

# Semantic Enabled Social-Collaborative Research Framework for Proteomics Domain

Nazmul Hussain<sup>1</sup>, Hai H. Wang<sup>2</sup>

Computer Science Department, Aston University, Birmingham, UK

Computer Science Department, Aston University, Birmingham, UK

hussan12@aston.ac.uk, h.wang10@aston.ac.uk

## Abstract

The importance of proteomics science is rapidly growing across all areas of biomedical research undefined for the measurement of cellular function. However, the most advanced proteomics technologies, knowledge and datasets have been widely unavailable to the biomedical research undefined communities. These limits make it harder to support the disease and drug-related research progress. Moreover, developing common data standards and the validation of deposition is difficult, slow, and an arduous task for any single group. Therefore, proteomics domain necessitates collaboration with diverse researchers for meaningful research progress. The lack of effective collaboration and sharing facilities among researchers with necessary biological expertise is another hurdle for properly designing studies, or developing data standards. The social network centric collaborative research framework could bridge that gap by bringing together expertise in different areas of proteomics, and providing a shared workspace for them by enabling research artifact sharing. Besides, it is a challenging task for a human administrator to manually manage, undefined undefined analyses, and take decisions in a traditional collaborative system. A promising solution to this challenging research problem is the application of the IBMs Autonomic Computing (AC) system. This paper presents a social networking enabled collaborative research framework that allows not only automatically finding relevant researchers, making social connections, co-authorships, but also collaboratively create and share research artifacts, forming a community that could influence research contribution. We also offer semantic richness to the framework by constructing ‘OntoProteomics’ ontology. Finally, we highlight the concept of the AC system that could enable automated functionalities in the collaborative environment.

## 1 Introduction

The application of Proteomic’s science in the biomedical study is useful to develop new diagnostic methods, identification of disease markers, or development and assessments of drug targets [1]. A challenge faced by researchers is the exchange of incompatible data that could lead to serious issues affecting result’s interpretation, reproducibility and analysis. To overcome this issue, it is important to create a common data standard for MS proteomics, with large acceptance by worldwide research communities [18]. However, proteomics data standard development is difficult and slow, because of no common understanding of requirements, little willingness to cooperate, continuous introduction of new data types and technologies. Moreover, developing data standards and validation of deposition is an arduous task for any group, and it is imperative for repositories, journals, and funding agencies to collaborate [1]. Because, collaborative approach becomes a valuable supporting system to provide easily collect, publish, and share complex scientific knowledge or data across research communities. The underlying idea of collaborative research is that different contributors can bring in key knowledge and skills into the system; increasing the likelihood of creating a meaningful innovative solution [33]. Proteomics research increasingly demands widespread collaborators with diverse laboratories, experimental workflows, instrumentation, software and data types, which require frequent exchange of data [18], sharing information, and new technologies. Sadly, the most advanced proteomics technologies, experimental methods, analysis tools, datasets, etc. have been widely unavailable to the biomedical undefined research community, which limits to aiding and support the disease and drug-related research progresses in the biomedical undefined field. In fact, collaborative research of proteomics frequently struggles with finding relevant contributors with the necessary biological expertise and skills to properly designing studies and analysing the data [1]. Obviously, developing more advanced tools for support-

ing collaborative research could bring a positive change in scientists' attitude and viewpoint [25]. Several studies have shown that the lack of effective tools, and a common framework are the main obstacles in finding relevant contributors [36, 37, 30], and difficult to collect and utilise necessary knowledge, information and resources not only for the biomedical community, but also proteomics community themselves. On the other hand, the lack of active connections, limited communication and interaction among heterogeneous research groups, resulting diversity of tools, knowledge, and resources in proteomics. As a result, undefined undefined most of the valuable knowledge, technologies and experimental datasets buried in the data-rich web without realizing its full potential and reuse by the biomedical research community. To bridge that gap, it is a great demand to bring diverse research communities or individuals together, and provide efficient knowledge creation and sharing platform that could enable designing and implementing proteomics data standard, appropriate study design and so on. At present, most collaborative work limited to between two or three research centres due to the complexity and unavailability of computational tools, resources, or platforms. For instance, proteomics research groups such as HUPO (Human Proteome Organisation) and PSI (Proteomics Standards Initiative) have been only working collaboratively on different types of proteomics investigation such as sample processing, mass spectrometry, protein modifications and so on. Moreover, the existing collaborative solution used by these groups is a seriously time-consuming process, expensive, and is unsuitable for problems that require quick and decisive action. Another complication of these existing systems is that the various interested groups may not be equally involved throughout the core development, such as data exchange and modelling standard, annotation guidelines, project plan, etc. [6]. Existing Web platforms (e.g. Research gate, Academia.edu, Github) and collaborative tools (e.g. Wiki, Google Docs, etc.) already are undefined undefined undefined undefined undefined undefined in place to support collaboration and knowledge sharing between scientific communities. These platforms are not sufficient to provide all necessary functionalities to fulfil all collaborative requirements from a single interface. For instance, undefined some of the system may only support the sharing and storage of general information, published paper, or bookmark; or some of them may only support basic social networking such as blogging, tagging, friendships, discussion forums; some of them only allow collaborative document writing or editing, code sharing, etc. Therefore, undefined it is difficult to use different systems for different purposes, which make the research progress time consuming, undefined undefined undefined expensive and management issues. Moreover, those solutions do not support the collaborative creation and shar-

