PROTEXPRESS 1.0

Use Case Summary

National Cancer Institute

Center for Biomedical Informatics
and Information Technology

February 13, 2009

protExpress Development and Management Teams			
Development	*Quality Assurance*	*Documentation*	*Project and Product Management*
Krishna Kanchinadam[2]	Tom Boal[5]	Carolyn Kelley Klinger[4]	Carl Schaefer[1]
Scott Miller[2]		Krishna Kanchinadam[2]	Anand Basu[1]
Steve Matyas[2]		Mahidhar Narra[6]	Xiaopeng Bian[1]
			Bill Mason[2]
			Brent Gendleman[2]
Systems & Application Support			
Bob Wysong[3]	Andrea Johnson[3]	Ralph Rutherford[3]	Sriram Kalyansundaram[3]
Nimish Shah[3]			
[1.] National Cancer Institute Center for Biomedical Informatics and Information Technology (NCI CBIIT)		[2.] 5AM Solutions Inc.	[3.] Terrapin Systems
[4.] Lockheed Martin	[5.] NARTech	[6.] Stelligent	

Contacts and Support	
NCICB Application Support	http://ncicb.nci.nih.gov/NCICB/support Telephone: 301-451-4384 Toll free: 888-478-4423

Contents

Introduction

This document serves to identify the use cases for protExpress. It is intended to be a living document, modified throughout the iterative development life cycle.

Purpose

The purpose of this document is to collect, analyze, and define high-level needs and features of protExpress. It focuses on the capabilities needed by the stakeholders and the target users, and why these needs exist. The details of how protExpress fulfills these needs are detailed in the individual use case specifications.

Overview

An ever increasing amount of proteomics research data, especially mass spectrometry data, has been made available in the past few years. Several large scale repositories have been created to host proteomics experiment and protocol data. The National Cancer Institute has implemented a public proteomics repository, based on Computational Proteomics and Analysis System (CPAS), developed at Fred Hutchinson Cancer Research Center. An instance of the system is currently hosted at NCICB and serves valuable data from NCI-funded cancer proteomics initiatives to the scientific community. Currently, the repository provides public access to 10,160,229 peptide identifications from 687 ms-ms runs from Mouse Proteomics Technology initiative (MPTI).

One problem for transporting the data from labs to public repositories is the lack of standard format and tools for capturing experiment and protocol annotations. For example, CPAS uses eXperimental ARchive (XAR) format, while other tools use proprietary and custom formats. The current NCICB proteomics repository lacks the appropriate tools to capture proteomics experiment data to generate a XAR file. Currently, these XAR files are manually created by the data curators, which is a very time consuming and error-prone process.

protExpress system aims to resolve this issue. It will be a web-based annotation tool for proteomics experimental information. protExpress will be an open source project utilizing industry standard best practices and developed on the J2EE platform. It will provide a framework for describing experimental procedures and steps. An intuitive web-based interface will enable the users to input and manage experiment and protocol information, enabling researchers to specify an experiment with a series of protocols, with specific inputs and outputs. The software will allow for exporting the experiment information into the XAR format used by CPAS, while allowing for addition for support of other additional formats in the future. It will also provide programmatic API to allow access to the underlying data.

