Operational Concept Description (OCD)

Soccer Data Web Crawler

Team No. 02

First Name	*Last Name*	*Role*
Trupti	*Sardesai*	*Project Manager*
Wenchen	*Tu*	*Prototyper*
Subessware	*Selvameena Karunamoorthy*	*System/Software Architect*
Pranshu	*Kumar*	*Requirements Engineer*
Zhitao	*Zhou*	*Feasibility Analyst*
Yan	*Zhang*	*Operational Concept Engineer*
Qing	*Hu*	*Life Cycle Planner*
Amir ali	*Tahmasebi*	*Shaper*

Version History

Date	Author	Version	Changes made	Rationale
10/07/14	TS	1.0	• Original template for CSCI 577a: Tailored from class OCD Template.	• Initial Document for Evaluation Phase
08/11/14	SK	2.0	• Added section 3.2, 3.3	• Section 3.2 was added to provide traceability for the outcome in the Benefits Chain
11/11/2014	TS	3.0	• Modified section 1.2, 3.1.2	• Final document for Development phase

Table of Contents

Table of Tables

Table of Figures

1. Introduction

1.1 Purpose of the OCD

The purpose of the OCD is to capture the shared vision of the success critical stakeholders of this project.
The success critical stakeholders for this project along with their roles are as follows:

- SporTech B.I. - owner
- Club managers, soccer club owners, club presidents, soccer coaches – End users.
- SporTech B.I. contractors - maintainers.
- Members of Team 02 - developers.

This document provides an initial reference for benefits expectation, benefits chain, current system and environment assessment, system objectives, constrains and priorities, new operational concept, organizational and operational implications.

1.2 Status of the OCD

The version number for this document is 3.0 and the project is currently in its final version of OCD. Previous versions have settled down the shared version, system objectives, constraints and priorities as well as the organizational and operational implications. In the version, we revise the shared vision and system boundary, refine some workflows and operational conceptions of the project and fix some mistakes.

2. Shared Vision

In order to understand or know what projects or related initiatives are required for program management, we create a Program Model that helps in designing and managing programs.

<table>
<tr><td colspan="4">Assumptions
• Increased need for access real-time data
• Soccer clubs losing money because of inefficient data entry process</td></tr>
<tr><th>Stakeholders</th><th>Initiatives</th><th>Value Propositions</th><th>Beneficiaries</th></tr>
<tr><td>• Developers
• Maintainers(SporTech B.I. contractor)
• SporTech B.I.</td><td>• Develop a web crawler to fetch data from required sources.
• Promote the apps that will be used by the users of the system.
• Train users on how to go about using the system and the apps.
• Maintain the system after its development to provide continuous Support.</td><td>• Enable user make well informed decision about soccer club operations and performance
• Increased time saving to gather data and automate
• Increase operational efficiency by Data gathering, Web crawling and lowering the errors
• Increase accessibility of real-time data and information</td><td>• Club Managers
• Club owners
• Club presidents
• Soccer coaches
• SporTech B.I.</td></tr>
<tr><th colspan="2">Cost</th><th colspan="2">Benefits</th></tr>
<tr><td colspan="2">• To maintain the system after development.
• Train SporTech B.I contractors.</td><td colspan="2">• Time-saving in gathering data and data ingestion.</td></tr>
</table>