ing of research artifacts (i.e. Experimental workflows, research methodology, data undefined exchange standards), collaborative authorship, forming community for a particular project. There is also a lack of semantic description of information and data about social entities, content, and resources on those platforms. Social networking concepts have the enormous potential to support collaborative research work, which could enhance connectivity among distributed contributors, improve communication, increase awareness, and facilitate sharing, organise and dissemination of findings [25]. A study reported that knowledge sharing within the community is influenced by the contributor's social relationships that motivate participation and enhance performance on the network [36]. Social tagging on an online platform has become one of the best ways of associating metadata with Web contents and useful for social interest discovery [13]. Social networking activity such as participation, tagging, management, optimization becomes extremely complicated and context-based [8]. undefined However, social network environment is very dynamic and complex, and social networking constantly grows and changes as new information and content created, collected, and stored, which require monitoring and re-analysis for each change [7]. Therefore, the analysis and decision-making in social networking are another challenging task for an individual; even with the help of computational tools [7]. The application of IBMs AC system could help to resolve this problem by enabling automated functionalities in the collaborative environment. In this paper, we propose a social network centric collaborative research framework for proteomics domain, which is semantically enriched. This framework, called 'SocialProteomics', undefined is departed from the existing idea in that sense, it will serve not only for social networking purposes undefined undefined undefined such as making friendship, communication, or tagging, but also collaboratively to create, publish and disseminate new findings, published paper, research artifacts, etc. An important feature of the framework that allows automated discovery of contributors with similar research interest by social tagging, bookmarking, and research artifacts. We outline the principles of AC system that can be utilised to automated discovery of a researcher, analyse the tags and bookmarks, notify to relevant people, and suggest necessary action. Finally, we provide the semantic richness of the framework by constructing 'OntoProteomics' ontology to describe the entities, their relationships, and activities within the collaborative network. This paper is organised as follows: Section 2 describes the background information on existing web tools and technologies used to facilitate collaborative research. Section 3 highlights social-collaborative aspects of proteomics research, entities involved, and core features within the framework. In Section 4, we proposed 'SocialProteomics' undefined frame-

work implementation based on proteomics research scenarios with the basic system interfaces. Section 5 shows the construction of ‘OntoProteomics’ prototype ontology with an example link to ‘SocialProteomics’. Section 6 outlines related work of the collaborative research support. Section 7 discusses the overall outcome of this research work and compares the proposed solution with the existing one. Finally, Section 8 draws a conclusion from this work in a direction for future work.

## 2 Background

This section presents a background study of state-of-the-art of social-collaborative research, and overviews the most relevant theories, approaches and tools that can serve as the basis for the designing and implementing a social network centric collaborative research framework.

### 2.1 Collaborative Research of Proteomics

In the post-genome era, proteomics research has advanced rapidly, with the exponential growth of the diverse experimental method and generation of large amounts of data [38]. Proteomics is a -omics science (e.g. Genomics, proteomics, epidemics), and sub-discipline of biomedical research that encompasses a large-scale and comprehensive study of the protein sequences, structures and functions, post-translational modifications, and complex formation in a cell, tissue or organism [1], and other processes relevant to understanding human health and various disease states. Collaboration is important with contributors of similar research interest or expertises to discover knowledge from rapidly accumulating biological data [37, 34, 25]. Collaboration can be defined as a process of participation of an individual, groups, organisations, or research centers, who work together in a sociable environment to systematically solve undefined existing and emerging problems that could not be solved by an individual or group alone [4]. Recently, the advancements in Web technologies made possible for easier collaboration, widespread access, and share cutting-edge scientific research data. For instance, Web 2.0 technologies (i.e. Wikis, blogs, social network, RSS, tagging, bookmarking) turned into a well-established medium for collaborative research, communication and interactions [23]. Social networking approach is a most promising Web 2.0 tools undefined undefined undefined undefined undefined to support collaborative research by encouraging participation and linking them together with similar research interests [37].

### 2.2 Social Network and Web 2.0 for Collaboration

Knowledge sharing is important in any scientific practice that allows other researchers to access, validate and re-analyse someone’s findings, the identification of errors and allows the results to be used in novel ways [32]. The current social networking system changes the way to share and communicate by bringing together broad new range of technologies [23]. Thus, this approach becomes a very popular infrastructure to support scientific collaboration through knowledge or data sharing and reuse, content annotations, and maintaining the social connections and communications among researchers [25]. A social network is a model that consists of a set of people, organisations, events and locations, which are connected by a set of social relationships, such as friendship, co-working, information sharing [10, 27]. Social networking service can be modeled as an undirected graph with nodes for the entities and an edge, which connects two nodes for the relationships of an event model. The nodes in the graph can represent not only people or groups but also their work artifacts and the edges can encode all kinds of relationships including friendship, professional or organizational memberships [10, 2]. The contemporary social networking service is considered as Web 2.0 tools. The major Web 2.0 tools that support general purpose undefined information and content sharing among research communities are described as follows:

- *Blogs* are a useful and easy way to build a profile as a researcher, trigger collaboration and establish group communication. [28] defines a blog as an informal online journal of the author’s thoughts that usually will focus on a specific topic or issue.
- *Microblog* is typically smaller that allow users to exchange small elements of content such as short messages, images, or video links. Researchers can use microblogging for their work in reaching a wider and broader audience quicker.
- *Social Network* is typically composed of users, groups, events and links between them located anywhere with similar interest. Social networking mainly involves making friendships, sharing ideas or opinions, and so on, which helps researchers keep track of their co-workers activities, communicate with them in an informal manner.
- *Wiki* is a very a powerful collaborative tool that encourages researchers to undefined collaboratively authoring scientific papers, edit documents, and discussing ideas in general.
- *Tagging* is a collection of keywords or a descriptive phrase, assigned to a content by an author or a user to

classify data, and makes information easier to search online. By tagging content and sharing these tags, researchers should be able to automatic discovery of contributors with similar research interest.

