and is typically loaded from those systems at regular intervals. Data warehouses contain historical information that enables analysis of performance over time and can capture data at census dates, end-of-term, and other important milestones. Enterprise data warehouses typically glean data from across the institution and should be designed to be easily extensible for incorporating additional data elements.

- **A multi-dimensional data warehouse** is a powerful database that enables users to analyze large amounts of data by capturing/presenting data in a specific type of multi-dimensional structure. This structure is much different from that found in transactional databases and is optimized for reporting and analysis.

- **Operational data stores** are databases that typically draw data from relatively limited operational data sources, with little, if any, major structural transformation of the data, and are not nearly as extensible and powerful as data warehouses.

**Dynamic Modeling.** A method used to define and analyze data requirements needed to support the business functions of an enterprise. The data requirements are recorded as a conceptual data model with associated data definitions. Data modeling defines the relationship between data elements and data structures.
Dynamic Viewing and Drill Down. Using an analytic application, users can dynamically change data elements in a report, query, or analysis (like rotating the elements of a Rubik’s Cube—or, more prosaically, like using a Pivot Table in Microsoft Excel), and then view the changed result dynamically. Users can then “drill down” to identify individuals (students, faculty, staff) who are in particular cells captured by the analysis.

End-to-End Web Solution. A solution wherein all user components are delivered via a web browser—where every form, including end user, architectural, and administrative are all thin-client, browser-based solutions. A number of applications still exist today that have Web components, or light versions of their tool on the Web, but still require users to install desktop/client application(s) for administration. Some thin client solutions can be or have been “web-ized”, although this is not considered a true ‘end-to-end’ web solution, and it can be an expensive workaround.

ERP (Enterprise Resource Planning) Systems. The basic transactional systems of the institution (Student—includes Financial Aid in PeopleSoft, Finance, Human Resources, Financial, Advancement). CCC has deployed a single instance of Banner SIS, FIN, FINAID and integrates with the State’s Oracle/PeopleSoft HR/Payroll application.

Extensibility. The capacity to extend a data mart or data warehouse to include additional data elements that draw from additional operational data sources, shadow systems, external data, and/or assessment data.

External Data. Systems like the Connecticut Community Colleges need to draw data from a variety of external sources such as NSC, Pell and Department of Labor in order to better analyze and spot trends as they relate to Achieving the Dream/At Risk students. Over time, CCC will want to compare and benchmark themselves with peer and/or aspirational institutions, state and national initiatives, and disciplinary/programmatic comparisons.

ETL (Extract, Transform and Load). Three database functions that are combined into one tool to pull data out of one database and place it in another database:

- **Extract**: the process of reading data from a database.
- **Transform**: the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data and in some cases can be rather complex.
- **Load**: the process of writing the data into the target database.

OLAP (Online Analytical Processing). A type of reporting and analytics that uses data stored in a special format that allows users to perform fast, multi-dimensional analysis across different points of view. The data are often sourced from a data warehouse and pre-aggregated to improve querying response time.

Information. Elements of data that are organized together to yield greater meaning.

Interactive Reporting Model. A flexible reporting solution that can be used for analysis as well. Interactive reporting allows the end user to perform functions such as drill down, pivot, and create personal computed columns. It supports creating KPI dashboards and alerts as well as ad hoc queries and advanced data visualization.

Knowledge. Information presented and understood in particular contexts. Knowledge is best understood through the give-and-take of conversation, exploring the impact of context on meaning.

LMS (Learning Management System). A learning or course management system is a set of software applications that manages the creation, storage, use, and reuse of learning content. CCC has a single instance of Blackboard Vista hosted by Blackboard as its courseware application for its credit and non-credit offerings.

Near Real Time. Pertains to the delay introduced by automated data processing or network transmission, between the occurrence of an event and the use of the processed data; e.g., for display or feedback and control purposes. For example, information presented in near real time represents the data as it existed at the current time less the processing time (which may only occur once a day or at a certain time each week). This data may be referred to as Current Data.
Optimized Data, Information, and Analytics Resources. Optimization means that the data are fully defined, mapped, understood through an analytic application that enables a wide range of users to access, analyze, and act on integrated information in the context of their business processes and tasks that they manage within a given functional area, department, program or college.

Positive Restlessness. A desire to continue to excel; a motivation to make the best better. It’s never resting.

Power Users. Most BI tools require an in-depth understanding of complex data structures, rules, procedures, and protocols, making them usable only by a select group of “power users” in offices of institutional research and information technology, and/or some departments/functional areas.

Predictive Modeling. Encompasses a variety of techniques from statistics and data mining that analyze current and historical data to make predictions about future events.

Query and Reporting. Accessing the institution’s data and information resources to provide insight through standard reports (What is happening?), ad hoc reports (Change views on “What is happening? Insight on “How many, how often, where?”), query/drill down (Where exactly is the problem? Who is affected?), and alerts (What actions are needed? Who is at risk?). Query and reporting can be combined with predictive modeling to launch real-time alerts and interventions based on analytics.

Query and Reporting Tools. Software tools designed to support query activity and report writing. CCC’s IT and power users make use of Oracle SQL query and reporting tools to create/run reports against the Banner Oracle database.

Real Time. In an application, “real time” means that the information is received and immediately responded to without any time delay. May also be referred to as synchronous. The distinction between “near real time” and “real time” varies and the delay is dependent on when the data are captured and how often. Currently, CCC analytics environment relies on frozen data created from Banner and other external data extracts. Over time, CCC would like to retire extracts in favor of access to near/real time current data.

Shadow Systems. Departmental systems that are developed in order to work around or compensate for limitations in ERP systems and other institutional data sources. Over time, an optimized data, information, and analytics capability will eliminate the need for many shadow systems and/or enable data to be extracted from shadow systems for incorporation in the data warehouse.