Table 1: The Program Model

2.1 Benefits Chain

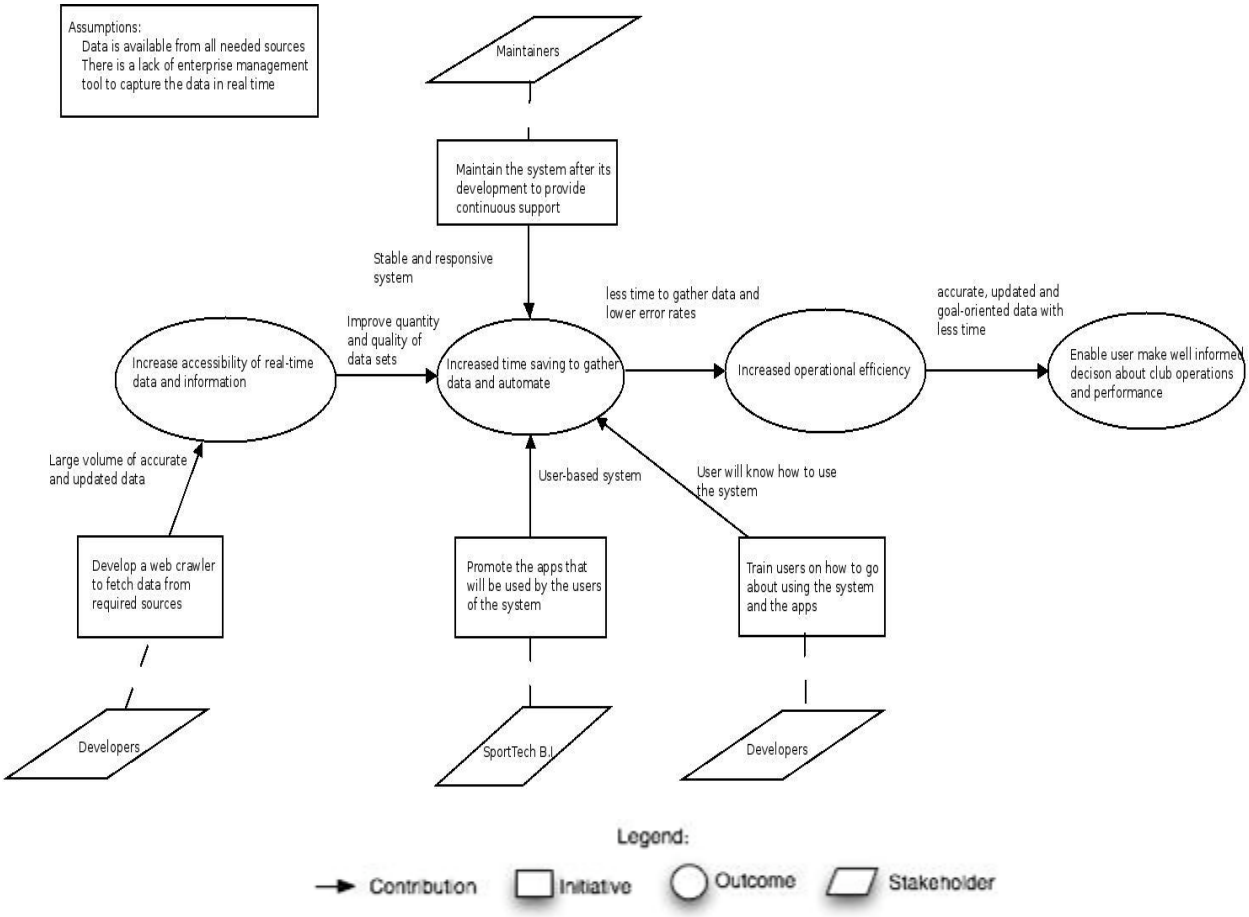

Figure 1: Benefits Chain Diagram for Web Crawler.

2.2 System Capability Description

- Our main task of the project is to design a web crawler for SporTech B.I. which will gather key soccer data from various specified websites and populate the SporTech database to feed the ecosystem of SporTech B.I. applications.
- The system also gathers data about players from social networking sites (Facebook and Twitter) and ingests it into SporTech B.I database. Data to be gathered includes number of retweets, likes, fans, the comments, name and number of group members.
- The target customer for our project will be the developers/contractors of SporTech B.I. and the ultimate target consumer will be the Soccer Clubs.

2.3 System Boundary and Environment

The system boundary and environment diagram contains a list of services and functions that the project team will be responsible for developing and delivering, as well as the system environment showing the stakeholders' organizations and other systems for which the project has no authority or responsibility, but with which the delivered system must interface in order to deliver the desired benefits. The figure below shows the system boundary and environment diagram.