- *Social bookmarking* commonly refers to as social tagging exploit by an individual to categorization of larger sets of information and knowledge resources. It provides an opportunity to find more websites related to the topic of interest and learns about people interested in the same topic.
- *Feeds* is a technique, mostly used by enterprises to provide a subscriber with frequent updates from the website instead of visiting websites. It is also used to achieve awareness in collaborative development settings, and continuously monitor relevant information.

### 2.3 The Semantic Web and Ontology

Semantic Web, undefined also known as Web 3.0, proposes a new generation of the Web in which information is given well-defined meaning, context, description better enabling collaboration between humans and machines. The ultimate goal of the Semantic Web undefined is to build a generic framework that will allow data and knowledge to be shared and reuse effectively with wider communities, and enabling data to be linked from one source to another in a machine-readable format. The common standards and framework have been developed for sharing and reusing knowledge on the Web is briefly introduced as follows: *Resources Description Framework (RDF)* is a language has been developed in order to provide a flexible mechanism for describing web resources and their relationships. However, RDF provides very few mechanisms undefined undefined to address the problem of understanding the meaning of the terms used in annotation. Hence, *RDF Schema (RDFS)*, an extension of RDF, which has been designed to overcome the limitation of RDF specifications. Although, RDFS allows the hierarchical specification of classes and properties that enable simple inference, the expressive power of RDFS is yet quite limited and unable to support a number of commonly required features, such as negation or disjunction. To overcome this issue, *Web Ontology Language (OWL)* has emerged as an extension of RDF/RDFS to fulfil such types of demands. OWL is recommended by W3C to represent the web ontologies that have greater machine interoperability of the web content than the XML, RDF, and RDF Schema. *SPARQL Protocol and RDF Query Language (SPARQL)* is a semantic query language that ensures that the information stored in the form of triples can be retrieved from the repositories with high precision and recall, similar to the use of SQL in traditional relational

databases. *Ontology* is a very widespread mechanism to encode terminology standards, which provide definitions for terms in the controlled vocabulary as well as properties and relationships between terms [6]. The ontology is a formal conceptualization of a domain that specifies the classes of objects that exist, the relationships among those classes, the possible relationships among instances of the classes, and constraints over those instances [10].

### 2.4 Semantic and Social Network

Modeling of social networks can be aided by ontologies and reasoning to aggregate it for the specification and explication of concepts and social relations among them. Researchers can build models of social based online research network using social ontologies like FOAF (Friend of a Friend), SIOC (Semantically-Interlinked Online Communities), and SKOS (Simple Knowledge Organisation System) [9]. FOAF ontology allows to describe and publish people profile information online, including their social links and activities [10, 24]. In contrast, SIOC describes the social network contents such as blogs, forums, mailing lists and facilitates interconnection of these contents via publishing of metadata [24]. Moreover, the Dublin Core (DC) is the most important metadata standard of the library world. The use of DC terms and its refinements in the metadata attached to each blog post can efficiently index created data and then to easily retrieve undefined undefined information [22].

### 2.5 Autonomic Computing

Autonomic Computing (AC) is an emerging paradigm that offers a promising solution to manage IT environment by enabling self-management capabilities based on situations they observe [29]. This approach was inspired by the way human Autonomic Nervous System that operates the body-related parameters, such as heart rate or temperature, thereby releasing the brain from controlling them in a fully aware manner [35]. Autonomic computing can be employed to improve collaborative information processing, and data aggregation in the social network so that expensive communication costs of large amounts of data can be substantially reduced, and the system-wide self-management is achieved [16]. An autonomic computing system described as having four key properties, which coined as self-CHOP [16], and are described in the social networking context as follows:

- *Self-Configuration*. The system is able to dynamically configure itself, adapt to changes in the IT environment [15]. This property deal with the new or existing people in the social network such as suggest some prioritized contact with the new member based on profile characteristics [29].

- *Self-Healing*. The system will detect the problematic operation and then initiate corrective action without disrupting system applications [15]. This property can be used to address issues in the social network such as knowledge that does not flow or the weakly linked members [29].
- *Self-Optimization*. The system will efficiently maximise resource allocation and utilisation to meet end users' needs and addresses the complexity of managing system performance [15]. This property is applied to address the issues related to balancing such as suggesting new relationships, optimise the personal interests and available communication resources [29].
- *Self-Protection*. The system protects itself from unauthorised people to access data, malicious attacks, and so on [15]. This property can address the issues related to people entering the social network such as detecting, reallocation or even shield the problematic people [29].

IBMs suggested a reference model for autonomic control loops [17], which is sometimes called the MAPE-K (Monitor, Analyse, Plan, Execute, Knowledge) loop [14]. The autonomic loop MAPE-K is used for four major operations: monitoring, analysing, planning, and executing, which operate taking into account the *knowledge* the autonomic element has on the environment and itself.

### 3 Social-Collaborative Aspects of Proteomics

In this section, we present 'SocialProteomics' a collaborative research framework within the context of the proteomics research domain.

#### 3.1 Entities of 'SocialProteomics'

The common types of entities usually engaged in the 'SocialProteomics' collaborative research frameworks are:

- *Person*. This type of entity expresses an individual researcher, who is willing to contribute to the problem area of his interest and wants to share personal knowledge, skills, or findings. For instance, entities generally may include academics, research students, domain experts, and end users.
- *Organisation*. This category of entities can be funding organizations, research centers, Universities, Hospitals, Journal or book publishers who have identifiable existence and involved in collaborative research in various ways.
- *Research Community*. A research community is a self-organizing group that meets to share information and gain insight about a particular issue or challenge and to learn from one another. A number of researchers with similar interests, aims, and attributes together form a community by establishing working relationships with each other.
- *Events*. This type of entity may include a conference, seminar, meeting, or a training session where a large group of researchers formally gathers to present their work.
- *Resources*. The resources may include published papers, datasets, analysis tools and so on, which are needed for meaningful research progress.
- *Research Artifacts*. Artifacts are the active component and necessary to coordinate and even trigger or initiate the collaborative research work by providing a shared workspace. The example of proteomics research artifacts may include (analysis pipeline, digital material and methods [26], where many researchers may be connected together by interacting and making contributions through the research artifact.