Chapter 1   Actors

The use cases for protExpress are defined in this section. An overview of the various actors involved is depicted in the following diagram:
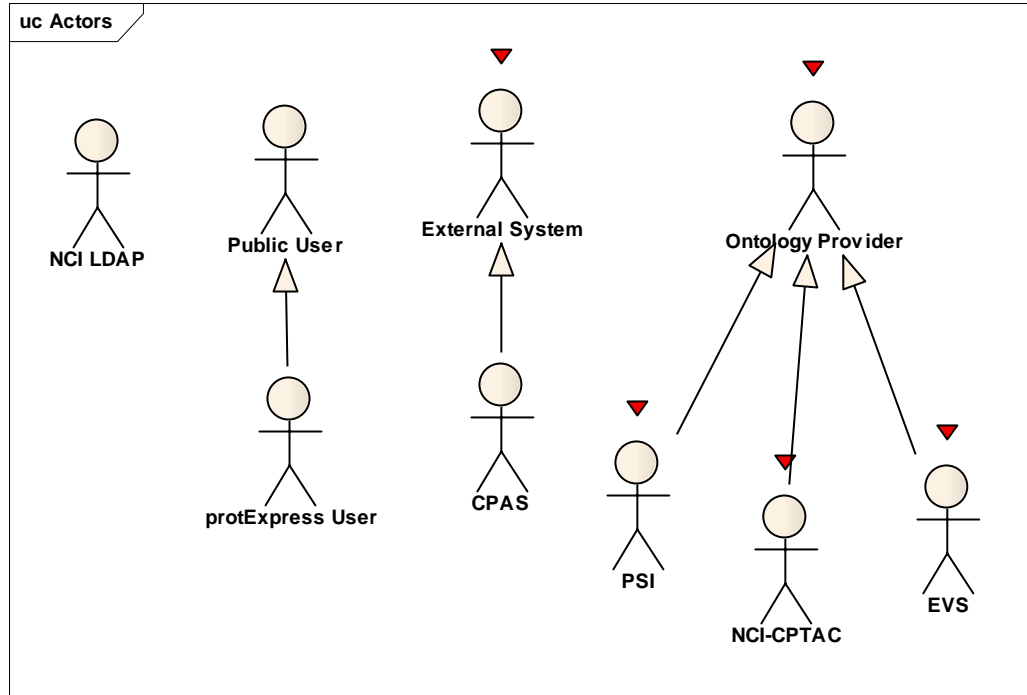
Public User

This actor represents a public user who can register for an account on the system.

protExpress User

This actor represents a registered user of the system. This actor has acquired the minimum credentials/privileges for accessing the non-public areas and functions of the system.

External System

A generalized actor intended to represent those applications/systems that programmatically communicate with protExpress. Examples of external systems include CPAS. Examples of programmatic access might include an API, web service, or grid service – specific mechanisms that will be incorporated have not been decided.

CPAS

This actor represents the CPAS system.

NCI LDAP

LDAP (Lightweight Directory Access Protocol) is a software protocol for enabling anyone to locate organizations, individuals, and other resources such as files and devices in a network (either public internet or a corporate intranet). LDAP is a "lightweight" (smaller amount of code) version of Directory Access Protocol (DAP), which is part of X.500, a standard for directory services in a network. LDAP is lighter because in its initial version it did not include security features. LDAP originated at the University of Michigan and has been endorsed by at least 40 companies.

Ontology Provider

A generalized actor that represents applications/systems that will provide ontology lookup services utilized by protExpress. Examples of such systems that we know protExpress will use are PSI, NCI-CPTAC and EVS.

PSI

This actor represents the Proteomics Standards Initiative (PSI) Controlled Vocabulary and Ontology. Background can be found from the following sources:

http://psidev.sourceforge.net/mod/

http://www.psidev.info/

NCI-CPTAC

This actor represents the NCI CPTAC Ontology.

EVS

This actor represents the NCI Enterprise Vocabulary Service. Additional information can be found at the URL:
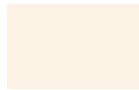
http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary

Chapter 2   Use Cases

The use cases for protExpress are defined in this section. An overview of the use cases is depicted in the following diagram.
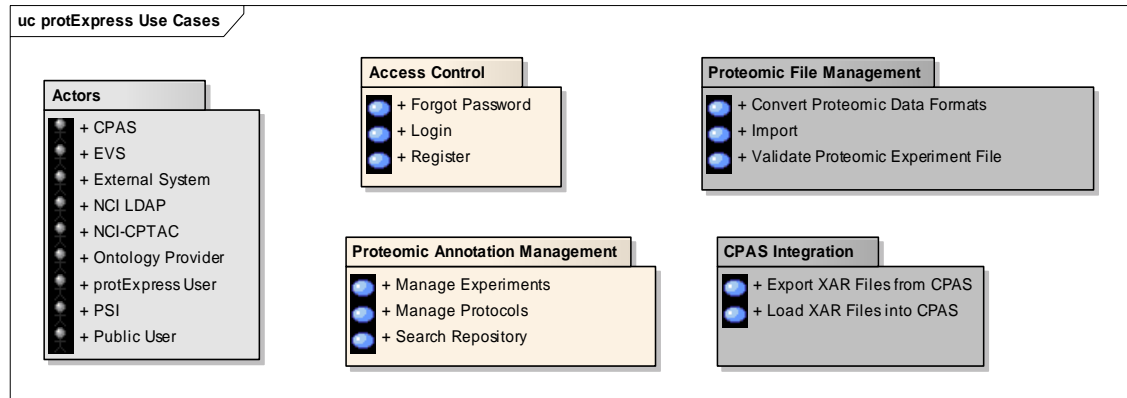
- Denotes actors in the system.

