KNOWLEDGE DISCOVERY IN MICROARRAY DATA OF BIOINFORMATICS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

FAHRİ SALİH KOCABAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF HEALTH INFORMATICS

JUNE 2012

KNOWLEDGE DISCOVERY IN MICROARRAY DATA OF BIOINFORMATICS

Submitted by **FAHRİ SALİH KOCABAŞ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems, Middle East Technical University** by,

Prof.Dr.Nazife BAYKAL ______________________
Director, **Informatics Institute**

Assist.Prof.Dr.Yeşim AYDIN SON ______________________
Head of Department, **Health Informatics**

Prof.Dr.Nazife BAYKAL ______________________
Supervisor, **Information Systems, METU**

Examining Committee Members:

Assist.Prof.Dr.Yeşim AYDIN SON ______________________
Health Informatics, METU

Prof.Dr.Nazife BAYKAL ______________________
Information Systems, METU

Assist.Prof.Dr.Aybar Can ACAR ______________________
Health Informatics, METU

Assoc.Prof.Dr.Tolga CAN ______________________
Computer Engineering, METU

Assoc.Prof.Dr.Hasan OĞUL ______________________
Computer Engineering, Başkent University

Date: 21 June 2012

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Fahri Salih KOCABAŞ

Signature : ____________________

ABSTRACT

KNOWLEDGE DISCOVERY IN MICROARRAY DATA OF BIOINFORMATICS

Kocabaş, Fahri Salih
Ph.D., Department of Health Informatics
Supervisor: Prof.Dr.Nazife Baykal

June 2012, 153 pages

This thesis analyzes major microarray repositories and presents a metadata framework both to address the current issues and to promote the main operations such as knowledge discovery, sharing, integration, and exchange. The proposed framework is demonstrated in a case study on real data and can be used for other high throughput repositories in biomedical domain.

Not only the number of microarray experimentation increases, but also the size and complexity of the results rise in response to biomedical inquiries. And, experiment results are significant when examined in a batch and placed in a biological context. There have been standardization initiatives on content, object model, exchange format, and ontology. However, they have proprietary information space. There are backlogs and the data cannot be exchanged among the repositories. There is a need for a format and data management standard at present.

We introduced a metadata framework to include metadata card and semantic nets to make the experiment results visible, understandable and usable. They are encoded in standard syntax encoding schemes and represented in XML/RDF. They can be integrated with other metadata cards, semantic nets and can be queried. They can be exchanged and shared. We demonstrated the performance and potential benefits with a case study on a microarray repository.

This study does not replace any product on repositories. A metadata framework is required to manage such huge data. We state that the backlogs can be reduced, complex knowledge discovery queries and exchange of information can become possible with this metadata framework.

Keywords: Microarray Repository, Metadata Card, Semantic Net, Knowledge Discovery

ÖZ

BİYOENFORMATİK MİKRODİZİ VERİSİNDE BİLGİ KEŞFETME

Kocabaş, Fahri Salih
Doktora, Tıp Bilişimi Bölümü
Tez Yöneticisi: Prof. Dr. Nazife Baykal

Haziran 2012, 153 sayfa

Bu tez ana mikrodizi bilgi bankalarını analiz ediyor ve hem mevcut sorunları adresleyen hem de bilgi keşfi, paylaşma, bütünleştirme ve değişim gibi temel işlemlerin desteklendiği bir meta veri yapısını sunuyor. Önerilen yapı, gerçek veri üzerinde bir vaka çalışması olarak gösterilmektedir ve biyomedikal alandaki diğer yüksek kapasiteli bilgi bankalarında da kullanılabilir.

Mikrodizi deneylerinin sadece sayısı artmamakta, aynı zamanda boyutu ve karmaşıklığı da biyomedikal sorgulara cevap olarak yükselmektedir. Ve deney sonuçları, deneyler bir grup olarak incelendiğinde ve biyolojik bir bağlama yerleştirildiğinde önemlidir. İçerik, nesne modeli, değişim formatı, ve ontoloji üzerinde standardizasyon girişimleri olmuştur. Ancak, bunların özel bilgi alanları vardır. Birikmiş işler vardır ve bilgi bankaları arasında veri değişimi yapılamamaktadır. Şu anda bir format ve veri yönetimi standardına ihtiyaç vardır.

Deney sonuçlarını görünür, anlaşılabilir ve kullanılabilir yapmak üzere meta veri kartı ve anlamsal ağlar içeren bir meta veri yapısını sunduk. Onlar, standart sözdizimi kodlama düzenlerinde kodlanır ve XML/RDF'te temsil edilirler. Onlar, diğer meta veri kartları ve anlamsal ağlar ile bütünleştirilebilirler ve sorgulanabilirler. Onlar değiştirilebilir ve paylaşılabilir. Bir mikrodizi bilgi bankası üzerinde bir vaka çalışması ile performans ve potansiyel faydaları gösterdik.

Bu çalışma bilgi bankalarındaki herhangi bir ürünün yerini almamaktadır. Böyle büyük bir veriyi yönetmek için bir meta veri yapısı gerekmektedir. Bu meta veri yapısı ile birikmiş işlerin azaltılabileceğini, karmaşık bilgi keşfi sorgusu ve bilgi değişiminin mümkün olabileceğini ifade ediyoruz.

Anahtar kelimeler: Mikrodizi Bilgi Bankası, Meta Veri Kartı, Anlamsal Ağ, Bilgi Keşfi,

To My Family

ACKNOWLEDGMENTS

I express my appreciation to Prof.Dr. Nazife Baykal for her continuous support and reliance on my research from the very beginning. Assoc.Prof.Dr. Tolga Can has perfectly guided where I needed. The administrative assistance of Ali Kantar and Sibel Gülnar is acknowledged. I would like to thank my wife, Neslihan, for her endurance as well as for her subject matter expertise throughout this research. I thank my children, Alara and Almina, for understanding my unscheduled absences.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

There is huge and ever increasing, complex, variable and heterogeneous data produced by microarray and other high throughput experiments. The amount of data on microarray repositories becomes increasingly unmanageable unless proper standards are adopted as the number and content of submissions grow. Numerous groups around the world perform microarray experiments for several reasons by using differing methods, tools, and formats. The publishing media and repositories host these disparate results. The annotations and the metadata additions add much to the existing experimental content. However, these contextual data are unstructured and do not follow defined standards. The request for experiments is piling up while the cost of experimentation is decreasing comparatively. Not only microarray community but also whole biomedical domain has an interest over the interpretation of microarray results, be it biological, statistical, or computational. With this trend, the size, complexity, and heterogeneity have become an issue for the microarray domain. This mode of operation complicates the main activities such as search, knowledge discovery, data exchange not only in the microarray domain but in whole biomedical sciences. Due to the further processing requirements, the microarray experiment result data sets need to be annotated with supporting data and accompanying OMICS (*genomics [the quantitative study of genes, regulatory and non-coding sequences]*, *proteomics [protein expression]*, *transcriptomics [gene* *expression], metabolomics [metabolic networks]*) findings so as to put them in meaningful biological context. We state that the backlogs have been an overhead on repositories, and the experiment results are lacking adequate visibility, understandability, and usability. Even reasonable exchange between the repositories is not yet possible. The major issue is now to share the findings and derive new knowledge from it. The ‘exchange’ is a key operation before ‘understand’ and ‘use’. The pressing need for verification, re-use, meta-analysis, and further interpretation of microarray results for knowledge discovery requires organization, storing, and processing to pave the way for machine understanding.

Scientists including academics and industry will continue to create their own local standards or formats which will worsen the problem of sharing and exchanging biological information. Since the data centers are not interoperable and the results are in different formats, limited collaborative and meta studies can be done and most of the published and unpublished data cannot be used in knowledge discovery efforts. There have been fruitful standardization initiatives which have gained wide acceptance on different areas of the microarray domain as listed subsequently. For example, MIAME (Minimum Information about a Microarray Experiment) has been a widely accepted content standard for microarray experiments especially following the conformance declaration by principal microarray repositories and some notable publications (1). MIAME has been developed into MINSEQE (MINimum Information about a high-throughput SEQuencing Experiment) to address new generation sequencing experiments (2). Some others, as detailed in the Functional Genomics Data Society website (http://www.mged.org/), are minimum dataset checklist, MIBBI (Minimum Information for Biological and Biomedical Investigations) (3); object model, MAGE OM (Microarray Gene Expression Object Model) (4) and FuGE (Functional Genomics Experiment) Data Model (5); exchange platform, MAGE-ML (Microarray Gene Expression Mark-up Language) (6); application with a standard format, MAGE-TAB (Microarray Gene Expression Tabular) (7) and ISA-TAB (Investigation Study Assay Tabular) (8); ontology, MGED Ontology (9), FuGO (Functional Genomics Investigation Ontology) (10), OBO (Open Biomedical Ontologies) Foundry (11), OBI (Ontology for Biomedical

Investigations) (12), and GO (Gene Ontology) (13). These initiatives and their developments have been presented in review articles to reach wider audiences (14-18). The adoption of common standards is emphasized for the management and sharing of microarray data by (19-22). The power of the linked data in RDF (*knowledge representation and semantic web language*) format has been presented by Bizer et al (23). Brors (24) and Aalai et al (25) have studied the functional annotation mechanisms, related databases and linking literature abstracts. Microarray annotation by ontology and ontology based analysis paradigm has been explained in detail by Brors (24), Schober et al (26), Chen et al (27), and Pasquier et al (28). There are 3 primary microarray repositories, which are NCBI GEO (National Center for Biotechnology Information Gene Expression Omnibus) (http://www.ncbi.nlm.nih.gov/geo/) (29a-h), EBI (European Bioinformatics Institute) ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) (30a-b), and CIBEX (Center for Information Biology Gene Expression Database) (http://cibex.nig.ac.jp/) (31a-b). They not only host the experimental data but also present tools for querying and analysis. Although there are standards for content, representation, storage, and exchange, GEO and other microarray repositories have their own information space and use mainly proprietary formats. There are also public-domain applications developed on the BioConductor platform (32) such as GEOmetadb (33), GEOquery (34) to extend the functionality of the GEO repository. There are several software platforms developed to implement MAGE OM and support MAGE-ML such as MeltDB (35), SAMS (36), EMMA (37), and MARS (38). However, the exchange model and mechanism among these applications is difficult to implement with less bioinformatics support. Therefore, exchange and common understanding of data among disparate repositories continues to be an issue. There is mediating software between microarray repositories and systems biology/pathway databases (39). There is also public microarray analysis software developed on the BioConductor platform (40, 41) as well as commercial ones like NextBio (http://www.nextbio.com/) (42). They typically suffer from above-mentioned issues, which are common for such big projects. The lack of a common data format in microarray domain, as stated by Larsson and Sandberg (43), is still valid. Practical solutions such as MINiML (MIAME Notation in Markup Language) (http://www.ncbi.nlm.nih.gov/projects/geo/info/MINiML.html) (29f) and MAGE-TAB (Microarray Gene Expression Tabular) (7) have been proposed. Nevertheless, they are lacking standard syntax and semantics. The solution is standard related and can be provided with data management and using architectural frameworks. We believe that data level standardization efforts such as employing data management, structuring data, and extending the semantic power of the data will play a very important role in addressing the above mentioned problems in biological community. Third-party platforms including other high throughput repositories will benefit from such solutions besides microarray repositories.