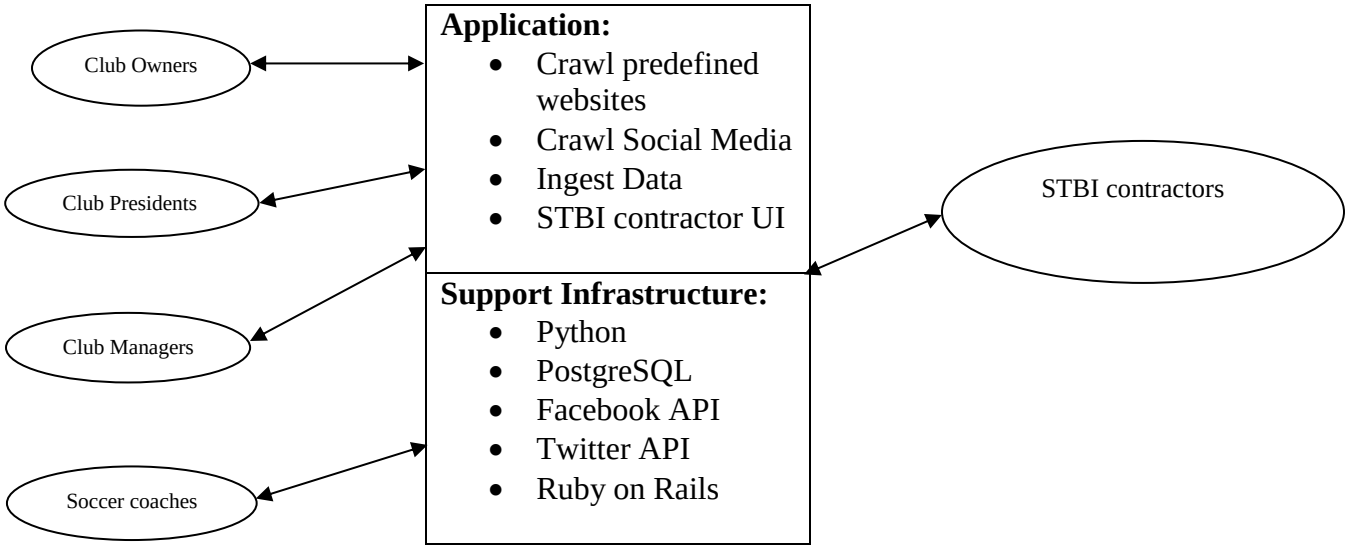

Figure 2: System Boundary and Environment Diagram

3. System Transformation

3.1 Information on Current System

3.1.1 Infrastructure

The client currently just owns the IP and a firm prototype of the application.
The current system uses a tool named 'import.io' to crawl a list of websites that contain soccer data. The major disadvantage of this system is that the tool doesn't automatically ingest data into a database. The need for manual data entry creates a time difference that is significant enough to lose real time accuracy of the data fed into the database.

3.1.2 Artifacts

Artifact	Description
Customer facing product card	Explains overall product/services that might be interest of the team.