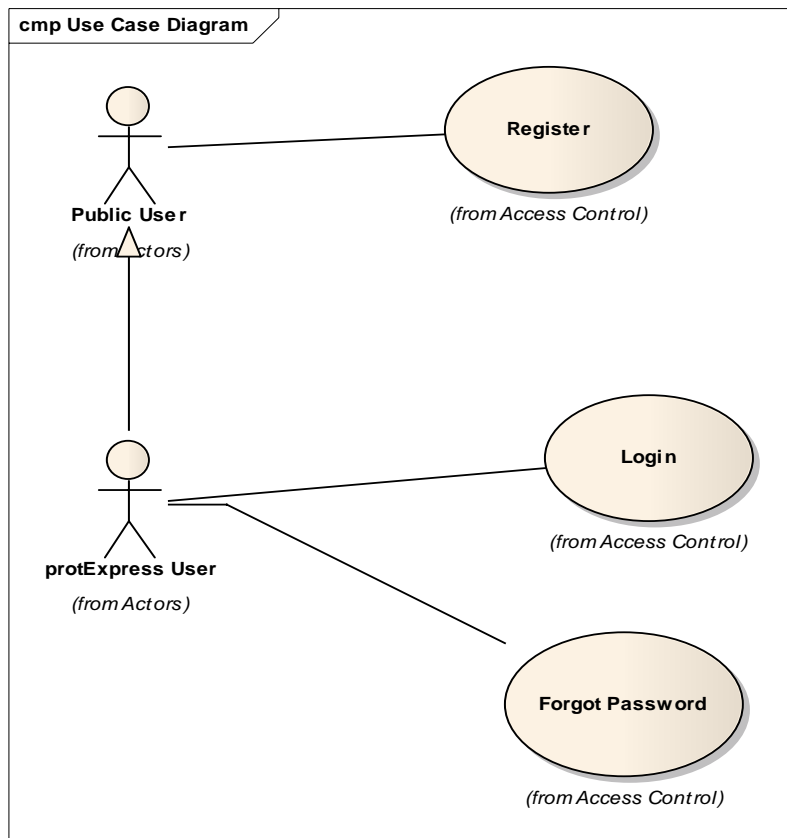- Denotes the use cases that will be implemented in protExpress 1.0.

- Denotes use cases that have been deferred and will not implemented in protExpress 1.0.

protExpress Use Cases

Access Control

Initiated by a user, the login use case allows for registration of a non-registered user for a new protExpress account, allowing them to login to access the application.

cmp Use Case Diagram

Register
(from Access Control)

Public User
(from Actors)

Login
(from Access Control)

protExpress User
(from Actors)

Forgot Password
(from Access Control)

Register

Initiated by a non-registered user, this use case allows the user to request a new account. This will enable the user to login to the system and perform certain actions for which they have the appropriate privileges.

Login

Initiated by registered protExpress user, this use case allows for the validation of the authenticity and authority of the given user either against a networked (LDAP) set of users or a local set (database). As a result of a successful login, the registered user is presented with their home space and the set of operations they have privileges to perform.

Forgot Password

Initiated by registered protExpress user, this use case allows the user to request help from the System Administrator to recover and/or reset their forgotten password, to enable them to re-access the system.