NCBI GEO repository has been selected for this study. The GEO is a curated, online resource for gene expression data. The GEO is not only one of the main submission areas for the microarray experimenters but also a primary information space for other biomedical domains in other studies. There are three records (*Platform, Sample, and Series*) as supplied by submitters on the GEO Database schema. There are currently 30 thousand Series records (GSExxx) on GEO. A GEO Series record summarizes an experiment. It links a group of related samples and provides a focal point. The GEO staff (*curator*) reassembles this data into the GEO Datasets (GDSxxx) to support further investigation. A Dataset represents a collection of comparable samples processed using the same platform. GDS differs from GSE where it excludes some parts and is reassembled, reformatted (29g). Considering the standardization issues, current overhead and backlog, we decided to deal with these challenges in this dissertation. We have examined the Series records on the GEO in this study, because the GEO Series record is the description of the overall experiment and it is also presented in a specific XML format, MINiML. The MINiML is a file which includes both data (*such as summary, platform, and sample data*) and metadata (*such as title, description, and contact information*) for Platform, Sample, and Series records and is defined and validated by a schema file, MINiML.xsd. The MINiML file is said to be a metadata card, but it is not named and designed as such. Although, we analyze the GEO in this thesis, we state that other microarray and high throughput repositories have similar structures and challenges. The framework presented in this study can be used in these repositories.

1.2 Current Issues

Some of the current issues over the GEO records are listed subsequently to provide both the rationale and driving factor behind this study.

The MIAME is a content standard, not a format standard. Because the MIAME lists the minimum content with no implementation details and format guidance. The data is mostly in proprietary formats on microarray repositories. MIAME offers several checklists; however, there is not yet a common reporting structure, although there are studies such as MICheck and MIBBI (44). The MAGE OM is a reference object model. However, its implementations differ greatly and are not easy to implement for laboratories without adequate bioinformatics support (6). Therefore, practical products such as MINiML, ISA-TAB, and MAGE-TAB have been developed. However, they have not been standardized yet and been proprietary implementations for their user communities. A study to propose a common format for adoption in this field is needed.

The microarray repositories are not connected and they cannot regularly exchange experimental data. Thus, the records that are on different repositories are not visible as a whole. The type, content, format, and availability of data and metadata on different repositories are at varying degrees and not comparable. Therefore, the regular exchange of data as it occurs among DNA repositories does not happen. There is an initiative by the ArrayExpress staff to import the GEO records with Affymetrix and Agilent platforms (*approximately 10% of the GEO records*), on a weekly basis following a manual re-curation (http://www.ebi.ac.uk/ microarray-as/ae/). However, they are not synchronized and if the records in GEO are updated, this will not be automatically reflected in the corresponding ArrayExpress entry.

We argue that the metadata for microarray records should be visible, usable, and understandable by both humans and machines. The microarray records are lacking these data tenets because they are not structured accordingly. For example, available analysis tools on the GEO (*e.g. GEO Profiles*) look into only the GEO Datasets (*not all GEO Series records*) and cannot search other microarray repositories. Moreover, some records are ambiguous and incomplete violating the integrity of microarray repository. Therefore, the scientists find it difficult to review the accomplished studies to plan their experiments. And, meta-analyst cannot reach and analyze comparable data from different repositories. The details for these data characteristics are given in Table 1.

Table 1 - The tenets of data strategy

Characteristics	**With**	**How**
Visible	*Resource metadata such as title, description, submission date, contact, security classification*	*Creating discovery metadata for precision search*
Usable	*Structural metadata as defined in the schema*	*Making data available in shared spaces*
Understandable	*Semantic metadata through the use of CV, ontology, and semantic nets*	*Publishing associated semantic and structural metadata in a metadata registry by consolidating the meaning and context to support semantic interoperability and knowledge discovery*

The current issues with these characteristics can be stated as the following, Table 2.

Table 2 - The current issues and associated data tenets

Characteristics	**Issues**
Visibility	not available and not discoverable
Usability	redundant data, sometimes different entries for the same record violating the integrity, differing formats, not machine-processable
Understandability	ambiguous, incomplete so that intended meaning cannot be conveyed

The quality and category of a GSE record is not clearly labelled at the submission and throughout its lifetime. The quality metric (*values such as "verified" and "SCI>10"*) and state (*values such as "incomplete" or "retired"*) information can add essential meaning to the records. For example, some experiments are published on a publication with high citation, done by a respectable group, verified with RT-PCR (Real-time Polymerase Chain Reaction) and repeated with success. On the other hand, some records may be identified as poor studies since they are contradicted by the experiments with high quality. There are also comparability issues between different platforms as pointed out by the MAQC (MicroArray Quality Control) project (45).

The metadata about the records are not structured in accordance with the DC (Dublin Core) metadata standard (http://dublincore.org/). There are entry anomalies, inconsistent terminology and even incorrect entries within metadata e.g. on contact information (*names, organizations, country names, Date Time Group*) or in the summary part. This can be handled with structured data entry that is based on standard code list, controlled vocabulary, and ontology. Mandating patterns could also be included in a relevant schema file as tested in OpenSDE projects (46). The experimenters should select from lists (*Codes, Controlled Vocabulary, and Ontology*) in their submissions and make structured entry rather than free text. The experimenter could enter more of the experimental findings including metadata on contributors, experiment settings, biomaterials, data analyses, and especially on the results section if there were a structured format.

Below, we give some examples about flawed and ambiguous entries on the GEO records.

- There are inconsistent, incomplete and even incorrect entries for the same information element. For example, there are seven different spellings (*United States of America, United States, USA, US, U.S., U.S.A., U.S.A*) in address data of contact details for the country name ‘USA’, although there are ISO country names and codes available as a CV (*60% of the contacts are from USA*). There are city names in the country field. There are different patterns for the names of the same person, organization, and date.
- There are 3 different versions of MINiML files for the same Series record which contain different content (*1: MINiML equivalent of HTML Series record, 2: MINiML_family link within the HTML Series record, 3: programmatically extracted Series data for the whole database*). For example, one of the contributors is missing in Series Record GSE362 at 1:. The Summary, Pubmed ID, and Overall Design information fields are not available in programmatic access in the same Series Record at 3:. All files included in one GPL record are non-platform files whereas GSE record (GSE473) contains sample and platform data for GPL96.
- The related experiments (*super and sub series records*) are not visible. A super Series record includes individually submitted subset records, which all belong to one experiment, and provides the base for GDS record creation. Some Series records about an experiment are

submitted separately over time without explicitly displaying if they are related or not, although there is an entry in parent Series record for a list of child records. Therefore, it is difficult to trace Series records for an experiment. For example, Vijay G. Sankaran submitted three Series records on the 5th of December 2008 which seem to be related but was not obvious whether they were part of a single experiment or they were related to different experiments. However, once we examine the details of these Series records, we see that although they are separate individual submissions, they are connected to a single experiment and Series GSE13285 is a super Series, which includes subset Series GSE13283 and GSE13284. Their MINiML files are included in each record; however, these three MINiML files need to be considered together to get the overall results from this experiment. In essence, the GEO curators transform these related Series Records into a GEO Dataset record, but this association is not clearly visible to the researcher. Another example for this case is by Alexander V. Dmitriev on his multiple submissions.

- MIAME guideline states that the summary part of the microarray experiment record and the abstract part of its related publication should be the same. Some records (*for example, GSE3570, GSE5483, GSE9628, GSE12459, GSE15280, and GSE15808*) have different summary information than the abstract portions of their related publications. The publication/citation information is missing in some records where we found out that the related publications have actually been published; but, their related GEO Series records have not been updated (*for example, GSE5546 was submitted in 2006 and has no citation information but its related publication was published in 2008 – PMID18271932*). It is the same case for GSE7135 and GSE 18770 for which their related publications (*PMID 18092324 and PMID 19965665 respectively)* have been published a couple of months later; but, the corresponding GSE records have not been updated. In some records, GDS, GSE, and related publications have different summary/result information (*for example, GDS848 (GSE763, PMID15205334), GDS2755 (GSE7754, PMID 17540599*), *GDS2861* (*GSE8562, PMID 17660348), GDS2785 (GSE7463, PMID 17505532)*). In some records, there are no summary information where there are entries in their GDS counterparts or associated publications (*for example, GSE7621* (*GDS2821, PMID17571925), GSE17101 (PMID19808959), GSE16779 (PMID19956606)*).

There are cases such that web site of the submitter's lab, subsequent entries to related publication(s), and other specific databases such as GO, public biological pathway databases (*for example, Patika (Pathway Analysis Tools for Integration and Knowledge Acquisition, www.patika.org/) (47), Reactome (48)*) have inconsistent contents for the summary/result data. The links among these repositories are not clear or not established. The microarray records, related publications, and relevant data fed into databases (*such as gene and biological pathway*) should be consistent. The microarray repository should be the primary reference for other platforms and there has to be a link between these platforms throughout the lifetime of the relevant record(s) in response to modifications. The semantics is not addressed in the design of microarray repositories. Thus, understandability and usability is weak. Lifecycle management to include version and change management is not in place, either. There are also ambiguous, different, or incomplete content within different parts of the text of the experiment results among similar experiments. This is a data integrity issue which can be resolved by employing data and configuration management.

More automation is required for addressing slow curation work and increasing backlogs. For example, the GEO is experiencing a significant backlog in Dataset (GDS) creation and most of the Series records (80%) do not have a corresponding Dataset. At present, there are about 2720 GDS records and 30497 Series records. The number of GDS records remains unchanged for almost a year. There are usually 2 subset records in a superset. One GDS record may contain more than one GSE record and equally one GSE record may be the source for multiple GDS records (*two GSE in one GDS on average. There are 600 super Series records and 45 GDS records include super Series records*). Therefore, we can label the number of GDS records as visible and GSE records as not yet visible microarray experiments. The current ratio between these two categories (*GDS records and GSE records, which are not yet transformed to GDS*) is about 1 to 5 which amounts to 80% backlog. Also, approximately 15% of submitted Series records have not yet been published due to ongoing curation work. As the content of the records is standardized with the use of data and syntax encoding schemes, the increase in machine involvement will facilitate and expedite the curation work. Inference and knowledge discovery operations are also promoted as experimental results are placed in a biological context within a machine processable structure and format.

1.3 The Approach to Overcome the Listed Issues

Biomedical studies are multidisciplinary. The biomedical projects are usually tightly scheduled and commercial or custom-built software is produced. Mostly, such software is proprietary and biomedical data is not organized, structured, and represented in accordance with information technology principles. Such projects indeed require coordinated activities of related domain experts.

Microarray repositories have records for submitted experiment results. Some data are raw data such as numerical values or scanned images. Some data are metadata such as title, description, summary, contact or publication. However, the metadata organization and representation is not in line with the established standards. Thus, the access, analysis, search, exchange of this metadata has long been an issue.

A standard for format in representation, exchange, and reporting is currently missing. Not only these knowledge-based operations are inefficient but also there are serious backlogs publishing the submitted data.

In this thesis, we propose, the concept of metadata card and semantic nets to address these issues within a metadata framework. In this framework, we demonstrate structuring and encoding the content to deal with syntactical issues, and representing the domain knowledge to handle semantic issues. We also bring and discuss the concept of configuration management (*for example lifetime change management of each record*), structured data entry and metadata registry. We propose the framework presented in this study as a format and data exchange standard for inclusion into microarray standards. We state that such a resource, when compared to its return, is worth spending. We emphasize that such a work is composed of a series of data management operations.