#### 3.2 The Social Features of 'SocialProteomics'

In this study, we considered several key features of the 'SocialProteomics' collaborative research frameworks that could support proteomics research by enabling various social activities are briefly discussed below:

##### 3.2.1 Connections

Researchers establish social connections with others by following their updates, responding and commenting on their content or ideas, by tagging or bookmarking the contents of interest. Thereafter, the interactions turned into a strong social relationship like friendship, membership, co-authorship and even an artifact's relationship. The common type of social connections may include:

- *Friendships*. A researcher may request or accept others as a friend (mostly considered direct relationships [36]) and then leverage their relations to share, exchange and discuss knowledge, ideas, or content about a research topic more informally. This will allow the researcher to keep track of each other's research activity, status updates, new publications, or findings.
- *Memberships*. A group of researchers with similar interest may establish an online research community,

and start a new collaborative project, and invite and accept membership from the relevant participants interested to contribute to the project.

- *Co-authorships.* In collaborative research, co-authoring, a scientific article is one promising approach. A scientific paper is co-authored if it has more than one author. One of the fundamental features of the framework is that it allows participation in co-authorship of publishing scientific paper or Web article.
- *Human-Artifacts Connection.* A researcher connects with research artifacts depending on his or her interest and skills on the topic. Certainly, the connection also establishes between not only researchers and artifacts, but also between artifacts.

### 3.2.2 Communications

The primary ability of the framework leads informal conversations that make it a powerful communication tool for collaborative work. The communication link could establish through discussion, review changes, engage in peer debate, disseminating findings or ideas.

### 3.2.3 Sharing and Dissemination

Proteomic's knowledge, resources, technologies, research artifacts could be created, shared and disseminate among communities. This can be either directed sharing, or targeting one or more specific researcher within the network.

### 3.2.4 Store and Access Data, Information or Resources

The scientific paper, experimental datasets, lists of bookmarks are stored and accessed to and from distributed sources for privacy, security and confidentiality purposes. For example, BibSonomy [3] a third party publication management system that serves to store and share published papers as well as bookmarks created by researchers, and ProteomeXchange [1] an international consortium, aspires to serve as a single point of MS data submission for PeptideAtlas, PRIDE, and Tranche to facilitate data exchange between major proteomics dataset repositories.

### 3.2.5 Tagging and Bookmarking

An author of a paper or an owner of a dataset can generate tags to locate and keep track of interaction of others with the generated content or shared resources. Using social bookmarking services, a researcher can create a comprehensive resource list, simplifies the sharing of resources among worldwide biomedical research communities. The

researcher can also add semantic tags and edit information about bookmarks such as a title of the web page.

### 3.2.6 Semantic Annotations

Annotation is the process of associating metadata (e.g. MS experimental conditions, data terms, and data structures [1]) with datasets, resources, artifacts, images, whereas a semantic annotation attaches semantic metadata to the resources, pointing to concepts and properties in the ontology. The framework offers a simple yet robust infrastructure of semantic annotations of the content with links to the ontology.

### 3.2.7 Publish and Subscribe

A researcher could publish his content or interest to the network, and other researchers can then subscribe to receive regular updates about new content or the latest updates on research activities. Researchers also subscribe to feed on numerous web resources relating to their research interest, which helps to uncover the latest knowledge, data or publications that may be valued in his own research.

## 3.3 Autonomic Functionalities

In 'SocialProteomics', autonomic features could be achieved through the utilisation of IBMs Autonomic Management Engine (AME) tools, which includes built-in representations of the four parts of the control loop (monitor, analyze, plan, and execute) [15]. We highlight the basic autonomic features of the framework as follows:

### 3.3.1 Managing Entity Behaviour

Entry or exit of different entities in the network, their interactions, communication, collaboration, AME monitors such event, trigger analysis, plan and execute actions based upon the policy determined by a domain expert. For example, this policy could suggest some prioritized contact with a new member based upon the similarity of user profile, annotations, and interest they have. This can trigger a collaboration session such as which researcher should collaborate with who for a given purpose.

### 3.3.2 Managing Research Artifacts

Managing research artifacts is a complex task for an individual because researchers may work with collections of items associated with an experiment. Optimizing such event is important in order to maintain shared workspaces for the participants for a particular research project.

### 3.3.3 Monitor, Analysis, and Notifications

Incorporation of tagging and social bookmarking idea into the collaborative environment could improve the decision-making and notification of events to relevant individuals. For instance, adopting AME can track the status of an entity's behavior, analyses the interactions and similarities with all tags, and bookmarks generated by an individual, and then notify the relevant individual for taking some action.

## 4 'SocialProteomics' Framework Implementation

In this section, we present a general architecture and implementation of a prototype 'SocialProteomics' framework to depict social-collaborative aspects of proteomics research discussed above. We demonstrate a few example scenarios for establishing connections and communication among researchers with similar interest. However, the implementation of autonomic functionalities, and the interoperability with BibSonomy and ProteomeXchange is not the scope of this study.