3.1.3 Current Business Workflow

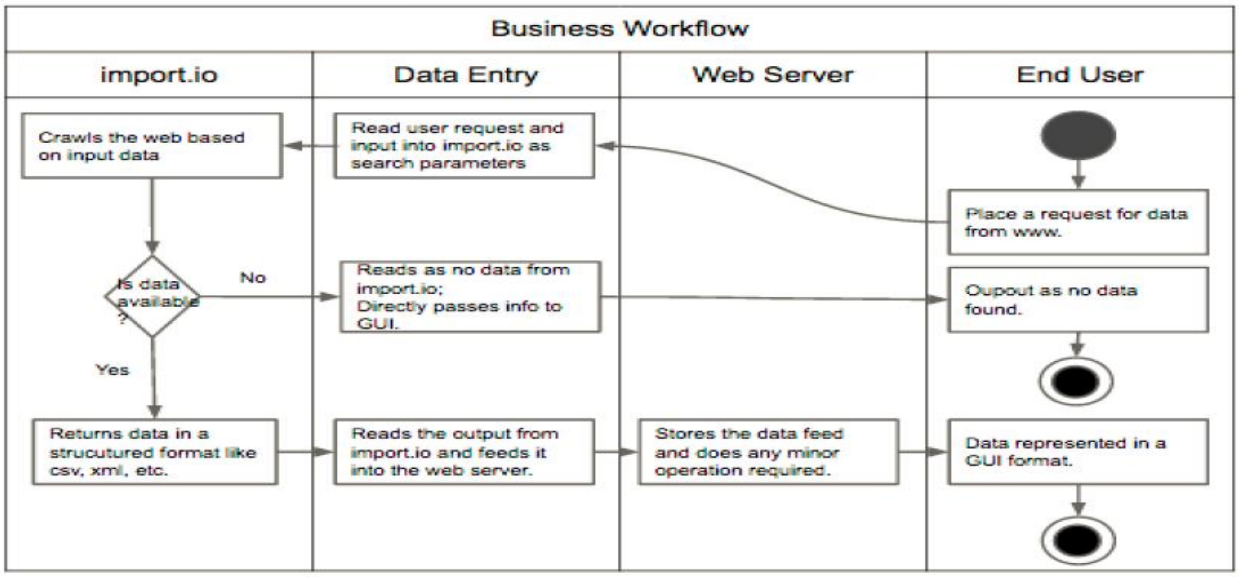

Figure 3: Business Workflow of current System

3.2 System Objectives, Constraints and Priorities

3.2.1 Capability Goals

Capability Goals	Priority Level
OC-1 **Crawl predefined websites**: The web crawler shall gather team information from the websites in the website list.	Must have
OC-2 **Crawl predefined websites**: The web crawler shall gather player information from the websites in the website list.	Must have
OC-3 **Collect Social Media**: The web crawler shall get comments, name and number of members, likes from specified Facebook pages.	Must have
OC-4 **Crawl Social Media**: The web crawler shall get number of followers, the comments and the number of retweets for a specified twitter account.	Must have
OC-5 **Ingest Data**: The crawler shall ingest crawled data into PostgreSQL database.	Must have
OC-6 **STBI Contractor UI**: As a SporTech B.I contractor, I can update/revise the player data as the season progresses.	Must have
OC-7 **STBI Contractor UI**: As a SporTech B.I contractor, I can add, delete, update the specific websites visited, fields to capture from the website and frequency of crawler refreshes for each specified website.	Must have
OC-8 **Crawl Social Media**: The web crawler shall gather Instagram pictures, number of likes and the comments from particular Instagram account.	Would Like
OC-9 **Crawl predefined websites**: The web crawler shall gather videos from the pages being crawled and ingest into STBI as is so that the coach and fans is able to watch the relevant videos.	Would like
OC-10 **Crawl Social Media**: The web crawler shall crawl YouTube to gather videos of specific players.	Would like

3.2.2 Level of Service Goals

Level of Service Goals	Priority Level	Referred WinWin Agreements
Flexibility The system can crawl and scrape any given URL into database.	Would like	WC_3414
Efficiency The system can crawl and scrape Facebook and Twitter data for a player in a time proportional to the amount of comments and post the player's account has. The system can crawl and scrape	Would like	NA

specific website in an hour averagely.		

Table 2: Level of Service Goals

3.2.3 Organizational Goals

OG-1: To enable the end users to make a well-informed knowledge about the players/team.

OG-2: To increase time-saving to increase operational efficiency.

OG-3: To increase accessibility of real-time data/information.

3.2.4 Constraints

CO-1: Operating System: The new system must be able to run on Windows, Linux and Mac platform.
CO-2: Zero Monetary Budget: The selected NDI/NCS should be free or no monetary cost.
CO-3: Python and PostgreSQL as a Development Language: Python will be used as a development language and PostgreSQL will be that database.

3.2.5 Relation to Current System

Capabilities	Current System	New System
Roles and Responsibilities	The Current System does not have any roles and responsibilities	The new system will have a maintenance team responsible for keeping the site up and financial officer to verify donations.
User Interactions	No user interactions	The new system will interact with contractors of SporTech B.I.
Infrastructure	Infrastructure does not support real time data analysis.	The new system will make available real time data about soccer players to the end users.
Stakeholder Essentials and Amenities	The new system is being used by club managers, club owners, club presidents and gym owners	The new system will be used by club managers, club owners, club presidents and gym owners
Future Capabilities	NA	The end users can make a well informed decision about soccer players/teams as they real time data about