1.4 Related Standardization Initiatives

Overview

The goal of this study is to propose a framework to support knowledge discovery by increasing the visibility, usability, and understandability of the experiment records on microarray repositories. The proposed framework MAdmf, includes metadata card and associated semantic nets about microarray record(s). The experiment itself is out of the scope of this study. This dissertation covers the lifetime of the microarray experiment result data sets. The problem area which is addressed by our framework encompasses the publication of the results in a journal (*or in a specialized repository*) which are not formatted in machine readable manner. There has been a research direction on describing and exchanging data within microarray domain. Data sharing has been successful among sequence databases. A similar infrastructure for microarray experiments will be more beneficial, because the data are meaningful only in a biological context. Potential data standards are visited for a possible use. It is because there are established standards and techniques on data management, configuration management, metadata management, knowledge representation, and structured data etc. (18). And, there is proliferation of community driven data standardization activities (17). Many biologists hardly know about architectural and management standards that exist in information technology. Equally, there are few IT people who have expertise in biology. Because there are few experts on this area (*bioinformatician or microarray informatics expert who is specialized in both biology and information technologies*), either IT people or biology people lead such studies. Commercial companies within short termed projects do not invest in standards as required, as well. The storage, representation, and management of data should take precedence to implement efficient discovery, exchange, and integration operations on this vast and heterogeneous information space. The important step is the adoption of existing or the development of new cross-domain standards.

A tremendous amount of data, which is largely disconnected, has been outputted from high throughput technologies like microarray experiments to examine gene expression profiles. The ability to extract information from this globally produced data falls short because the amount of data overtakes the efforts to store (49). Spellman (50) indicates that if microarray data continues to be in different format with their own meanings, the advantage of public data sharing would be lost. The solution is standardization in all major activities of the business process namely, representing, storing, exchanging, and interpreting that are related to the syntax and semantics of data. The degree of structuredness can be varied from some enhancements to an optimum format such as structured message exchange in (North Atlantic Treaty Organization) environment or Clinical Document Architecture (CDA) based reporting in HL7 (Health Level Seven, http://www.hl7.org/). The applications usually are limited for both the producer in expressing the totality of his/her findings and the consumer in extracting quality information out of it for complex queries. The bottom line is the data management and creating a separate data layer for domain information so that domain experts can work on it. The benefits of such a work will be propagated to the subsequent phases such as modelling, application development, and ontology development. The importance of standardization has been recognized by the microarray community. Therefore, similar and related groups are merging, the standards, repositories, publications are freely accessible. There are mail groups, forums, workshops, trainings, conferences to share and improve the performance and adoption of the existing standards initiatives. There is enough momentum in this area to move forward. We believe that standardization studies such as ours is needed in microarray domain.

Brazma (49) emphasizes that the result of an experiment can be combined from many independent ones leading to more knowledge than each experiment can yield individually. However, such an operation is dependent on the use of standards. The need for standardization becomes obvious when we try to integrate annotations of the functional roles of the genes. To obtain new insights and knowledge, the huge amount of information generated by high-throughput experiments needs to be transformed into meaningful executive summaries. That is the metadata card in our framework. Testing the hypotheses based on these summaries is a main task for computational biology. Biological interpretation of the experiment results will not be possible without adopting standards. That is why; we propose RDF/XML based metadata card and semantic nets which may take different representations with the use of established syntax encoding schemes.

Standardizing the Standards

Although there are many standards, most of them are incompatible and the required standards are not used (*not preferred or not known*) by biomedical community. Quackenbush (51) titled his relevant article as "standardizing the standards" to discuss this issue. He states that two conflicting but correct descriptions of the same experiment can be encoded in MAGE-ML due to the flexibility of XML. To address this issue, Wang et al (52) suggest the use of the semantic web, especially RDF as data representation in the design of OMICS standards. They indicate the value of the syntax and semantics of RDF. However, they use neither metadata standards, metadata card concept, nor semantic nets.

Biological Data Management Challenges (http://crd.lbl.gov/html/BDMTC/overview.html)

Biological data management involves the traditional areas of data generation and acquisition, data modelling, data integration, and data analysis.

There is no standard way of moving results between research groups yet. This applies to both the underlying gene-expression data and the descriptive biological annotations that provide context for the gene-expression measurements. Numerous papers have been published, but a few of them have presented data in the same format, and almost none has provided adequate contextual information to allow verification and reproduction of experiments and further use of the results. Because the microarray data is meaningful if combined with other systems biology data, the microarray data needs to be shared by other communities. No one should expect that there will be one common standard for the format and exchange of the data. We need approaches to work with existing practices. It is not wise to come up with a new standard and expect all parties to adhere it. It does not work that way. Therefore we took references from successful implementations based on established standards. This is structured data concept with data management focus.

Relevant Organizations and Standardization Initiatives

Microarray Gene Expression (MGED) Society was founded in 1999. Its name was changed to Functional Genomics Data Society in 2010 to reflect its current mission which embraces functional genomics. This society works for sharing functional genomics, proteomics, and microarray data, (http://www.mged.org). It has established the data standards (*data quality, data management, annotation, exchange*) in that regard. The MGED standards are shown in Table 3.

MIAME is a content standard to declare the minimum information about a microarray experiment to support its repeat and correct interpretation. MAGE-OM is an object model describing microarray experiments. It is a database schema which presents a reference model to facilitate the exchange of microarray information. MAGE-ML is a markup language used to describe microarray experiments. It is an exchange format, which is based on MAGE-OM. MAGE-ML is an XML representation of the MAGE-OM. MAGE-TAB can be used for annotating and communicating microarray data in a MIAME compliant fashion. It is expected to enable laboratories without bioinformatics support to manage, exchange, and submit annotated microarray data in a standard format using a spreadsheet instead of

MAGE-ML. MAGE software toolkit (MAGEstk) is an open-source software toolkit that helps to construct and use MAGE files.

Table 3 - MGED Standards

MIAME (WHAT to store)	**M**inimum **I**nformation **A**bout a **M**icroarray **E**xperiment specifies what information is necessary to understand a microarray experiment.		
MAGE (**M**icroarray **a**nd **G**ene **E**xpression)	**MAGE-OM** (**MAGE O**bject **M**odel) (HOW to store)	**MAGE-ML** (**MAGE M**arkup **L**anguage) (HOW to communicate)	**MAGEstk** (**MAGE S**upporting **T**ool **K**it) (HOW to develop)
MO (**M**GED **O**ntology)	MGED Ontology provides standardized descriptions of biological and experimental properties to better facilitate interpretation and comparison.		

The goals of MGED efforts can be listed as,

- Determine information needed for experiments (*What to store: MIAME and Minimum Information Checklists*)
- Provide a means to share this information (*How to store: MAGE-OM*)
- Provide a microarray exchange data format (*How to communicate: MAGE-ML*)
- Provide a common language for experiments (*How to annotate: MGED Ontology*)

MGED's solutions to achieve above goals are the introduction of scientific guidelines (MIAME), data communications standards (MAGE) and biological annotation systems (MO). As costs go down and experimental systems become more automated, microarrays have increasingly been used in biological research. ArrayExpress and GEO tend to double in size every year. The interpretation of microarray results depends on context and experimental variables. They have a profound impact on what information can be extracted from the data (50).

Within the scope of this study, MGED standardization initiatives fall into one of the following categories: Capturing (MIAME); Representing and Storing (*MAGE-OM, MO - Structured and expressive tools*); Exchanging (*MAGE-OM, MAGE-ML - The latest technology*). The basic approach is two pronged: One being ‘more structuredness', the other being 'high expressiveness'. The platform should be XML based which hosts specifications and technologies to structure, to represent and to exchange data.

MIAME

MIAME is a list of information objects that researchers should share to describe their experiments. It includes information about experimental design, biological samples, and features on the microarrays, experimental protocols, data acquisition and processing. It identifies the minimum information necessary to understand and repeat the experiment, verify the results (1). The MIAME checklist (http://www.mged.org/Workgroups/MIAME/miame_checklist.html) lists the information elements that should exist and many journals require it for the submission of microarray experiment results. The MIAME framework standardizes six areas of information:

- experimental design
- microarray design
- biological samples
- hybridization procedures
- image analysis
- data analysis

The MIAME describes content without a format; there may be numerous different formats to support MIAME requirements. The lack of format and exchange standardization in microarray experiment process indicates an issue for interoperability. Several projects use different tools and methods to produce, store, and exchange microarray data. However, their data are heterogeneous and not structured. The MIAME provides basis for an exchange format. However, the format is considered to be model based or application based so far. Therefore, the data content standard is developed without a format. We believe that the format proposed in MAdmf may serve many purposes such as data exchange, integration, and knowledge discovery. Due to the characteristics of representation scheme (RDF/XML), the syntax and semantics are provided by default.

MIBBI (Minimum Information for Biological and Biomedical Investigations, http://mibbi.org/)

Minimum information checklists are based on MIAME guidelines. The MIBBI offers a portal to proliferating Minimum Information Checklists (3). This has been a standardization initiative on content. There are more than 30 checklists in this project. These prescriptive checklists specify the key information when reporting experimental results. Such checklists enable the submitter to ensure that the methods, data, analyses and results are described to a level sufficient to support

- unambiguous interpretation
- experimental corroboration
- reuse
- the extraction of maximum value

However, the checklists conflict in scope, structure, and conventions. Therefore, they cannot be used in combination. This is important in complex studies such as metabolic pathways, and systems biology which combine information from multiple biological domains. As a solution, MIBBI serves as a one-stop-shop for those exploring the checklist projects.

MAGE-OM AND MAGE-ML

The MAGE-OM is the establishment of a data exchange format and object model for microarray experiments. The MAGE-ML is an exchange format in XML which is generated from this model. MAGE-ML packages correspond to MIAME sections (6). MAGE-ML files can be submitted to ArrayExpress (53). Stanford Microarray Database (SMD) makes MIAME-compliant data submission to ArrayExpress on MAGE-ML format. The ArrayExpress repository complies MAGE-OM model.

The difficulty with MAGE-ML has been the complexity of the MAGE-OM. Another difficulty is the possibility of encoding the same information in different ways (18). Although the authors (18) list the overhead of using XML representation as additional content, we believe this is not valid anymore due to fact that XML is the representation for exchange with many advantages. And, there are efficient XML studies which aim at using binary format (www.w3.org/XML/EXI/).

A spreadsheet-based format, MAGE-TAB and XSD (XML Schema Definition) based format, MINiML have been proposed in place of MAGE-ML to describe and exchange microarray data. We state that MINiML and MAGE-TAB are alike and they both suffer from similar deficiencies. They may continue to be used for reporting purposes. However, the metadata and semantics should be represented within metadata card and SemNet as presented in our framework, MAdmf.

MAGE-TAB

Sharing of microarray data within the research community has been facilitated by the development of the content and communication standards, MIAME and MAGE-ML. However, the use of MAGE-ML format has been impractical for laboratories lacking dedicated bioinformatics support. MAGE-TAB, as simple tab-delimited format, can enable laboratories with insufficient resources to manage, to exchange and to submit well-annotated microarray data in a standard format using a spreadsheet (18). However, since its inception, it has not been recognized by majority of microarray community and has remained as a proprietary format. Because it is useful if all parties agree to use it. The microarray community is in need for solutions which are based on mature, already established standards.