### 4.1 'SocialProteomics' Framework Architecture

The architecture of the framework consists of a social-collaborative environment with the space for collaborative research artifacts, and 'OntoProteomics' ontology extended with the vocabulary term and properties from DC, FOAF and SIOC enhance the semantic capabilities. This framework is also interoperable with BibSonomy and ProteomeXchange repository to access, store and share published papers, bookmarks and MS proteomics data. The IBMs AME tool monitor and analyses the node's events and suggests appropriate actions. The architecture of 'SocialProteomics' framework is shown in **Figure 1**.

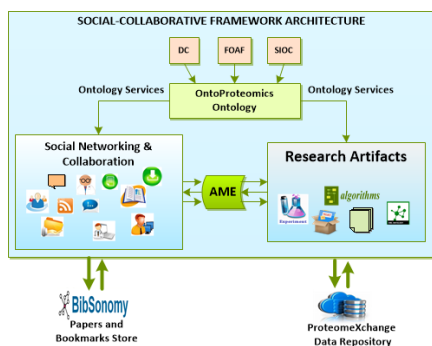


Figure 1: The general architecture of 'SocialProteomics' Framework

### 4.2 Establish Connections between Entities

The social connection is established between entities according to their research interest and activities. Each of the entities represented as nodes in the graph and their relationships using edges between the corresponding nodes. In this section, it illustrates example connection scenarios based on the tagging and social bookmarking.

#### 4.2.1 Scenario A - Connection based on Tags

As shown in **Figure 2**, *Researcher A* has a profile page *http:Profile A* with a set of tags *Tag A* with some useful keywords (e.g. proteomics, ms experiment, biomarker) that indicates his skills, expertise, research interest, and so on). He designed and published a *Experimental Workflow* for *Biomarker* identification. Shortly, *Researcher B* newly joined to the network and create a profile page *http:Profile B* with his research biography and created another set of tags *Tag B* with keywords (2D gel page, biomarker, proteomics, ontology, etc.).

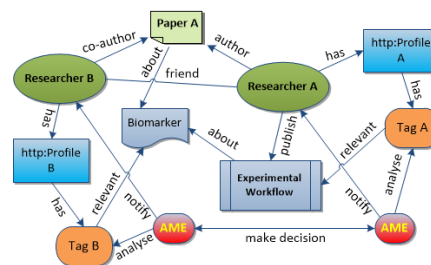


Figure 2: Connection by similar tagging

AME is an IBMs autonomic toolkit attached to the framework, which analyze the both sets of tag and find many similarities between *Tag A* and *Tag B*, and make a decision that both researchers are expertise and interested in the similar research topic and should be connected together. At this point, AME analyses the similarities and notifies the both researchers to take some action. Thereupon, both researchers establish social connections by requesting and accepting friendships with each other. After becoming friends, they started communication, interaction, and sharing knowledge or published paper. In the end, they may start writing a research paper collaboratively about the topic *Biomarker*. In this way, each of the researchers easily can discover each other within the network that has similar research interest.

#### 4.2.2 Scenario B - Connection based on Social Bookmarking

As shown in **Figure 3**, Scenario B represents extended connections among researchers. It shows that *Researcher C* is also a participant of the network who has a profile

page *http:Profile C* with a set of tags *Tag C*. He saves a bookmark link *Bookmark A* in his profile page with a set of tags *Tag D* with the keywords (mass spectrometry, protein, peptide, biomarker), which is about *Biomarker* topic, and stored the bookmark in **BibSonomy**, which can be viewed and tagged by other researchers too.

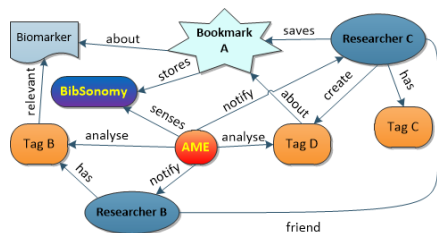


Figure 3: Connection by bookmarking

*Researcher B* already have specified *Tag B* in his profile page. Now, **AME** senses the changes in **BibSonomy** about newly added bookmarks and then analyse the tag sets *Tag B*, *Tag D* and finds the similarities between the tag sets, which indicates *Researcher C* also has similar research interest. **AME** makes a decision and sends notification to each of the creators of the tags to take some action. Both researchers establish a social connection by becoming a friend of each other.

#### 4.2.3 Scenario C - Connection by Research Artifacts

Another way to make connections among researchers by using research artifacts in the network. Research artifacts allow connections not only among researchers and their work artifacts, but also between artifact to artifact, where researchers can use it to find projects ideas, the creator to make a contribution to the problem.

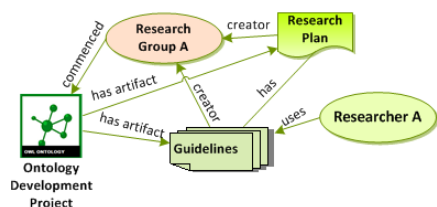


Figure 4: Contributors connected with artifacts

**Figure 4** illustrates an example of making a connection between contributors and research artifacts. Assume that *Ontology Development* is a collaborative research project for protein ontology, commenced in the network by a research community *Research Group A*. The project has artifacts *Research Plan* and a *Guidelines*, which describes the project details, and how to use the project plan. *Researcher A* uses the guidelines to contribute to the project. In this way, *Research Group A*, *Researcher A* become a

contributor of *Ontology Development Project*, which establish a connection as a member of *Research Group A*.

#### 4.2.4 Forming Research Community

Forming research community involves a wide range of social and research activities, and creating free spaces for researchers to come together for collaborative work. As follows from the scenarios shown above, demonstrates that researchers automatically discovered, notified and get connected together via tagging, bookmarking and research artifacts with the aid of AME. After building social connections, all of the connected researchers become aware of each other’s status, skills, activities, and research interest. Hence, they form a research community and begin to collaborate, sharing data, knowledge or ideas for solving a specific proteomics problem of their interest.

