Business Intelligence and Predictive Modeling (BIPM) Mission and Data Standards

1. **Mission:**
   To provide University personnel data-derived information for evidence-based decision-making related to internal policies, auditing, review and assessment, and to empower analysts by having a comprehensive data center from which to access integrated student/course data to meet their needs for information and to increase information efficiencies.

2. **Data Scope:**
   
   **Unified Student and Course Data System**
   
   A centralized data system that spans 1980 to the present providing data integration along horizontal and vertical dimensions related to student migration through the University. BIPM data encompasses horizontal migration through incoming (recruitment and admissions), in-process (registration, enrollment and course activity) and outgoing (degrees) stages; vertical integration includes supplemental student and non-student data systems (financial aid, housing, payroll, and accounts receivables) that cross divisional units that are joined to the horizontal migration tables. Data are stored at the detailed student and/or course level and allows for data aggregation and disaggregation as needs warrant. BIPM tables are refreshed daily via IBM SPSS Server from native Banner tables to include new and modified student and non-student data. In addition to the Core, Pre-stage and Supplemental tables that comprise the BIPM integrated centralized student/course data system, BIPM creates custom tables for support of reporting service views, daily Banner production table updates, debugging of Banner modules, maintenance of coding schemas (used in both BIPM and native Banner), and the storage of non-native Banner data.

   Dynamic models producing several scoring values are refreshed daily to adjust for changes in student-level/course-level data and changes in aggregate probability samples. Student and structural (e.g., high schools, community colleges, geographic regions) scores are uploaded to Banner daily via IBM SPSS for use among BIPM and native Banner users. Dynamic scoring allows for the identification of students at academic risk, courses which place at-risk students in academic jeopardy, high matriculation yield rates of high schools and colleges for targeted mass recruiting, student probabilities of matriculation for targeted recruitment and admission decisions, and the use of BIPM data to increase student retention. Modeling also allows for the identification of statistically significant predictors in student success which can be used to take institutional action. In addition to score modeling, forecasting models are automatically scheduled daily to produce estimates on future recruits, applications and enrollment outcomes for planning and targeted unit accountability. Tuition revenue models are produced at several different structural levels ranging from the course-level, Departments, Colleges and to the University-level in identifying revenues and cost for budgeting and resource allocation. BIPM data also are employed to provide daily reports on pre-20thday registration and tri-semester year round application activities. Data integrity reports are produced daily to identify data entry errors in native banner tables that offices can correct within a 24 hour period of original data entry.
3. **Data Tables:**
Authenticated users have access to a set of denormalized tables from which to perform data exports, reporting, analysis and forecasting. Denormalized tables are constructed in consultation with offices that have jurisdiction over affected base tables in order to have knowledge of office business practices related to data entry and to create denormalized tables that meet the data exporting, reporting, analysis and forecasting needs of offices and University personnel. Denormalized tables may be used singularly or joined, and can be joined with native Banner tables to access data necessary to inform decision-making processes. Reporting service views, custom tables and applications can be created from BIPM tables by offices to meet their reporting and business practice needs.

4. **Data Standards:**
BIPM is an open source (in terms of data content) data system based on authenticated user access to data with complete transparency in data derivation, execution and content.

- **Open data source system (data, code, documentation)**
  - All BIPM data are stored on the University information system storage facilities for access by authenticated users
  - BIPM code files for denormalized and custom tables, reporting, and analysis are made available to all authenticated users
    - allows users to see code for table builds to fully understand data derivation
    - in the event the data architect can no longer execute the code, backup architects can step in & continue data production since code access & documentation are available
    - access to code allows users to make suggestions to increase code efficiencies and to identify incorrect table dependencies
- **BIPM data fully documented (source & build code, data dictionaries, data maps)**
  - Along with source code, data dictionaries detailing table columns (names, content, format, source) and sample data views are available for each BIPM table
  - Data maps showing table dependencies with Banner & BIPM tables are also available
  - BIPM User Groups and Power Users provide open source input and voice their needs for additional data elements within the BIPM data system
- **No offline data systems or manual data entry permitted**
  - Source data for BIPM tables for exporting, reporting, analysis, and forecasting are derived from the Banner data systems
    - In cases where data are not native to Banner, custom BIPM tables are created for data storage and access with source code available to all BIPM users
  - When Banner data are found to be in error, BIPM will work with the office who has control over the affected data to make corrections
    - In cases where corrections cannot be made in native Banner tables, custom BIPM tables are created to store the corrected versions along with source code
  - All Banner data customizations must retain pre-customization values when constructing BIPM tables
  - All ETL (extract-transform-load) builds of BIPM data tables (including custom tables) are code based, no manual data entry allowed
    - In cases of custom data entry (such as an adjustment of a student’s record), insertions are done in code rather than offline in static tables
  - Downloading of BIPM tables or conversions is discouraged; instead, code-based extracts from live BIPM tables is recommended
- **Data formats**
  - IBM SPSS
    - All BIPM tables and code (including embedded sql) are in IBM SPSS formats
    - University 500 seat client license means no additional software cost is incurred for adding BIPM users
    - Along with the IBM SPSS client, enterprise level software including IBM SPSS Server and IBM SPSS Collaboration & Deployment Services is available for BIPM data
    - IBM SPSS provides for automation of scheduled BIPM ETL data builds and updates including uploads to Banner and the automation of scheduled BIPM reports
    - Allows code builds in both native SQL and IBM SPSS syntax including code integration across multiple system platforms
    - Single product use for data importing, exporting, data management, statistical analysis, modeling and scoring, forecasting, graphics, custom tables and reporting (including integration in MS Reporting Services and Visual Studio)
    - IBM SPSS is fully integrated with IBM Cognos
For a selected subset of BIPM tables used for feeding reporting service views, Banner popsels, validation tables, and intra-platform transfers, dbo (database object) tables are available in MS sql Server & Oracle for Kokopelli, ODSP, PROD & RPTPRD storage spaces