MINiML (MIAME Notation in Markup Language, http://www.ncbi.nlm.nih.gov/projects/geo/info/MINiML.html)

The MINiML is a data exchange format for microarray gene expression data, as well as many other types of high-throughput data. It uses three basic records: Platform (*e.g., array*), Sample (*e.g., hybridization*), and Series (*experiment description*). MINiML follows the MIAME checklist and uses XSD as syntax. GEO supports data submissions and retrievals in MINiML format.

MAGE-ML is the standard. However due to its complexity, MAGE-TAB and MINiML have been the substitutes. Thus, MINiML is an alternative to MAGE-ML. We believe that it is a metadata card and it should conform to metadata standards. It should also be amended with knowledge representation to improve its semantics. This is the essence of our study. So that microarray records can become visible, usable and understandable with the use of machine processable metadata cards and semantic nets. XML schema based definition of data in biomedical domain is not new nor in the realm of MINiML. There are studies on XML schemas on biomedical domain such as Seibel et al (54). They developed XML schema to include common data types in a Java library to enable the applications to interoperate seamlessly. Note that we propose RDF/XML syntax, which is SemWeb knowledge representation language, to encode metadata.

MGED Ontology (MO, http://www.mged.org)

The MGED ontology is a framework for describing functional genomics experiments. The primary purpose of the MO is to provide standard terms for the annotation of microarray experiments (55). The MO includes terms, definitions and concepts. The terms are organized into classes with properties and relationships are defined.

The ontologies are structured as directed acyclic graphs, which are similar to hierarchies but differ in that a child term can have many parents. For example, the biological process term “hexose biosynthesis” has two parents, “hexose metabolism” and “monosaccharide biosynthesis”. The MO includes terms and annotation rules for microarray experiments. Thus, several operations such as structured queries and data exchange can be done without loss of meaning. Ontology concepts are derived from the MIAME guidelines and MAGE-OM. The MO uses structured fields and controlled terms. For example, MO is used in generating MAGE-ML files to describe a microarray experiment (55). MO currently comes in html and OWL (Web Ontology Language) format and contains 228 classes (*BioAssayData, ExperimentDesignType, BiologicalProperty*), 110 properties (*has_accession, has_bioassay_data, has_initial_timepoint*) and 658 instances (*image_acquisition_software, disease_state, control_biosequence*). There are about 40,000 terms in it. The use of these terms enables both unambiguous description of how the experiment was performed and structured queries on elements of the microarray experiments.

Microarray Quality Control Project (MAQC, http://www.fda.gov/ScienceResearch/BioinformaticsTools/MicroarrayQualityControlProject/**)**

This is a project of the US Food and Drug Administration (FDA) to develop quality metrics which will enable the use of microarray data in medical science. The purpose of the MAQC project is to provide quality control tools (*process, software, hardware, principles*) to the microarray community in order to avoid procedural failures and to develop guidelines for microarray data analysis. Thus, MAQC addresses the concerns about the reliability of this technology regarding the publication of studies with contradictory results, obtained using different microarray platforms to analyze identical RNA samples, as well as other performance and analysis issues. The MAQC study provides a resource to build consensus on the use of microarrays in research, clinical and regulatory settings (45).

As Eggle, D. et al stated in their submission to GEO (GSE10929), the MAQC project showed that microarrays with comparable content enable inter and intra-platform reproducibility. They demonstrate that not only experiment design but also database content and annotation strongly influence comparability and performance of subsequent generations of microarrays. The MAQC initiative now has become another major customer for microarray data.

OBO (The Open Biological and Biomedical Ontologies) **Foundry (**http://www.obofoundry.org/**)**

In general, we can say that OBO is a collection of controlled vocabularies. Such a controlled vocabulary may take the form of ontology as it is refined and integrated. OBO intends to present a core of participating ontologies to support interoperability by a common design philosophy. By providing controlled vocabularies, OBO aims to extend GO (Gene Ontology) development principles to biological domains. There are currently over 40 ontologies in OBO, covering domains such as anatomy, development, genomic and proteomic information, and taxonomic classification (11).

GO (*Gene Ontology for biological function*), OBI (*Ontology for Biomedical Investigations for experiments*), SBO (*Systems Biology for biochemistry*), MO

(*MGED Ontology for microarray experiments*), and PW (*Pathway Ontology for biological process*) are among participating ontologies.

BFO (Basic Formal Ontology, http://www.ifomis.org/bfo**)**

BFO is an upper ontology which can be used in support of domain ontologies developed for scientific research within the framework of the OBO Foundry. BFO does not contain physical, chemical, biological or other terms which would fall within the special domains. And, OBI utilizes BFO as upper ontology in its design.

OBI (Ontology for Biomedical Investigations, http://obi-ontology.org/**)**

OBI models the design of an experiment with its protocols, the material used, the data generated and the type of analysis performed on it. MGED is also participating to OBI Consortium. The OBI is developing an integrated ontology for the description of biological and clinical investigations.

BioPortal (NCBO Portal, http://bioportal.bioontology.org/**)**

Bioportal provides access to the ontologies that are actively used in biomedical communities. One can annotate his/her text in its annotator tool, NCBO Annotator (56). It is a web service that converts free text (*e.g., a journal abstract of 200 to 300 words*) into a set of recognized and related ontology concepts and terms. Since 2002, the call for papers for the International Semantic Web Conference has contained the following sentence: "Authors of accepted papers will be required to provide semantic annotations for the abstract of their submission for the Semantic Web". Today, the content is still mainly composed of unstructured text that is not re-usable by software agents or semantic engines.

GO (Gene Ontology, http://www.geneontology.org**)**

GO provides structured CVs for the annotation of gene products with respect to their molecular function, cellular component, and biological role. The Gene Ontology project develops and uses a set of structured, controlled vocabularies for community use in annotating genes, gene products and sequences. The Gene Ontology is one of the ontologies held in the OBO.

FuGO (Functional Genomics Investigation Ontology, http://fugo.sourceforge.net/ontologyInfo/ontology.php**)**

Functional Genomics Investigation Ontology is a collaborative, international effort that will provide a resource for annotating functional genomics investigations. FuGO will contain both terms that are universal to all functional genomics investigations and those that are domain specific. In this way, the ontology will serve as the "semantic glue" to provide a common understanding of data from across these disparate data sources.

Biological ontologies aim to overcome the semantic heterogeneity and they are both machine and human readable. Ontology terms provide unambiguous descriptions and enable structured queries. However, proliferation of ontologies has become an issue because they do not follow reference structures, common design architecture and principles. Note that terms are not mere words but come from a hierarchy of class structures with properties. Within SDE (structured data entry) paradigm, the use of CV and data encoding patterns (*address, DTG, contact information etc*) support semantic annotation besides the use of ontologies. Note that we also propose syntax encoding schemes to complement structured data entry in our framework. In modelling the domain semantics, we state that the use of ontology should be supported with the use of metadata cards and semantic nets. In that regard, the efficiency of RDF and OWL could be extended with the use of rules. The results of the experiments, in form of facts, rules, hypotheses, could be encoded into structured statements such as RuleML Datalog, as used in our metadata framework.

BioConductor, (http://www.bioconductor.org**)**

The Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. It uses the R statistical programming language. The Bioconductor project produces an open source software framework that will assist scientists especially working on microarray.

Terminology Discussion, (http://vanrees.org/research/papers/2003_cib.pdf**)**

The terms classification, taxonomy, and ontology are explained comparatively to ensure common understanding in Table 4.

Table 4 – The comparative explanations of classification, taxonomy, and ontology

Classification	A grouping of entities according to some external criteria. Classification is a set of boxes with labels to sort things. It can be used as a user-friendly view in a taxonomy or ontology
Taxonomy	A hierarchical grouping of entities according to internal data. When used as a simple ontology, the taxonomy's hierarchy should be based upon a subclass hierarchy
Ontology	A set of well-defined concepts describing a specific domain. The concepts are defined using a subclass hierarchy, by assigning and defining properties and by defining relationships between the concepts. Ontology’s goal is to provide a common set of concepts for use. Therefore, it is common to use multiple ontologies each providing concepts for a particular domain to form a rich vocabulary

To gain a clearer understanding, the terms are offset against each other.

Differences between taxonomy and classification: Taxonomy classifies according to properties internal to the data, a classification can be made according to external criteria. For example, the decision to place yoghurt in the category dairy products is based upon data inherent to the entities, so this would be a piece of taxonomy. An external reason could be for instance classification of building components according to the branches of the building industry. This would lead to a classification, not taxonomy. They provide a list but a classification basically stops at that point. Taxonomy includes more descriptions. A classification tells you in which box your piece of data is, taxonomy tells you what your data is.

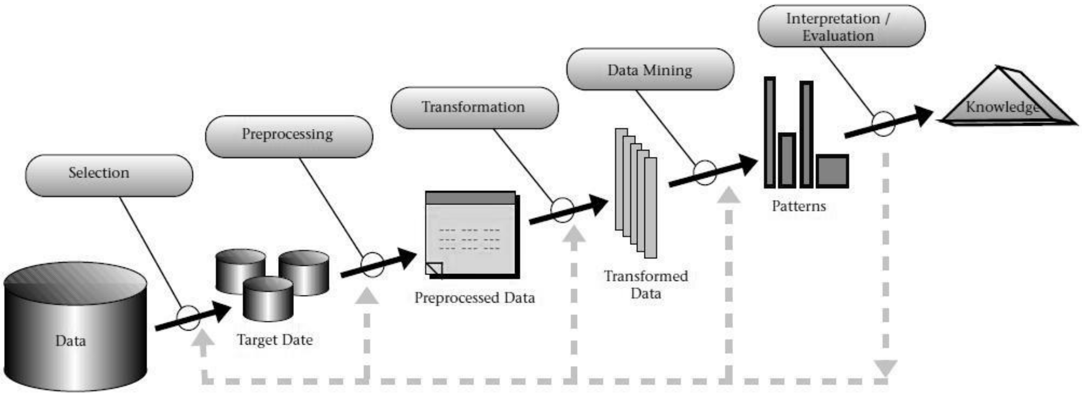
Differences between taxonomy and ontology: Ontology provides you with a lot of information about the concepts, including their relationships. There are rich properties and relationships in ontology whereas taxonomy is a sub-class hierarchy. However, ontology often contains a subclass-based taxonomic hierarchy. And, people uses taxonomy interchangeably with simple ontology. The difference between taxonomy and ontology is in the richness of information available. Once a lot of properties and relationships are added to a hierarchical structure, the term “ontology” is better suited than “taxonomy”.