#### 4.3 Establish Communications between Entities

Creating a dissemination forum could be a persuasive method that can be utilized to establish a communication link with experts to disseminate new findings. In this example, we show how two or more researchers with similar interest communicate with each other to disseminate research finding.

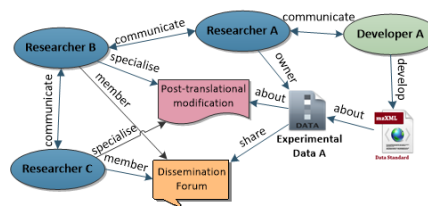


Figure 5: Communication through Dissemination Forum

**Figure 5** illustrates, researchers communicate through the dissemination forum. Assume that *Researcher A* is an owner of *Experimental Data A*. This data is about *Post-translational Modification* and shared with *Dissemination Forum* for experts’ opinion. *Researcher B* and *Researcher C* are the two specialists in *Post-translational Modification* and member of the forum. *Developer A* involves developing *Data Standard* for experimental data and communicate with *Researcher A* for requirements. They review, discuss, and provide feedback to the owner. In this way, communication could begin among the researchers.

#### 4.4 System Interface

We present a simple system interface of the ‘SocialProteomics’ framework, as follows:



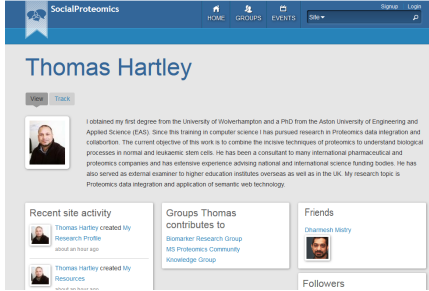


Figure 6: Example profile page of a researcher

The **Figure 6** shows the profile page of “Thomas Hartley” with his research interest.

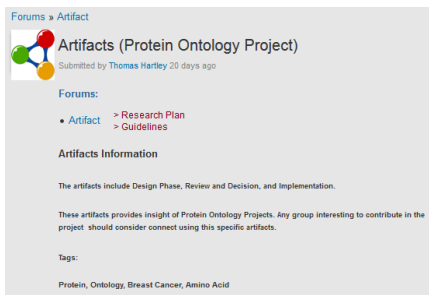


Figure 7: Example of a Research Artifact

The **Figure 7** shows an example published artifacts (i.e. Research Plan, Guidelines) for “Protein Ontology Project”.

## 5 Prototype Ontology Model for ‘SocialProteomics’

Of course, a particular ontology is not able to provide all the required representation of domain knowledge. Therefore, we considered to construct a ‘OntoProteomic’ prototype ontology to describe researcher’s biography, interests, activities, publications, and research artifacts. The ontology model, encoded in OWL using Protégé 4.3.

### 5.1 ‘OntoProteomics’ Ontology Construction.

In ‘OntoProteomics’ ontology, the core modules defined as base classes that represent the concepts of proteomics entities, and the properties describe their relationships. Additionally, the ‘OntoProteomics’ ontology has been extended by borrowing vocabulary terms from FOAF and SIOC ontologies to describe social relationships and activities of entities. This merging approach could enhance the capabilities of ‘OntoProteomics’ ontology to provide

maximum benefits for the ‘SocialProteomics’. **Figure 8**, shows the basic base classes of the ontology.

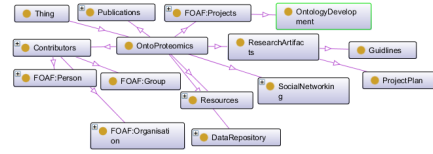


Figure 8: Overview of ‘OntoProteomics’ Ontology

### 5.2 Linking ‘SocialProteomics’ with ‘OntoProteomics’

In this section, we demonstrate the relatedness between ‘SocialProteomics’ framework and ‘OntoProteomics’ Ontology. We give an example representation of social-collaborative network information with the ‘OntoProteomics’ ontology. In this example, same colour coded oval shape represents class-subclass relationships with text *rdfs:subClassOf* property, and the rectangular shape represents an individual of the class.

**Figure 9** shows, *OntologyDevelopment* is a subclass of *foaf:Project*, *ResearcherA* is a subclass of *Contributors*, which is a subclass of *foaf:Person*. *Project-Plan* and *Guidelines* is a subclass of *ResearchArtifacts*.

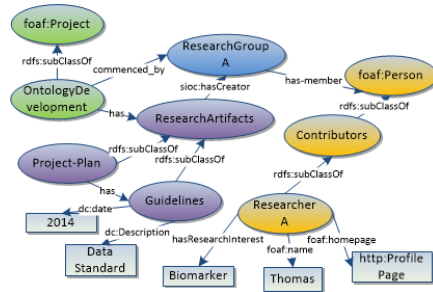


Figure 9: Example of ontology extension with DC, FOAF and SIOC

The project *OntologyDevelopment* commenced by a proteomics research group *ResearchGroupA*. This project has *ResearchArtifacts* such as *ProjectPlan* and *Guidelines*, and the creator of these artifacts is *ResearchGroupA* and the relationships denoted by *sio:hasCreator* property. Properties of an artifact *Guidelines* describe using the DC terms such as *dc:Date* is “2014”, and *dc:Description* is “Data Standard”. The *ResearchGroupA* has some members from *foaf:Person* class, such as *ResearcherA*, the properties described using the FOAF vocabulary terms such as *foaf:name* is “Thomas”, *foaf:homepage* is “http:Profile Page”, and *foaf:hasResearchInterest* in “Biomarker”.