Table 3: Relation to Current System

3.3 Proposed New Operational Concept

3.3.1 Element Relationship Diagram

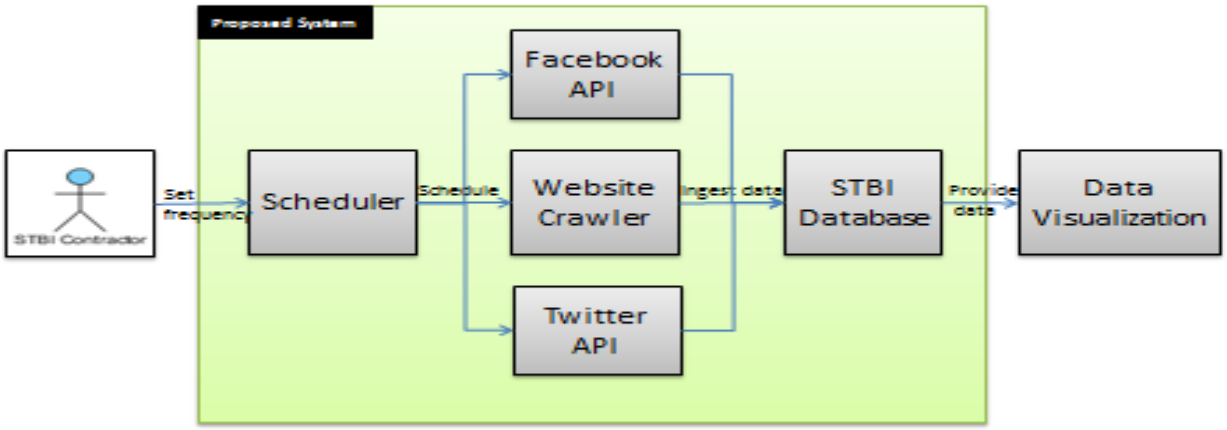

Figure 4: Element Relationship Diagram

3.3.2 Business Workflows

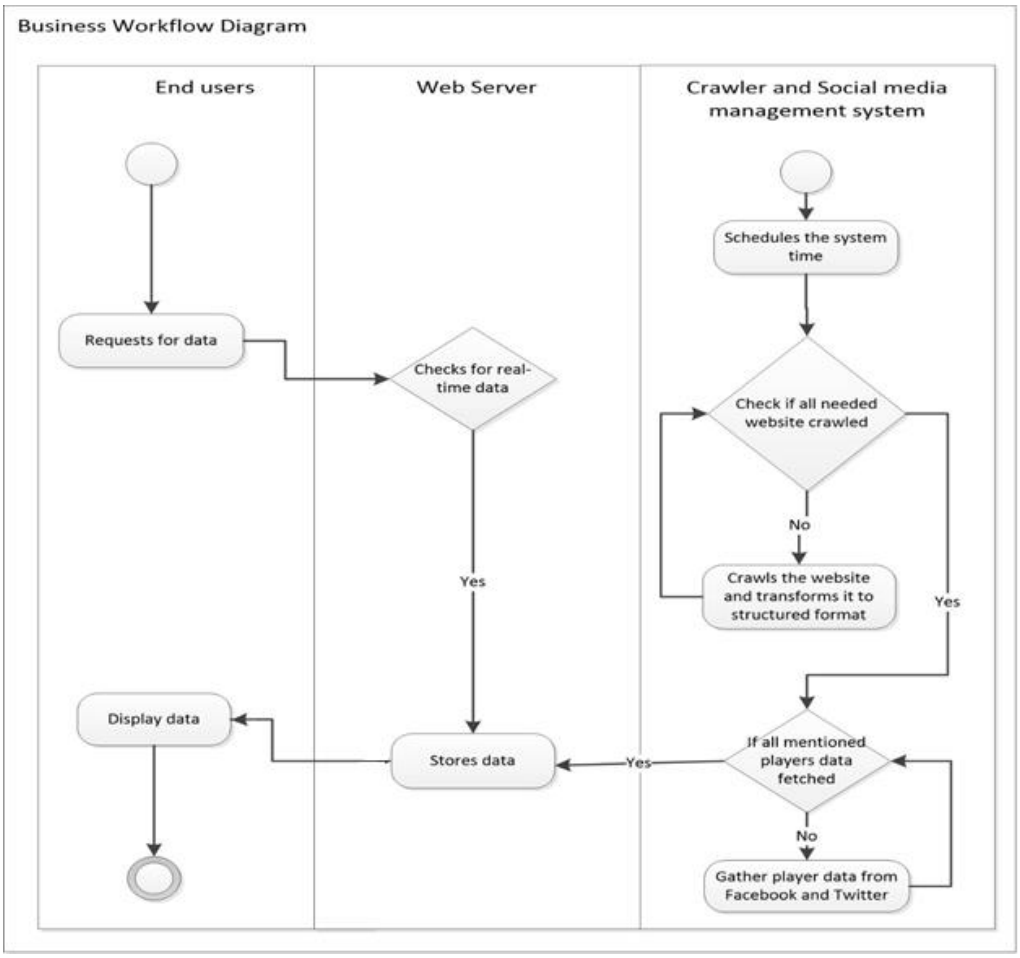

Figure5: Business Workflow Diagram of Web Crawler System

3.4 Organizational and Operational Implications

3.4.1 Organizational Transformations

- To hire maintainers to take care of the system.
- To hire UI developers to integrate the system being developed with data visualization component and engineers to maintain the system.
- The need to create tie-ups with soccer clubs.
- The need to hire marketing team to market this app amongst the soccer world.

3.4.2 Operational Transformations

- Members will have benefit of access to real time data so that they can make a well-informed decision about hiring new soccer players.
- This will become the first app to gather player details from sports websites as well as from social media.