1.5 Knowledge Discovery and Microarray Informatics

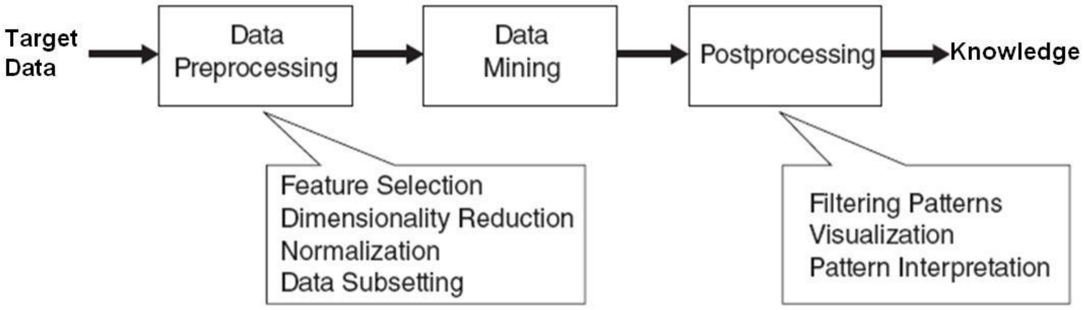
Knowledge Discovery Basics

Knowledge discovery (KD) is an interdisciplinary (*machine learning, database management, pattern recognition, statistics, artificial intelligence, expert systems etc.*) process of searching large volumes of high-dimensional data for patterns of domain knowledge. It is the process of deriving (*automated extraction*) useful information (*knowledge*) from such huge data. It is a continuous process such that the knowledge generated out of this process can also be used for further discovery. The data is exponentially growing and becomes more complex in both structure and semantics. The knowledge discovery from this huge amount of heterogeneous complex data requires preprocessing such as introduction of new types of data as well as new representation schemes.

The KD is defined with the stages of selection, preprocessing, transformation, data mining, and interpretation as depicted in Figure 1.



1.a: An overview of the steps that compose the KD Process (57)



b: An abstraction of KD process (58)

Figure 1 - Two layouts for the process of knowledge discovery

The data mining is the analysis step of KD process. The terms data extraction and data harvesting are also used in place of data mining. KD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. We believe that sole application of data mining out of the concept of KD may yield incomplete or incorrect results. The knowledge discovery technology is in its infancy, as it is still far from adequate for handling the large-scale and complex data (59).

Frawley et al (60) stated in 1992 that “computers have promised us a fountain of wisdom but delivered a flood of data”. This is valid today for all domains including microarray data. And, there is a growing gap between data generation and data understanding. Data is collected and stored at enormous speeds. But, it is not appropriately represented in terms of its syntax and semantics so that knowledge discovery operations are not as successful as expected.

KD is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It is mainly related to better use of domain knowledge. The methods for KD have been improved but the handling of target data is often neglected. The data preparation including the organization, structure, syntax, and representation of the target data as well as search method has an important effect on the success of KD process. These operations correct data problems and improve the syntax and semantics to support the KD process. Pyle states (61) that without adequate preparation of data, the return on investment becomes disappointing. Because the discovery is computationally expensive, one needs to have guidance and constraints for an efficient search. The domain knowledge assists discovery by focusing the search. We present semantic nets to capture domain knowledge for this purpose besides the use of controlled vocabulary and ontology. However, the accuracy of such background knowledge as well as ethical, legal, and social issues (ELSI), needs to be assured with timely refreshments. The challenge is to use domain knowledge in such a way that we don't block the discovery of unexpected solutions.

The knowledge representation should be appropriate such as natural language, formal logic, visual depictions, or a combination for the intended user. Therefore, we used RDF/XML as basic knowledge representation language and some encoding schemes such as FOAF and RuleML as higher level representation syntaxes.

Microarray Informatics Overview

Microarray usally forms a part of a greater research. The results of the microarray experiment may address new research directions. Today, there is a strong challenge in microarray data analysis. The number of experiments is on the rise and the amount of information produced by these experiments is enormous. Mining microarray data sets has become important in biomedical research. And, such an analysis needs a common understanding of the experimental results as a prerequisite (62). There is a requirement for efficient and meaningful biological interpretation of microarray data. Such initiatives will enable efficient meta and teoretical studies. The recent research direction has become how to analyze a large amount of microarray data and make biological sense of them. The statistical and computational methods are equally employed besides biological methods. For example, all information about the bacteria “rickettsia conorii” has been produced by using computational methods in place of experimental study. Today, all above-mentioned methods have been an element of microarray informatics. Data archiving was the task in early days, extracting biological knowledge has now become the main task. Active research area in microarray informatics needs to be data management that will support organizing, presenting, exchanging, understanding, and using the experimental results. Such a trend will open the way to interpret the experimental evidence to derive meaningful biological results.

The work of microarray informatics requires orchestrating different disciplines like molecular biology, math, computer science, statistics etc. to have a united focus on its objectives in a team oriented work environment. Experimentation, computational and statistical studies should be synchronized not only to provide required information but also to verify the findings. Once the result of a microarray experiment is published, it becomes a part of biological knowledgebase. The ambiguous, incomplete, even incorrect results could also be propagated into other application areas that use these results. In that regard, there may be false negative and false positive results. False negative can be the act of declaring a gene expression change to be insignificant when in fact a change has occurred. False positive constitutes the opposite situation. And also, statistical measure of the performance of a microarray test could be determined and this value may be a part of the quality of the results. Thus, qualitative data as well as quantitative data should be taken into account in interpreting the experiment results. Such an approach requires the creation and management of intensive metadata.

Gottlieb et al (63) showed that there might be a need to verify the genome-wide association studies based on DNA obtained through blood samples. Because, genetic material (DNA) may vary between blood cells and other tissues for a single individual and may yield different results when used in microarray experiment. Many genome-wide studies use DNA from blood samples. But this study suggests that one can miss some results by looking only at blood. Such a case by itself shows the importance of the evaluation of the experiment setting and their results.

Today, rapidly expanding heterogeneous biological data is considered to be an issue in biology research. Main focus has been the development of methods and technologies supporting high-throughput generation of biological data, such as DNA sequence and gene expression data. Compared to the rapid developments in the area of instrumentation and methods, biological data management is still relatively immature (64).

Dr. Helen Parkinson (65) from EBI says “The information should be structured so that querying and automated analysis and mining are feasible”. It has still not been realized despite intensive standardization initiatives have been conducted for a decade. Only annotation and content standard have successfully been implemented. Our framework is the first comprehensive structured framework to address this challenge.

Microarrays allow simultaneous study of thousands of genes in a single experiment. The generation of large datasets present unique challenges in the representing, sharing, and analysis of that data. It is found useful for solving important biological problems, regarding metabolic pathways, functions of unknown genes, diagnosis of diseased states, as well as the development of individualized drugs. There is much more demand from a microarray experiment than gene annotations. Functional associations between genes are at demand. The goal is to identify metabolic pathways or genetic networks that regulate the cell. Considering the impact of microarray data in biological research, this technology needs the assistance of computational methods for interpreting and utilizing the raw information. The challenge for this field is to couple the microarray experiment data with annotation information implying functional association between genes (66). This is what we do in the proposed metadata framework.

The standardization serves for interoperability, integration, data sharing, and data exchange operations. The standard interpretation of terms and concept contained in the results of a microarray experiment is required. Nothing but only the intended meaning is supposed to be inferred by the reader. Therefore, the writers should pay attention to their terminologies either by referring to a known repository or by explicitly giving a clear definition where necessary. Interoperability, as being the ultimate goal for standardization, has three main aspects : operational to deal with the content, meaning of the information to be exchanged; procedural to deal with the format and representation; technical to deal with functional, electrical, and physical characteristics of the equipments in use. We study the operational and procedural aspects in this study to include organization, structure, syntax, and semantics of the experiment results.

The focus is on the microarray experiment data that is submitted to a repository and published on a journal. This is the most neglected stage of the whole process. Data standards are involved in at this stage. The use of standards are expected to allow the producer to convey all his/her findings while the data becomes ready in content, structure, syntax, and structure to share so that the consumers of this data can discover the knowledge inside. In short, the standards should be used in shaping the data so that the experimenter can transfer all findings and customers can get what they look for. The standards are evolving by new requirements and technological advancements. Some standards are implemented within communities. However when it comes to cross-domain business practices, the lack of standards becomes evident because there are incompatible representations and exchange formats. The subject of this thesis is to look for a commonly understood reporting structure for a microarray experiment. This constitutes a main problem area for microarray repositories. We should not reinvent the wheel and exploit the established standards in this process. These standards are from data management, configuration management, metadata management, knowledge representation, semantic web, and electronic business. There are two main operations on microarray data which are data management (*giving the service*) and knowledge discovery (*receiving the service*). Without a proper data management, the producers cannot get the desired services.

The contemporary software development advocates layers such as presentation, content management, database access, business logic, and integration. Thus, different experts can be involved in the development of different layers. Knowledge based development paradigm requires a separate data layer, similarly. The representation scheme includes domain knowledge with a common format and it may include information ranging from individual information objects to a whole information model based on its maturity. Once the representation scheme is ready, the application code, database model development, or ontology development teams can use this domain data. This structured and semantically powered data will propagate through the information analysis channels for the consumers. There are benefits for both the producer and the consumer in using such a scheme.

We claim to propose a framework which can provide a common reporting format within a meaningful biological context for microarray experimental data. The information is the data in context and the information in a network, which is interconnected and combined with experience, is called knowledge. The data set becomes a data asset when included the services such as discovery and sharing. Without the context and web of relationships to other data sets, an individual data asset is less meaningful. The data whether proprietary application data or public data, has to be structured like data structures in a software or in a database in such a way that machines can understand them. There are four distinct types of microarray data:

- gene data
- sample data
- measured data
- metadata descriptions and annotations which form biological context

Metadata and annotational data, which are not originally present in the information space, are added to support representation, exchange process, and shared semantics. The microarray experiment results themselves are not adequate for biological interpretation. They should be augmented by metadata and annotations to put them in a biological context. Note also that microarray data has high variability. The lifetime management is required. The data sets should be organized such that any user could benefit from the results submitted and published. Currently, microarray data is redundant, incomplete, ambiguous, scattered, heterogeneous, and semi-structured. Structuring the data with encoding schemes, the use of structured data entry and, annotation by ontology, establishing a metadata framework, which are all techniques from data management paradigm, can form such a biological context. Then, this body of data can either be integrated into other data sources or can be a separate data layer of a knowledge based development. The goal is not only exchanging information but sharing meaningful information that can be understood by the consumers. Web services then can be attached to these data sets and one can achieve better integration and interoperability.

The basic goal in KD is to represent knowledge in a manner that facilitates drawing conclusions from knowledge. The expressivity of the knowledge representation is important. The goal in this study is not to analyze knowledge representation paradigm but to demonstrate exemplary representation schemes. XML-based knowledge representation languages such as RDF and OWL have been successfully used by semantic web initiative at our time. The RDF data model includes statements about information objects in the form of subject-predicate-object expressions. RDF, as a notation, has been widespread in many areas because RDF graphs which capture domain knowledge can be used by database modellers, ontology developers, and applications like reasoning systems. The metadata and metadata card is a key data structure for voluminous data. Most of the metadata today is in XML format. We believe that it should be represented in RDF/XML.

Microarray Data Management

Microarray technology is one of the most important experimental breakthroughs in molecular biology. Today, volume of generated data is overcoming the capacity for storing and analyzing it. As more laboratories use this technology, the problem will get worse. This avalanche of data requires standardization of representation and reporting techniques. It is an active research area about a standard format for microarray data, which will facilitate biological inferences (http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html).