- Conversion to other data management and statistical software packages are available from the IBM SPSS software systems when required

**BIPM Naming Standards and table types:**

- **CORE**—the main table builds that end users rely on for one-stop-shopping (e.g., recruits, admissions, student, course, degrees).
- **SUPPLEMENTALS**—tables that hold supplemental data for the CORE tables (e.g., SARD, payroll, financial aid, housing); these are normally repeating tables.
- **VALIDATION**—validation codes and descriptors for all table builds (e.g., mcg majors, ccg course, term codes); these are normally non-repeating tables.
- **PRESTAGE**—serve as pre-staging builds for the COREs (or what Oracle calls dimension staging tables), they allow us to separate legacy & banner data so that legacy does not require refreshes when daily CORE builds are performed. For example, the CORE student 20thday table has a student 20thday legacy table for data from 198030 thru 200630 (STU20_legacy) and a student 20thday banner table (STU20_inb), these two tables are joined along with supplementals to create the CORE table (STU20), having the legacy & banner pre-stage tables in separate code & tables allows for faster processing since the legacy data does not need to be refreshed.
- **REPORT**—tables used by Reporting Services, allows RS to render results faster since they don’t need to read the large CORE tables (essentially acting as specialized data cubes).
- **OTHER** (letters T and U )—tables that perform specially functions such as uploading scores into Banner PROD

When writing tables in the BIPM schemas use the following prefixes to separate tables by type:

1st digit=IS ownership/custodians (S=student/course; F=Finance; H=Human Resources)
S = Student/course (home is BIPMS)
F = Financial operations (home is BIPMF)
G = General/common shared tables (home is BIPMG)

2nd digit=table type (letters A & B are reserved, do not use)
C = Core (main denormalized tables)
P = Pre-stage (pre-builds for Core ETLs)
S = Supplemental (supplemental data shared across Core tables)
V = Validation (validation tables)
R = Report (reduced/purposed tables for reporting & reporting services)
T = Other (specialty functions)
U = Other (upload data to Banner)

Examples:
- SC_STUCRScur.sav (core student & crn current term data)
- SP_STUEot_leg.sav (prestage build for core file)
- SS_Finaid.sav (supplemental for cores)
- SR_adm_app_errors.sav (report error checking table for applications core)
- SR_AtRisk_STU.sav (report table for reporting services)

5. **BIPM Data Custodians:**

BIPM is a collaborative effort among many offices. The Office of Academic Affairs through their designated data architect controls BIPM data content and management of code builds including BIPM documentation and training. University Computing and Telecommunications (UCATS) provides security for authenticated user access, data storage, and development of BIPM related reporting service views. Several university offices provide input for what data are included in the BIPM tables (e.g., Undergraduate Admissions, International Education, Graduate School, Registrar, Financial Operations and Business Technology, Institutional Research, academic colleges and departments, advisors).
The BIPM data architect is responsible for all supplemental, pre-stage and core ETL builds including all documentation (code, data dictionaries, and data maps) in the BIPM production storage space. BIPM power users access data from the mirrored BIPM non-production storage space. The primary data architect will have at any time two data architect backups within different units in the event the primary architect can no longer provide service; the code-based BIPM system and documentation is meant to provide any trained BIPM user the ability to execute BIPM code files.

Revised June 15, 2012