Proteomic Annotation Management

Initiated by a registered and logged in user, this use case enables the input of protocols, experiments, protocol inputs and outputs, and annotation information into the system. The ability to search the data is also provided.
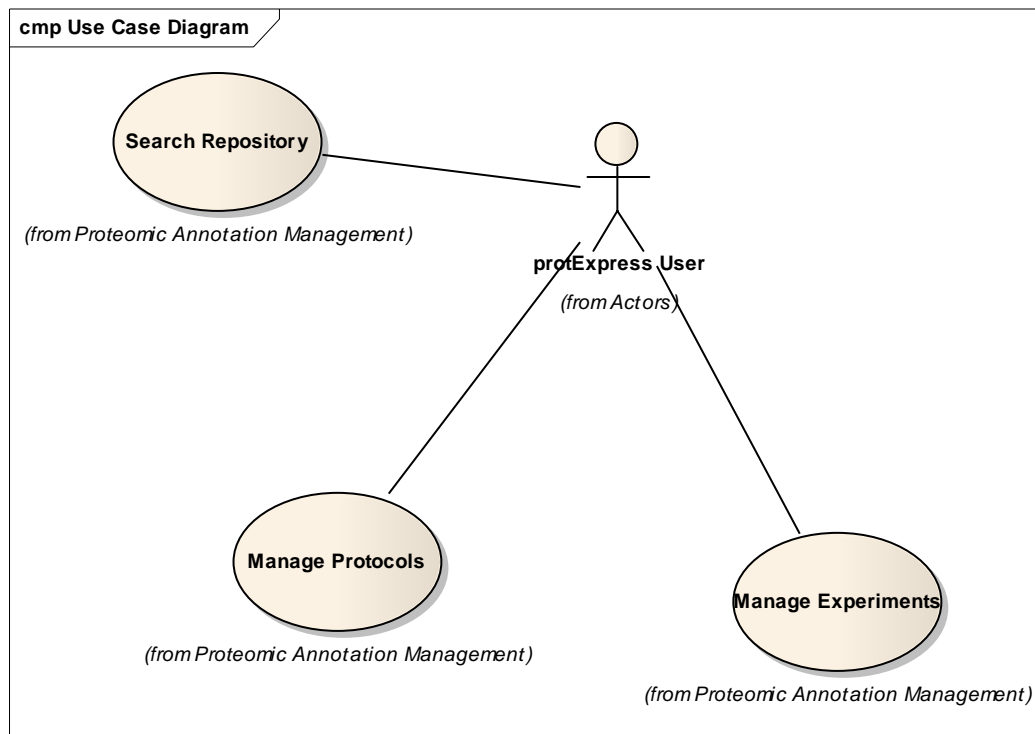
Manage Protocols

Initiated by a protExpress user, this use case allows the user to input and create new protocols, and modify/delete existing protocols from the system.

Manage Experiments

Initiated by a protExpress user, this use case allows the user to input and create new protocols, and modify/delete existing experiments from the system. The following functionalities are supported:

1. Create, modify and delete experiments.

2. Provide ability to repeat an experiment. This will result in multiple "runs" of the experiment, with each run representing an instance of the experiment. Each experiment has an individual set of protocol applications, with each protocol application having its own sets of inputs and outputs.

3. Export experiment data into an eXperimental ARchive (XAR) format used by CPAS.

   Note: It is envisioned that protExpress will be able to support a variety of formats. However, for version 1.0, only the XAR format will be supported by the system.

Search Repository

Initiated by a protExpress user, this use case allows a user of the system to search and view experiments and protocols present in the system. Checks are made to ensure that a user can only edit/delete the entities they have privileges to do so.