There are currently around 20,000 experiments hosted by only GEO. The statistics state that the number of microarray experiments more than doubles every two year. There could be a million experiment records on public repositories by 2020 with this trend. Considering the added annotation, metadata and contextual data in addition to the increasing experiment results, the demand for exchange, integration, and knowledge discovery operations will continue to increase. Our study deals with this issue and offers a data management perspective to resolve the issues concerned. The major problem is the loss of considerable amount of biologically significant information. The term loss here means neither not discovered by the experimenter nor shared by others. The isolated data is considered to be a loss for the rest of the community. Therefore, there is a clear requirement for standardization in capturing, representing, storing, exchanging, and interpretation.

The use of RDF (Resource Desription Framework, http://www.w3.org/RDF**)**

The semantic web is based on RDF. Because the challenge is to define and describe the relationships among data and RDF gives a formal definition for that interchange. RDF is used to represent information and to exchange knowledge in the web. It is the universal format for the data on the web. It allows a mixture of structured and semi-structured data. RDF language expresses data models which refer to objects, resources and their relationships. An RDF-based model can be encoded in XML syntax as well as "n3" or "turtle" formats. It allows to apply semantic descriptions to information resources and facilitates the computational processing of distributed information. RDF together with OWL (*ontology*) provide a formal description of concepts, terms, relationships within a given knowledge domain. RDF and OWL data can describe many things whereas XML schemas just describe documents. RDF provides the following features

- interoperability
- machine understandable semantics
- better precision in resource discovery than full text search
- future-proof applications as schemas evolve

RDF statement is a specific resource (*oval shape, subject, e.g. a web page with Uniform Resource Indicator -URI*) with a named property (*arc, predicate*), plus the value of that property (*rectangle shape, e.g. can be another subject*). RDF schemas can extend the expressive power of the simple resource/property/value triplet into network of related descriptions using types and classes. RDF is a derivative technology which is

- graph based (*whereas XML is tree based*)
- no explicit ordering
- a data model and in XML format

Consider the sentence below:

Fahri Kocabaş is the author of the resource "http://www.ii.metu.edu.tr/MAdmf".

This sentence has the following parts:

Subject (Resource)	http://www.ii.metu.edu.tr/MAdmf
Predicate (Property)	Author
Object (literal)	"Fahri Kocabaş"

The syntax is:

Statement := (predicate,subject,object)

Object = Predicate(Subject)

RDF statement is:

```
<rdf:Statement>
    <rdf:subject resource="http://www.ii.metu.edu.tr/MAdmf">
    <rdf:predicate resource="http://www.purl.org/dublin-core#author">
    <rdf:object>Fahri Kocabaş</rdf:object>
</rdf:Statement>
```

We can read the RDF statement in the Figure 2 as the following.

"http://www.ii.metu.edu.tr/MAdmf" has author "Fahri Kocabaş".

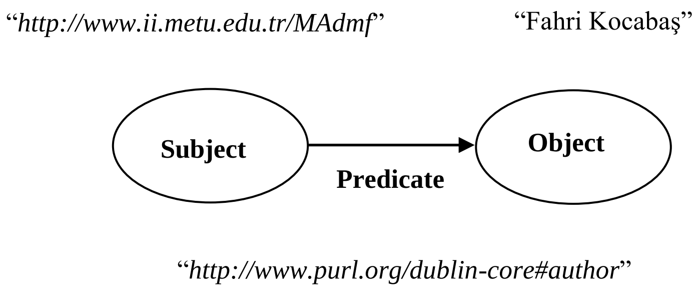

"http://www.ii.metu.edu.tr/MAdmf"
"Fahri Kocabaş"
Subject
Object
Predicate
"http://www.purl.org/dublin-core#author"

Figure 2 - Graph model for an RDF statement (*The RDF statement here is modelled as Directed Labelled Graph model with resources [subject and object] as nodes and predicate as the edge connecting subject to object*)

RDF statements are easy to create. One can describe information objects and their relationships in RDF. An RDF data is self-describing containing data as structured. Once integrated, RDF graphs then can form a model to represent domain semantics. We used RDF in developing semantic nets in the metadata framework. The semantic nets are knowledge representation schemes like concept networks or topic maps. We employed syntax encoding schemes within RDF statements to form the semantic net to capture the domain semantics. It is like using Arden syntax for representing medical statements in HL7 (Health Level Seven, http://www.hl7.org/). OWL, on the other hand, is more advanced. It is used to publish and share set of terms from ontologies supporting advanced web search, software agents, and knowledge management. OWL adds more vocabulary for describing properties and classes among others relationships between classes, cardinality, equality, richer typing of properties, enumerated classes. Ontology experts use OWL to edit ontology. RDF statements can be basis for ontology development as well as database development.

Common semantics is generally bound to a community of interest (COI) but cross domain processes should aim for shared semantics. This work can pave the way for describing data with different format and semantics to be understandable by all parties. The ultimate goal is common understanding. Such a goal can be reached with the structured an semantically powered data.

Biological conclusions and biological confirmation of microarray data is important. Microarray experiments are generating hypothesis. Biological validation of microarray results includes specific examination of identified genes by other high throughput techniques like quantitative reverse transcriptase polymerase chain reaction, RT-PCR. In many studies, this additional yet important aspect has been neglected. The confirmation of the results of an experiment can also be done by another recognized research. MAQC initiative has been launched to maintain the quality and correctness for microarray experiments.

The globally dispersed microarray data comes with varying formats, non-standard syntax and semantics. In order to take full advantage of the wealth of data being generated, data sets need to be collected and stored in a way that they can be searched; the experiments can be verified and reproduced. In computer science a data structure such as record, array, stack, queue, tree, file and graph is a way of organizing and storing data in computer so that it can be processed efficiently. Although data structures in programming arena have been mature and stable for a long time, some additional data structures can be designed to organize and store the data in today's interconnected world. There are structured (*database file*), semi-structured (*html/xml file*), and unstructured (*free text file*) forms of microarray data on experiment site servers, public repositories, and related publications. For example, MINiML and MAGE-ML are semi-structured documents. The structuredness of these XML representations can be extended by adding structural and semantic metadata in a metadata framework. The unstructured data may not communicate the exact meaning. The way we interpret what we read becomes largely subjective. Therefore a structured reporting format is investigated in this study. In the proposed metadata framework, the metadata is structured within metadata card concept and its meaning is represented with semantic nets. Ontology is specific in defining a concept and its relationships. However, the performance of ontology should be supported with additional tools such as rules and syntax encoding schemes to build a commonly understood semantic layer. E.g. gene can be a target (*CYP1A*), producer name (*Gene Hackman*), annotation name (*GO terms*), a word in free text in Turkish (*meaning 'again'*).

With the completion of the Human Genome Project, a new emphasis is focusing on the sequence variation and the resulting phenotype. The completion of human genome sequence has been the initial step in understanding the human biology. The number of data available from genomic studies addressing this relationship is rapidly growing. In order to analyze these data as a whole, they need to be integrated and annotated in a timely manner (67). Only 10% of ~20 thousand genes have been annotated so far. In order to understand biological systems as a whole, a single repository could not be sufficient to provide required data and functionality. We need to integrate several data stores and employ relevant operations to solve complex problems. Note that data exchange and knowledge discovery operations require structured and commonly understood data as input.

Knowledge representation should support the organization and the use of knowledge. That is what we aim to achieve in this study. The use of ontology has been an effective way to annotate the data. However, efficient representation schemes are required to manage biological knowledge (68). Current bioinformatics tools use a great variety of heterogeneous formats; therefore, a common set of defined data formats is needed (54). The MIAME declares the standardization in the content for microarray experiments and also states that it comes without a format. We state that such a format, especially reporting the results, should be in line with Dublin Core metadata standard. Currently there is no metadata card in microarray repositories and metadata standard is not referenced in reporting structures such as MAGE-TAB and MINiML.

There are very different customers (*scientist, journal publisher, legal bodies etc.*) for microarray data. Thus, the same set of data produced by microarray experiment team should be in such a structured report format that it could serve for different purposes in varying detail and format. The producer can improve the expressive power and value of the data by

- structuring the content of data (*data and syntax encoding schemes*)
- structured data entry (*controlled vocabulary, ontology*)

We propose metada card and semantic net to structure the metadata and expose the meaning of the data sets within a framework.

There are solutions in using MO and MAGE-OM for most of the standardization concerns raised in this study provided that they are uniformly implemented by all parties. But, this is not possible in practice. Therefore, we have made this study on microarray data that would be an input to the use of MO and MAGE-OM. The metadata framework described in this thesis is a product from data management arena. Such a work does not replace any ongoing studies. And any modelling or application development can make use of the proposed metadata framework. The data analysis process that can be executed at this framework is given in the Table 5.

Table 5 - Microarray data analysis process

Data Management	- Metadata card in a metadata framework - Structuring the mark-up tags (grouping of tags and building associations) and content - Semantic layer in form of semantic nets to increase expressive power - Microarray metadata registry on an ebXML like central registry.
Data Entry	The use of metadata card per experiment and semantic nets where required. Structured data entry with the use of vocabularies, ontology and patterns as defined in schema files.
Data Exchange	Exchange of metadata card and related SemNets on an ebXML like framework.
Knowledge Discovery	There are different reasons for analyzing microarray experiment results: • Comparative (*one experimenter looking for some finished studies*) • Collaborative (*one experiment that can be implemented with the contribution of different groups of experimenters*) • Meta-analysis (*one experiment that can cover all possible results of microarray experiments with certain properties such as region, age, gender, race, condition, number etc*) • Validation and Redo (*One experiment can validate an existing one as a whole or partially or reproduce it*)

There is a structured data entry initiative within Electronic Health Record (EHR) projects which has faced resistance from health professionals. We believe that this does not apply to the microarray community. Microarray experimenters will desire the visibility and undestandability of their findings. If their data is structured, expressive, and machine processable, then its value is higher for repositories, meta analysts, funding agencies, and publishers.

Semantic Web (SemWeb) and Semantic Interoperability

W3C director Tom Berners-Lee visions web as a universal medium for data, information, and knowledge exchange. Semantic web is web of data rather than web of documents. Semantic web is a technology that makes it easier for people to find and correlate the information they need whereever the data resides. It is based on

RDF. The challenge is to describe the relationships among data. And, RDF gives a formal definition for that interchange. The vision is to extend the public data by bringing the structure and semantics to it. The SemWeb is not a replacement but an extension of the current web. RDF data and related ontologies can be developed incrementally. Biology together with healthcare and life sciences form major application area to adopt semantic web technologies (http://www.w3.org/RDF/FAQ).

We have regulatory information, clinical trial data, genomics, and proteomics data together with our microarray experiment result data set that are all in different places and formats. A scientist needs to explore this distributed and heterogeneous data in a connected way. Semantic web allows the scientists to explore hypotheses, to access and correlate information in unprecedented manner. It contains XML, namespaces, schemas, RDF, ontology and rules. Semantic web uses

- RDF for representation
- SPARQL for queries
- RDF Schema and ontology for structuring
- OWL and rules for reasoning

Semantic interoperability means different things to different people because the context is always different. Different sources use different semantics. Semantics is resolved at the understanding and reasoning level. It starts with domain vocabulary level continues with concept representation level. This can be done with a universal semantic layer (69). We use syntax encoding schemes with RDF/XML notation to capture the domain semantics within metadata framework.