## 6 Related Work

Taking into account the need for a large-scale sharing of proteomics information and data, there is very little research exists on to construct the most effective research activity designs and encourages participants in group discussion for knowledge sharing [36]. [23] analysed the role of different web technologies in the enterprise social collaboration platforms. They argued that social collaboration keeping track of the relationships between events and people. Conversely, research carried out by [36] and reported the usefulness of social networking and Web 2.0 service for identifying communities of practice. [37] proposed a Web 2.0 based Scientific Social Community (SSC) model for communication, data sharing or collaboration among users. [21] investigated the structure of scientific collaboration networks and assumed that two scientists are considered connected if they have co-authored a paper together. However, [2] introduced in their Codebook project that people can also be connected by becoming a friend with work artifacts. [19] cited that the purpose of tags to extract the intent of the user and facilitate goal-oriented a social network. Potentially, [10] investigates the way that the semantic web is utilized to represent and process social network information. Their finding shows that FOAF ontology is the most widely used domain ontology on the semantic web. [12] stated that large web-based social networks are starting to share members' information and their social connections in FOAF format, making millions of profiles available online. The Semantic Web for Research Community (SWRC) ontology project is a big effort, which generally models key entities relevant to typical research communities and the relations between them [31]. On the other hand, BibSonomy system developed for sharing bookmarks and published literature among scientist [3]. In contrast, DBLP (Digital Bibliography and Library Project) is another joint service of the University of Trier and Schloss Dagstuhl that provides open bibliographic information on major computer science journals and proceedings [20].

## 7 Discussion

Collaboration is the most important capability for any research domain to maximise the problem-solving and discovery of new knowledge. Using traditional collaboration tools, many scientists are not enthusiastic to engage with the online based research contribution due to the lack of time, efficiency of the system, and not seeing the many benefits of employing in a research context. Our study simplified that social networking approach can offer a wide range of features for communication, dynamic collaboration, forming community, as well as knowledge and

research artifacts creation and sharing.

However, there are many collaborative tools and applications developed for bringing large numbers of people together, but these are not enough flexibility for a specific research field that could fulfil all required functionalities from a single interface. For example, existing system such as Research gate, Academia.edu, Github, Wiki, Google Docs support only collaboration and knowledge sharing among individuals for a general purpose. Despite the fact that, GitHub and Wiki have the broader approach to collaboration, is not ideal for co-authorships. On the other hand, CiteULike is a largest collaborative tagging service for bibliographic references only. In contrast to BibSonomy, only references imported from one of the supported digital libraries appear on the central web pages [3]. Moreover, these systems are lack of semantic and less autonomous, and do not fit purely for collaborative research artifact creation. Therefore, the development of alternative services is necessary for proteomics for the powerful community interactions. In terms of semantic issues, DC, FOAF and SIOC ontologies provide necessary vocabulary terms for representing metadata and social networking of online communities, but not sufficient to describe a collaborative research environment. Although, SWRC provides a useful concept in general research context, yet missing the representation of research artifacts and social networking in proteomics perspectives. BibSonomy is very useful for sharing published paper and bookmark, but DBLP do not allow these services and limited to sharing computer science bibliography only.

## 8 Conclusion and Future Work

The success of biomedical research is largely dependent on the quality and quantity of proteomics data, knowledge, and resources available and shared among scientific communities. The validation of data deposition, designing experimental workflow and disseminate new proteomics technologies with widely accepted standard, necessitate a collaborative research approach. In this study, we have chosen data and resource-intensive proteomics domain, which suffers from inefficient collaboration tools. We addressed a number of issues concerning the key barriers of the collaborative research progress in the context of proteomics domain. We believe that the 'SocialProteomics' framework will accelerate collaborative research of proteomics and increase the global impact. The uniqueness of our work is that social interactions can establish among not only researchers or communities, but also research artifacts. We have also shown the semantic capability of the framework by constructing prototype 'OntoProteomics' ontology. Additionally, we highlighted an interesting approach of IBMs autonomic computing paradigm

that could enable automated features in the framework.

The future work is to develop an autonomic computing adapter to achieve self-managed and adaptive functionalities in social-collaborative environment.