CPAS Integration

Initiated by a registered and logged in user, these use cases allow a registered user of the system to communicate with an installation of CPAS (Computational Proteomics Analysis System)

Note: This use case has been deferred and not implemented for protExpress 1.0.
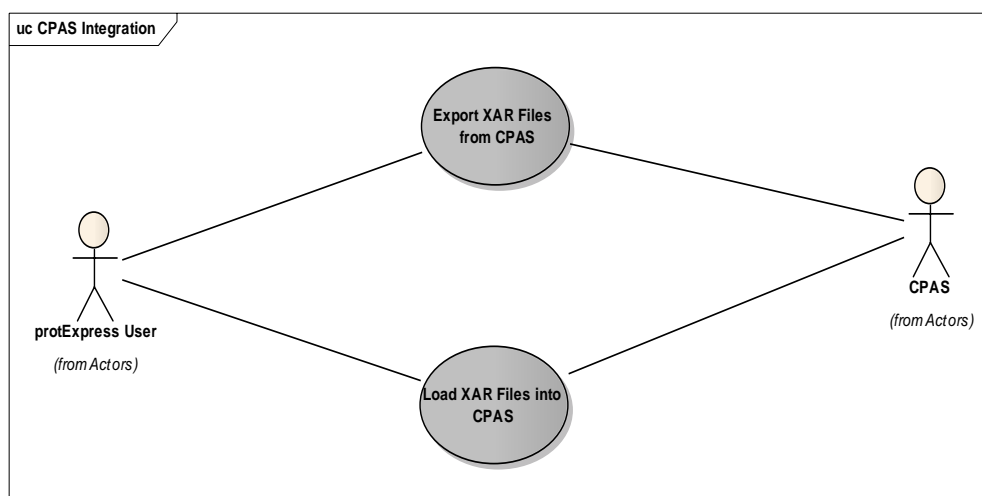
Load XAR files into CPAS

Initiated by a protExpress user, this use case allows the user to export the experiment information into a XAR formatted XML file. This file can be uploaded into a CPAS installation for further processing.

Note: This use case has been deferred and not implemented for protExpress 1.0.

Export XAR files from CPAS

Initiated by a protExpress user, this use case will allow an XAR formatted XML file (exported from CPAS) to be imported into a protExpress system.

Note: This use case has been deferred and not implemented for protExpress 1.0.

Proteomic File Management

Initiated by a registered and logged in user, this use case enables the conversion of experiment data from/into multiple formats supported by the system, in addition to validating the syntax and semantics of the generated format.

Note: This use case has been deferred and not implemented for protExpress 1.0.
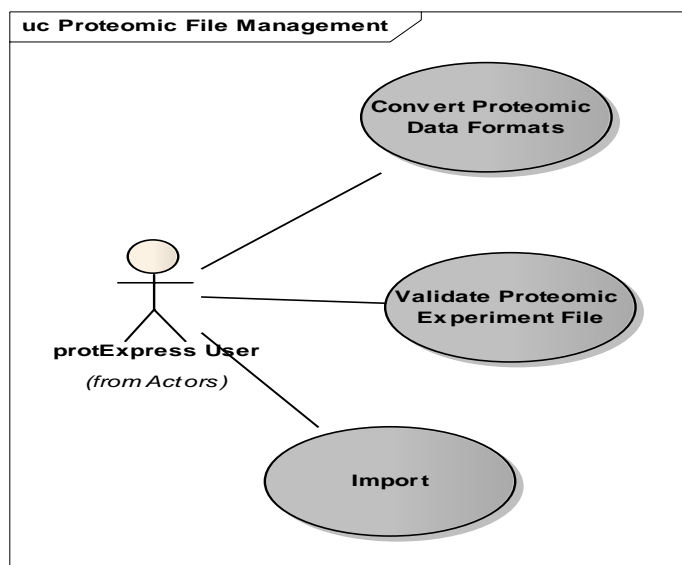
Convert Proteomic Data Formats

Initiated by a protExpress user, this use case will allow the user to retrieve an experiment data from the system and convert it to a supported format.

Note: This use case has been deferred and not implemented for protExpress 1.0.

Import

Initiated by a protExpress user, this use case will allow the user to import proteoemics experiment data from a supported format into the protExpress system.

Note: This use case has been deferred and not implemented for protExpress 1.0.

Validate Proteomic Experiment File

Initiated by a protExpress user, this use case allows the user to take a file containing proteomic experiment and annotation information, and validate the syntax and semantics of the experiment data using the protExpress system.

Note: This use case has been deferred and not implemented for protExpress 1.0.