INITIAL PROJECT PROPOSAL
(Research Data Warehouse)

EXECUTIVE SUMMARY – SECTION 1

Provide high-level, summary information about the project and why it is needed. This is submitted by the requestor and approved locally before technical research is started. This is a tool for the local college, department, etc. to make a project request and then prioritize.

Bakersfield College (BC) will utilize funds from the Innovation Award to support the creation of a data warehouse, in partnership with Phytorion, a Data Warehousing and Analytics software company. A data warehouse is a centralized database that integrates information from the entire organization (e.g., Banner, Starfish, CalPASS, eLumen, etc.) and provides snapshots of key data, a streamlined data system, and a single version of the truth.

The current operational data store (ODS) poses challenges when extracting data and are inadequate to meet the information needs of the College. The Research Data Warehouse (RDW) will allow the Office of Institutional Effectiveness (OIE) and other college constituents to easily and quickly extract data; thus, supporting the College's data request needs in a timely and efficient manner. In an era defined by performance-based funding and guided pathways work, Bakersfield College is critically dependent on high quality, highly available data.

BUSINESS PROBLEM

Describe the business problem or opportunity.

Overall, the Office of Institutional Effectiveness and other college constituents have issues obtaining reliable, consistent data in a timely manner.

There are several issues with relying exclusively on the current data system, our Operational Data Store or ODS:
- Lack of definitional clarity and visibility in data functions that are hard-coded by IT and not accessible by IR personnel resulting in data that cannot replicated or validated,
- Inconsistencies in the data,
- Messy data – the presence of multiple, similar elements with little documentation can result in confusion and errors when reporting out MIS data to the state, and
- Availability of the ODS – due to the complexity and comprehensiveness of the ODS, processes often run long or get hung up and fail to update, resulting in outages for IR personnel.

ANALYSIS

Describe business processes that will be improved or are not operating properly because of the problem.

Creating a data warehouse will improve the speed and efficiency of accessing and joining data tables to create custom data sets.

The RDW will enhance data and reporting consistency by implementing a streamlined, transparent, and well-documented data schema.

The RDW is different from the ODS because it will store snapshots of a wide assortment of historical information resulting in stable and consistent reporting over time enhancing the perception of data quality and the ability of the District to represent one single version of the truth (vs. the ODS which is constantly being updated due to its operational nature).

What is the impact of not implementing the proposed solution to the problem?

Inefficient use of the research personnel's time will result in a backlog of critical data requests which will impede the College's core mission of supporting students' attainment of their educational goals. Inability to optimize graduation, transfer and other key momentum points will result in less funding as the performance-based funding formula comes on line.

List all quantitative support in favor of eliminating the problem.

The RDW has full college support. The College has received an Innovation Award from the state to help fund development of a model research data warehouse that has enhanced capabilities to stage and deploy high school performance data for assessment, placement, and research purposes.

What timeframe are you hoping to get the solution\problem resolved within?

The RDW will be implemented in three six-month phases. The plan is for the RDW to take 1.5 years to complete. In addition, annual maintenance will be needed to support the RDW. Phytorion is contracted to provide support the RDW one year after going live.

Is the solution to this problem an opportunity to collaborate with other colleges?

Yes; as part of the Innovation Award, the RDW will be a data warehouse model for other California Community Colleges (CCC). Therefore, collaboration with other CCC who plan to build a RDW may take place.

SOLUTION

Describe the proposed solution in detail.

In partnership with Phytorion, a RDW will be built to meet the data and reporting needs of BC.

Phytorion will provide the following consultants to build the RDW: project manager, architect, Extract, Transfer, and Load (ETL) developer, Cognos developer/tester, and project director.

OBJECTIVES

List the high-level objectives of the proposed solution.

Provide the scope of the proposed solution

Note: Please be as thorough as possible in what you want implemented (for example, if there are multiple modules in a software/hardware solution, which of those are you hoping to implement?) The scope that is defined in this section will help determine the resource needed for the project and shouldn't be changed once the project is in the queue.

During this process, fit/gap sessions will be between Phytorion and the BC to help translate the identified gaps into tables and fields. Phytorion will meet with BC's management, functional, and technical experts to determine out business requirements for the following:

Phase I. Create within the RDW, a schema that relates to querying all available Chancellor's Office Management Information System (MIS) data tables, as well as external systems (CCCApply; student test scores (e.g., Accuplacer); high school performance data from Cal-Pass, CCCApply, and CCGI; eLumen Curriculum and Assessment data; and National Student Clearinghouse). RDW development will include ETL systems for updating all tables regularly. In addition, Phase I will include the creation and provision of one Cognos reporting portal.