## References

- [1] L. B. Becnel and N. J. McKenna. Minireview: Progress and challenges in proteomics data management, sharing, and integration. *Molecular Endocrinology*, 26(10):1660–1674, 2012.
- [2] A. Begel and R. DeLine. Codebook: Social networking over code. In *Software Engineering-Companion Volume, 2009. ICSE-Companion 2009. 31st International Conference on*, pages 263–266. IEEE, 2009.
- [3] D. Benz, A. Hotho, R. Jäschke, B. Krause, F. Mitzlaff, C. Schmitz, and G. Stumme. The social bookmark and publication management system bibsonomy. *The VLDB Journal—The International Journal on Very Large Data Bases*, 19(6):849–875, 2010.
- [4] A. Bergstrom, R. Clark, T. Hogue, T. Iyechad, J. Miller, S. Mullen, D. Perkins, E. Rowe, J. Russell, V. Simon-Brown, et al. Collaboration framework: Addressing community capacity. *Fargo, ND: National Network for Collaboration*, 1995.
- [5] J. Breslin and S. Decker. The future of social networks on the internet: the need for semantics. *Internet Computing, IEEE*, 11(6):86–90, 2007.
- [6] S. A. Chervitz, E. W. Deutsch, D. Field, H. Parkinson, J. Quackenbush, P. Rocca-Serra, S.-A. Sansone, C. J. Stoeckert Jr, C. F. Taylor, R. Taylor, et al. Data standards for omics data: the basis of data sharing and reuse. In *Bioinformatics for Omics Data*, pages 31–69. Springer, 2011.
- [7] R. T. da Silva, J. M. de Souza, and J. Oliveira. Autonomic analysis of social networks. In *Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on*, pages 508–515. IEEE, 2011.
- [8] D. Dasgupta and R. Dasgupta. Social networks using web 2.0. *ibm.com/developerWorks*, 2009.
- [9] G. Erétéo, F. Limpens, F. Gandon, O. Corby, M. Buffa, M. Leitzelman, P. Sander, et al. Semantic social network analysis, a concrete case. *Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena*, pages 122–156, 2010.
- [10] T. Finin, L. Ding, L. Zhou, and A. Joshi. Social networking on the semantic web. *Learning Organization, The*, 12(5):418–435, 2005.
- [11] A. Garcia-Perez and R. Ayres. Wikifailure: the limitations of technology for knowledge sharing. *Electronic journal of knowledge management*, 8(1), 2010.
- [12] J. Golbeck and M. Rothstein. Linking social networks on the web with foaf: A semantic web case study. In *AAAI*, volume 8, pages 1138–1143, 2008.
- [13] M. Gupta, R. Li, Z. Yin, and J. Han. Survey on social tagging techniques. *ACM SIGKDD Explorations Newsletter*, 12(1):58–72, 2010.
- [14] M. C. Huebscher and J. A. McCann. A survey of autonomic computing—degrees, models, and applications. *ACM Computing Surveys (CSUR)*, 40(3):7, 2008.
- [15] B. Jacob, R. Lanyon-Hogg, D. K. Nadgir, and A. F. Yassin. A practical guide to the ibm autonomic computing toolkit, 2004.
- [16] H. Kang, X. Li, and P. J. Moran. Autonomic sensor networks: A new paradigm for collaborative information processing. In *Dependable, Autonomic and Secure Computing, 2nd IEEE International Symposium on*, pages 258–268. IEEE, 2006.
- [17] J. O. Kephart and D. M. Chess. The vision of autonomic computing. *Computer*, 36(1):41–50, 2003.
- [18] K. L. R. M. C. T. A. P. R. A. P. H. M. I. A. J. Kevin Jones, James DeGreef and E. Nilsson. A collaborative approach to developing proteomics data standards, 2013.
- [19] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 195–202. ACM, 2009.
- [20] M. Ley. Dblp: some lessons learned. *Proceedings of the VLDB Endowment*, 2(2):1493–1500, 2009.
- [21] M. E. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.
- [22] A. Passant. Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs. In *International Conference on Weblogs and Social Media*, 2007.

- [23] M. Polaschek, W. Zeppelzauer, N. Kryvinska, and C. Strauss. Enterprise 2.0 integrated communication and collaboration platform: A conceptual viewpoint. In *Advanced Information Networking and Applications Workshops (WAINA), 2012 26th International Conference on*, pages 1221–1226. IEEE, 2012.
- [24] L. Razmerita, M. Jusevičius, and R. Firantas. New generation of social networks based on semantic web technologies: the importance of social data portability. In *Proceedings of the International Workshop on Adaptation and Personalization for Web*, volume 2, pages 79–87, 2009.
- [25] P. Romano, R. Giugno, and A. Pulvirenti. Tools and collaborative environments for bioinformatics research. *Briefings in bioinformatics*, 12(6):549–561, 2011.
- [26] M. Roos, S. Bechhofer, J. Zhao, P. Missier, D. R. Newman, D. De Roure, and M. S. Marshall. A linked data approach to sharing workflows and workflow results. In *Leveraging Applications of Formal Methods, Verification, and Validation*, pages 340–354. Springer, 2010.
- [27] D. Rosen, G. A. Barnett, and J. H. Kim. Social networks and online environments: when science and practice co-evolve. *Social Network Analysis and Mining*, 1(1):27–42, 2011.
- [28] J. A. Sagotsky, L. Zhang, Z. Wang, S. Martin, and T. S. Deisboeck. Life sciences and the web: a new era for collaboration. *Molecular systems biology*, 4(1), 2008.
- [29] O. J. Silva, R. and J. De Souza. Improving collaborative knowledge flow design on social networks through autonomic computing systems properties. 2010.
- [30] P. Singh and N. Shadbolt. Linked data in crowdsourcing purposive social network. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 913–918. International World Wide Web Conferences Steering Committee, 2013.
- [31] Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, and D. Oberle. The swrc ontology–semantic web for research communities. In *Progress in Artificial Intelligence*, pages 218–231. Springer, 2005.
- [32] J. A. Vizcaíno, R. G. Côté, A. Csordas, J. A. Dianes, A. Fabregat, J. M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, et al. The proteomics identifications (pride) database and associated tools: status in 2013. *Nucleic acids research*, 41(D1):D1063–D1069, 2013.
- [33] F. Wang, C. Rabsch, P. Kling, P. Liu, and J. Pearson. Web-based collaborative information integration for scientific research. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 1232–1241. IEEE, 2007.
- [34] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218. Acm, 2009.
- [35] M. Wódczak. Autonomic computing enabled cooperative networked design. 2014.
- [36] S. J. Yang and I. Y. Chen. A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network. *International Journal of Human-Computer Studies*, 66(1):36–50, 2008.
- [37] Z. Zhang, K.-H. Cheung, and J. P. Townsend. Bringing web 2.0 to bioinformatics. *Briefings in bioinformatics*, 10(1):1–10, 2009.
- [38] G. Zheng, H. Li, C. Wang, Q. Sheng, H. Fan, S. Yang, B. Liu, J. Dai, R. Zeng, and L. Xie. A platform to standardize, store, and visualize proteomics experimental data. *Acta biochimica et biophysica Sinica*, 41(4):273–279, 2009.