Metadata Concept (http://www.w3.org/metadata)

Metadata is said to be machine understandable information for cross-domain operations. The purpose of metadata is to enhance the understanding of the information to which it is related. All data has the potential to be metadata and is limited only by one’s perspective. What makes it metadata is its purpose and usage rather than its context or structure. Metadata is designed to support people and programs in locating and retrieving information resources. For example, metadata can be data set specific describing the experiments performed and what data was produced. It allows a meaningful analysis of data. Note that the quality of some data in a microarray record in one context may be inadequate in a different context.

There are different types of metadata. As it is captured in metadata card, they could be metadata for administrative purpose (*security, experimenter details*) and metadata for technical purpose (*annotations, discovery services, mark-ups, configuration management*). There may be metadata generated for several areas in microarray information space. For example,

- feature points for experimenters (*experienced, collaborative team*)
- type of the experiment (*meta-analysis, re-do, validation*)
- status (*novel, annotated well, validated with RT-PCR, biological, statistical, computational validations*)
- confidence level
- security, access control and authorization (*who can access for individual genetic data*)
- compartmentalization (*executive summary, results, statistical analysis, raw data*)
- discovery metadata services

The Structure, the Semantics, and Biological Interpretation of Microarray Experiment Results

We used SemNet as a knowledge representation scheme to capture and analyze the meaning of the microarray results. We used RDF graphs as semantic networks. Declarative rules allow integration and transformation of data from multiple sources. We used FOAF and RuleML Datalog in RDF/XML syntax to encode data. We state that when the resulting statements of microarray experiments are encoded in RuleML Datalog, they can be utilized by rule processing systems or can be integrated into SWRL (Semantic Web Rule Language, http://www.w3.org/ Submission/SWRL/) and RIF platforms (Rule Interchange Format, http://www.w3. org/ 2005/rules/).

Experts from CRD (Computational research Division), Berkeley University, USA state that partial but rapidly developed solutions may be more valuable than complete but time consuming solutions (*An overview of the Biological Data Management and Technology Center at http://crd.lbl.gov/html/BDMTC/overview. html*). We believe that such a statement is misleading. If it means to develop evolutionary, phased development in a standardization framework, then it is acceptable. If it means to produce non standard solutions, then it is not acceptable. Mature standards should be visited in facing the challenges in microarray information space. Data standards are the first ones to explore. Data standards can help to represent, organize, store, analysis, exchange microarray data. And the implementations including modelling (*database and ontology*) can benefit from it.

The MIAME is an information exchange requirement (IER) that provides the basis for the structure and semantics of a reporting structure. Such an IER should be shaped into a structured data format. There are unstructured, semi-structured, and structured data sets. Such a definition and classification contains both the syntax and semantics so that data can be represented, exchanged, or shared with common understanding. The work in this direction can be accomplished with data management principles. Such a study and its products create no conflict with any modelling and application development efforts. This is a precursor to all development (*modelling, coding*) efforts. Thus, all further projects can benefit from such a data level work. This is a core study. Thus, the extension is possible. The goal is to structure and organize the data meaningfully so that unambiguous, non-redundant data becomes visible, usable, and understandable by automated means. Once the syntax and the semantics of the data is managed within a framework, the expressive power and information quality is ensured. For this purpose, We present metadata card and semantic nets in a metadata framework for microarray repositories. We propose a framework so that the producers can add more and the consumers can extract more and furthermore new and advanced inferences can be done. This is in principal the semantic web approach which is to bring semantics and structure to the current sea of web content.

In computer science, there are data structures that host structured data with some level of semantics. This is the first step in organising data for an application development. A new data structure can be designed to organise the data. The concept of structuredness came with the advent of semi-structured data with xml tags. XML has brought semi-structured data concept. The most value is gained by bringing structured and unstructured data together. However, they may create more pollution with ambiguities. The challenges are:

- data itself is not consumable from a semantic level
- not necessarily gain insight into the context of the information. The way we interpret what we read is largely subjective

And, the redundancy and ambiguity can be controlled with the degree of structuredness. Without the context and the web of relationships to other data items, an individual data item is meaningless.

Unstructured data is the data that is not consumable from a semantic level. It includes

- no conceptual definition
- no inherent structure
- no rules for formulating queries
- incomplete most of the time

They are either image, video, audio file or textual objects. The unstructured data does not communicate the exact meaning intended (70). A typical example of unstructured data is the collection of documents found on the internet that do not have a common schema (70). Textual data mining can work for unstructured data but it is expensive and semantic inferences are usually not possible.

Semi-structured data is the one labeled by XML tags. And, XML itself has a semi-structured data model. They fall somewhere in between the structured and unstructured data. If the elements of separate documents are unrelated, we have a semi-structured data set. Different elements of the dataset may even require different treatment. It may be structured in the sense that the information content springs from its organisation rather than the individual data elements, but, unstructured in the sense that its component files are not self-describing and metadata will have to be stored to preserve the value of the information. For example, a set of related spreadsheets is an example to a semi-structured data set (71). That is the database of spreadsheets without a common data model (70). It is an attempt to reconcile database and document worlds. The data is available in database, file systems, web, and specific data exchange formats. It generally contains schema within the data which is self-describing, not often a separate schema. Semi-structured data models are

- based on labeled graphs where schema information is in the edge labels and data is stored at the leaves
- used for data exchange among heterogeneous data sources

Structured data is anything that has an enforced composition to include

- predetermined data types
- understood relationships
- defined format under a data model
- complete data sets (*completeness that identifies the structured data*)

Structured data is organized such that similar entities are grouped together with the same descriptions. Relational database and spreadsheets with data model are examples. The value of structured database lies in the relationships between data items rather than in the data items themselves. SQL is the querying language for structured data. The benefit of structured data platform is two fold. The producer could convey whole findings and then the consumer would be able to extract the meaningful results. In this study we treat the less structured data (*unstructured and semi-structured microarray metadata*) in two aspects.

- structure and organize the data in metadata card format
- creating a semantic layer to build concepts and their relationships to represent the meaning of the microarray metadata

The quick fix is for poor data management that can be realized with isolated solutions. However, when it comes to integration, sharing and interoperability operations, such a product starts failing. Identification of data assets and information sharing capability have been challenges for contemporary organizations. The data strategy is to move from privately owned and stored data in disparate networks and within legacy systems to an enterprise information environment where users can access information and derive conclusions. Data should be not only visible and accessible but also understandable and usable. The shared meaning is the basis for a common understanding. Although MIAME is important as a content standard, MIAME compliance can not ensure understandability and interoperability. Semantic and structural content should be added to data assets before the modelling and application development gets started. Microarray data and its supporting data to generate a biomedical context are now at different locations and with incompatible formats. They are redundant, incomplete, ambiguous, and without a lifetime data and change management. Today, the repositories are storing data with varying degrees of structuredness. If we have already identified the context and semantics of our unstructured data, we can bring them together within a structured framework. There is reward for this extra work. The direction in our research has been structuring the data by adding structural and semantic metadata, discovery metadata constructs in an architectural framework. In order to take full advantage of the wealth of data being generated, microarray metadata needs to be represented and stored in such a way that they can be extended, searched, integrated, exchanged, so that the relevant experiments can be analyzed, verified and reproduced. There are similar metadata frameworks in effect. For example, SIMILE (Semantic Interoperability of Metadata and Information in unLike Environments) by MIT (http://simile.mit.edu). It is based on semantic web technologies that improve access, management, interoperability, and reuse among digital assets**.** The rule is that the data assets should be considered together with their schema, vocabulary, ontology, and metadata.

1.6 The Contents of the Dissertation

In this thesis, we introduce a framework, MAdmf (Microarray Discovery Metadata Framework) to handle the issues stated in previous sections. MAdmf is introduced including the process and main elements in the Methods and Case Study chapter. The application of the framework is presented as a case study. We also demonstrate the efficiency of the framework in the Methods and Case Study chapter. The Results and Discussions chapter contains a summary of main findings, which are detailed with some statistics. An overall discussion, main conclusions of the research and explanation of their importance and relevance are given in the Conclusions chapter.

CHAPTER 2

METHODS AND CASE STUDY

2.1 Overview of MAdmf (Microarray Discovery Metadata Framework)

We took GEO (Gene Expression Omnibus, www.ncbi.nih.gov/geo/) at NCBI (National Center for Biotechnical Information) as the reference microarray repository in this work. It is because, GEO is the largest one and provides an XML file (MINiML) supported with an XML Schema file for each submitted record (*per Experiment/Series record*). The Curators transform submitted microarray experiment results into the GEO Datasets. We focused on MINiML in this thesis. We have analyzed the submitted data sets in the GEO repository. We proposed a framework, MAdmf, which includes a format for metadata in microarray results to address listed issues. MAdmf is a metadata framework with semantic capability to complement the current MINiML file. The metadata card, the semantic net (SemNet), and the metadata registry are the key elements of this framework. The metadata card is an index card for storing basic meta data elements about a microarray experiment. The metadata card will provide the reader with information to assist him/her in making a decision as to whether the record(s) might suit his/her needs. SemNet is a small data model to represent specific domain information. The metadata card and the SemNets are encoded in RDF/XML (*metadata language and Semantic Web knowledge representation language*) format. Syntax encoding schemes are used in SemNets. There is a schema for metadata card but the content of the SemNet can be populated based on the current design without a schema. The metadata registry is a shareable repository for metadata and its related SemNet(s). The structure, syntax, and definitions used in the proposed framework, MAdmf, can form the basis and supporting elements for a format and reporting standard for microarray and other high throughput experiment results. Free text within the microarray record is modeled in RuleML Datalog within RDF/XML data representation language. For example, Experiment (*title, type, overall design, result*), Sample (*type, organism, condition*), Condition (*summary*), Protocol (*hybridization, scan, manufacture*), and Platform (*title, technology, manufacturer*) sections are unstructured currently. We have processed these sections within SemNet.

The framework has four components as depicted in Table 6.

Table 6- Microarray Discovery Metadata Framework

Component	**What It Does**
MAdmc (Microarray Discovery Metadata Card)	Reports a microarray experiment through metadata elements
Semantic Layer (*Semantic Nets*)	- Details domain-specific topics, fortifies the intended meaning - Discloses otherwise hidden data
Query Layer (*Optional*)	SPARQL queries
MAdmr (Microarray Discovery Metadata Registry)	Main files for MAdmf are stored at this ebXML based shared space

The content of MAdmf is as follows:

MAdmc.xml: Microarray discovery metadata card
MAdmc.xsd: Schema file for MAdmc
Experimenter.rdf: SemNet (FOAF/RDF file) for experimenters
Result.rdf: SemNet (RuleML Datalog/RDF file) for result/summary section
MAdmc.rq: Query file in SPARQL to run on SemNets

First, we provide a metadata card to include common exchange elements in a standard format in accordance with the metadata standards. Thus, discoverability, semantic interoperability, and integration operations are supported. Since the use of microarray experimental data are exposed to cross-domain operations, shared semantics and understandability are required. The format and structure of MAdmc, as defined in its schema file, is based on Dublin Core (DC) Metadata Initiative, and Metadata Registry Standard (http://www.iso.org, ISO 11179). Second, SemNets are developed for experimenters and results for related experiments. Third, Queries, in SPARQL (Simple Protocol and RDF Query Language) (http://www.w3.org/TR/rdf-sparql-query/) (72), have been developed for information access and discovery operations. Finally, these products (*MAdmc, SemNets, and associated queries*) are stored in a common reference area for further use. They can be exchanged among microarray repositories. Such an exchange or share may reduce the need for multiple submissions and undesired redundancy where raw data resides at its original place. Thus, a microarray repository becomes the reference for further analyses and the next step entries to other specialized databases. The tools being used in the development of these products are from W3C resources, which are in the public-domain. Semantic Web, data management, structured reporting, electronic business management, configuration management (*lifecycle management*, *change management*), and metadata management standards are revisited, and respective concepts and techniques are borrowed from these already proven practices.