Phase II. Incorporate stakeholder requirements including additional Banner data, financial data, student accounts, and general student records and other requirements derived from the fit/gap sessions. Additional pertinent data beyond what is included in the MIS data, including more recent data and more detailed data (e.g., daily snapshot of critical enrollment data elements) to address requirements determined from the fit/gap sessions will be included in the RDW during the second phase of work. Further, one additional Cognos portal will be created.

Phase III. Cognos portals will be created for reporting purposes and further tuning of the RDW (i.e., RDW data flow, data tables, relationships, ETL, etc.).

Documentation

A document on design will be provided by Phytorion, including a list of tables and fields that contain the data elements used to populate the RDW and a list of data transformation rules that will be applied to the source data elements. In addition, a Data Dictionary, including the names, definitions, and attributes of the data will be provided by Phytorion.

Build ETL mappings using Microsoft SISS.

Phytorion will test each mapping, as well as the complete RDW set of mappings for a controlled subset of the data of a production-quality database.

Cognos portals

Phytorion will deliver up to seven Cognos portals.

Training

Phytorion will train BC in the operations of the RDW.

Transfer knowledge

Phytorion will provide overview meetings on technical (overview of the technology, architecture, ETL tool and operations of the RDW) and functional aspects of the RDW.

Maintenance and custom programming services

For a one-year period following go-live, Phytorion will offer maintenance and custom programming services (at no charge).

DELIVERABLES

List the project deliverables. A deliverable is a unique and verifiable product, result, or capability to perform a service that must be produced to complete a process, phase, or project. Requirements

- Fit/gap sessions with Phytorion consulting team and appropriate BC team.
- Phytorion will create ETL mappings that extract data from our systems, transform them to reporting structures, and populate the RDW. The jobs that submit the RDW ETL mappings.

- Phytorion will create the reporting tables in the RDW populated with source data.
- Phytorion will create documentation for
  - Requirements gatherings: a summary document with scope, assumptions, and priorities.
  - Design:
    - Visio diagrams of data model.
    - Sourcing documents that include a general table description, the source tables the data is coming from, the source fields for the data, the transformations applied to the source fields in order to create the target fields, and business descriptions for the target fields.
    - Go-live: a technical overview document.
- Training will be provided by Phytorion.
- Phytorion will provide knowledge transfer meetings.
- Cognos portals will be developed by Phytorion.

ESTIMATED COST

Provide high-level cost information or funding(s) for implementing the proposed solution. Items include Software, hardware, training, ongoing license\maintenance, purchase price.

Description	Estimated Cost
License/Contract	$75,000
Phase I – MIS	$100,000
Phase II – Additional Banner information, financial data, student accounts, and general student records	$120,000
Phase III – Cognos portals	$100,000
TOTAL ESTIMATED COST OF PROPOSED SOLUTION	**$400,000**

Phytorion will provide the services and deliverables for an amount up to $400,000. Payments will be based on the time it takes to complete the deliverables, as well as the materials provided.

AUTHORIZATION – SECTION 1

Date: 10/23/18 _____ Project sponsor approval

COLLEGE REVIEW AND APPROVAL – SECTION 2

Note: This section to be completed by the campus IT Director, or in the case of the district office, a district office IT Director.

Provide high-level, summary information about the project and why it is needed. This section is submitted to the college's technology committee and approved locally before submission to the district wide committee.

If it is determined it can be done locally no further submission into the district wide process is required and the college will proceed as needed.

The following are areas that need to be reviewed and verified prior to further submission into the process. Several of these sections will help indicate if District Office resources are needed.

☐ SSO (Single Sign-on) – Will staff need to have access to the system (internal\external to the district).

☒ Data integration – What other systems will this solutions' data need to access - both internal\external to the district.

☒ New application – Is this a new application in the district

☐ Security – Data\access security analysis

☐ Legal – Contracting language, FERPA, HIPPA, etc.

☐ Accessibility – ADA, 508 compliance

ESTIMATED TOTAL COST OF OWNERSHIP

This section will share how this product will be supported for the duration of the life cycle until it is discontinued. Key parts will include:

☐ Ongoing funding source – This is GUI, RP, grant, etc.

☐ Staff support – how will this be supported for ongoing maintenance of the solution

If this is a grant funded project the college will provide the resources to support this system once the grant funding has ended.

AUTHORIZATION – SECTION 2

Date: _10/15/18_ _Pamela K. Bayley_____ IT Committee Faculty Co-chair

Date: _10/23/18_ _____ College IT Director approval

COMMITTEE REVIEW – SECTION 3

ESTIMATED SCHEDULE

Provide high-level schedule key milestones.

Project Milestones and/or Phases	Estimated Completion Date
Start of Project	
End of Project	

This section still in development.