The lifecycle management of the records is important. The experimentation and its publication together with some updates on specific databases constitute the first part of the activities in the lifetime of the record. The biomedical community has been successful in this part. However, the important part, which has largely been overlooked, follows this first part and ends when the record is deleted. This second part involves in reuse, validation, modification and knowledge discovery (*for example*, *developing research hypotheses in meta-analyses*) operations. The weakness lies on this part currently as highlighted in several publications (73, 74). This study is performed on this part to make the results visible, understandable, and usable.

The framework in this thesis requires additional resources but such an effort will pay off in data-centric operations in knowledge space. The microarray domain is a data intensive one, and data centric operations are in place. We enforce data management by organizing and structuring data that would improve the quality of microarray data analysis. Therefore, a data management policy should be established.

Data architects should apply data management before system architects and developers do their part afterwards. Data management must be built into the process from the beginning to support information system development. Otherwise, the weakness at the data layer is propagated to other layers and subsequent system development stages are adversely affected. This is an engineering approach with a special focus on data management. The conceptual data model of this data layer can exist outside of the physical model and can be managed separately. It is, in fact, a knowledge-interoperable development that allows domain experts to build or contribute to a separate data layer which can then be incorporated into knowledge based design (75, 76). Although metadata cards can be queried efficiently, the SemNets are better suited for complex queries since they include domain specific content. For example, the domain expert may create a SemNet to include the information 'P53 gene related experiments which finds relevance on arsenite and apoptosis on breast cancer as verified by RT-PCR, published in SCI journal with SCI>10, sample size >20, curated into GDS record and inputted to a specialized repository (*such as GO (Gene Ontology), and Reactome (pathway)*)'.

The metadata card and its associated SemNet(s) (*per record or a group of related records*) may hold frequently accessed data patterns and disclose previously hidden or unavailable content in a standardized and structured format. Thus, much more automated processing is involved. The metadata card and the SemNets can be queried without a need for expensive, dedicated applications for discovery purposes where they contain patterns of data (*like stored queries*) in which one may match the query. It is because they are represented in RDF/XML that is extendable, integrable, queryable, and machine readable. We state that these files can be stored in the GEO and can be exchanged among microarray repositories. Alternatively, they can be stored in a separate metadata registry in the public-domain. So that, the experimenter will not have to prepare different content for different submission requirements for special functional genomics databases (*Gene, protein, microarray, pathway, system biology*). Note that this study deals with the data layer of the domain, and the proposed framework is about organizing and structuring the microarray metadata. The computers can access into the meaning of data which in turn improves the efficiency on knowledge discovery over microarray records by the contribution of this study. With the use of machine processable metadata cards and their related SemNets, not only the users may perform more complex queries but also current backlogs can be reduced. The producer could submit more contextual data in structured form with the proposed framework so that the consumer would be able to extract, infer results accordingly. Microarray analysis has already evolved into microarray informatics, and we believe that the microarray domain requires much more architectural and management wise solutions. The goal to reach shared semantics and common understanding can be realized by applying data management principles over structured and semantically enriched data.

The content and format of data for experiments from repositories greatly vary. With regard to the lifetime of the GEO records, a mechanism should be in place to transfer future modifications on the source to keep synchronization with the target.

We do not intend to discredit current curation and ongoing standardization and development work. But, much more automated means should be used to cope with the pounding expression data. We argue that shareable metadata cards which are semantically powered by SemNets can be a solution. The original records are stored in their respective repositories, but the metadata cards and their related SemNets can be exchanged and shared. The framework presented in this study can be used in other high throughput repositories as well as third party platforms.

2.2 MAdmc (Microarray Discovery Metadata Card)

MAdmc is a metadata card for a microarray experiment. The metadata card is a stable concept and mainly used for resource discovery. Within this framework, it not only facilitates the visibility and discovery but also the usability, and common understanding. With that goal in mind, we extended the structure, organization, and syntax of the MINiML file to produce MAdmc. The expressive power here is two fold: the producer is tempted to include more of the experimental findings and the implicit or previously unavailable data becomes discoverable by consumers who get the intended meaning. We propose the standardization of metadata on the MINiML file by including DC elements and by introducing the metadata card concept. We harmonized the DC metadata element set and relevant attributes with the MINiML file structure. The metadata card has administrative, descriptive, structural, and semantic elements. DC is a standard (ISO 15386) for cross-domain resource description. The use of DC elements in metadata definition also promotes structured entry, as well. Thus, it becomes easy to find and understand information resources. The MINiML seems to serve this purpose but its structure and content is not adequate to perform such functionality. Structuring the records and making structured entry for data elements within the records are closely related and complementing paradigms. The structured entry for the values is enforced by selecting a value from a controlled vocabulary or entering a value dictated by a pattern (*e.g. an expression in the schema file*). In addition to syntax encoding schemes, NER (Named Entity Recognition) and ontology should be used to encode the biological knowledge in microarray records. Within this format standard initiative, the metadata card restructures and extends the MINiML format. The elements and attributes, which are based on DC initiative, are layered within a metadata framework, and lifecycle management concept is introduced with the use of versioning and modification status information.

Microarray records pose more meaning when analyzed in a batch and placed in a biological context. Since the experimental settings, samples, methods, tools, and format widely differ; it is a challenging task for microarray repositories to offer such an analysis in an efficient manner. We bring new concepts with the introduction of the metadata card. We introduce the layers into the organization of metadata elements, and employed data and syntax encoding schemes. Repeatability and structural relationships between elements are defined. For example, the title may be repeated (*alternative title*) which is useful in addition to labels, aliases, and keywords. Or, the use of an element can depend on a condition of another one where the obligation category for the element is conditional. The lifecycle management and versioning are pursued. The lifecycle management covers the period from the submission until the retirement including modifications and versions, thus bringing up the living record concept. For example, it is implemented based on the relation element which may include the values ‘is version of’, ‘replaces’, ‘requires’, ‘part of’, or ‘referenced by’. Thus, this becomes a part of the microarray data rather than the software code. Contradictory and incomplete results, errors, and biases may exist in the records from poor studies. There are also comparability issues between different platforms (*biological samples under similar conditions, protocols, tools*) as pointed out by MAQC project (45). The human or automated users can modify, annotate, verify and re-use a record several times throughout its life-time within permissions allowed.

We developed an XML application (*MAdmc program*) so that the user selects the elements from MINiML document and add new ones from the DC Metadata Set and the attributes from Metadata Registries standard (ISO 11179) on the template provided to create MAdmc. The DC Metadata Set includes fifteen information elements. In MAdmc, we added four new information elements and detailed each element with the introduction of four attributes including an obligation category. We then placed them into four layers as shown in Table 7. MINiML does not conform to DC metadata standard. We added new metadata elements and attributes to make it compliant. These elements constitute the core of the metadata. However, metadata card and Semantic Nets include specific information elements about microarray experiments such as experiment, sample, condition, protocol, and platform as well as result data. Table 7 shows only DC based metadata elements that is only a part of the MAdmc content. Metadata card (*for individual record*) and created SemNets (*for individual and multiple records*) include microarray specific data such as sample, platform, condition, and protocol.

Table 7 - MAdmc elements and Obligation Categories for elements

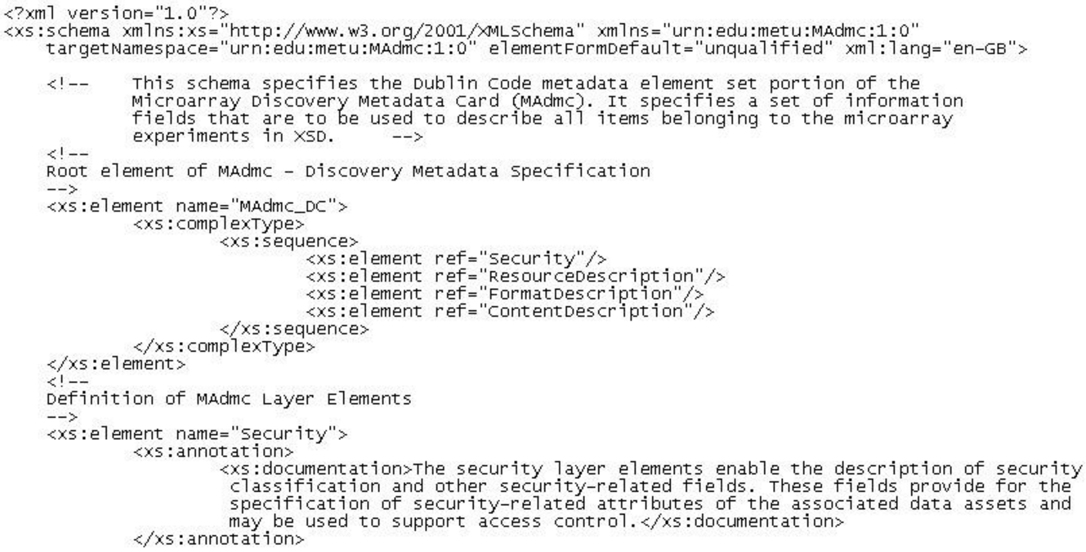
LAYERS (4)	**ELEMENTS (19)**	**ATTRIBUTES (4)**
Security	Policy	Definition Comment Obligation Category Max. Occurrence
	Classification	
	Category	
Resource Description	Title	
	Identifier	
	Creator	
	Publisher	
	Contributor	
	Date	
	Rights	
	Language	
	Type	
	Source	
	Relation	
Format Specification	Format	
	Version	
Content Description	Subject	
	Description	
	Coverage	

7.a : MAdmc elements

OBLIGATION	**DEFINITION**
Mandatory (**M**)	An element must be supplied with a value to comply with MAdmc
Conditional (**C**)	The usage of an element is dependent upon a particular condition
Optional (**O**)	An element may be supplied with a value, but it is not a requirement

7.b : Obligation Categories

The definition of the metadata card is given in MAdmc.xsd file, Figure 3. MAdmc.xsd is the schema file to define the structure and the syntax of the metadata card. The user can reference this schema file to create his/her own instance document (*metadata card*). The experimenter or curator can create the MAdmc file by using MINiML file and the MAdmc program, as explained in the Case Study section.



```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="urn:edu:metu:MAdmc:1:0"
    targetNamespace="urn:edu:metu:MAdmc:1:0" elementFormDefault="unqualified" xml:lang="en-GB">

    <!--    This schema specifies the Dublin Code metadata element set portion of the
            Microarray Discovery Metadata Card (MAdmc). It specifies a set of information
            fields that are to be used to describe all items belonging to the microarray
            experiments in XSD.     -->
    <!--
    Root element of MAdmc - Discovery Metadata Specification
    -->
    <xs:element name="MAdmc_DC">
            <xs:complexType>
                    <xs:sequence>
                            <xs:element ref="Security"/>
                            <xs:element ref="ResourceDescription"/>
                            <xs:element ref="FormatDescription"/>
                            <xs:element ref="ContentDescription"/>
                    </xs:sequence>
            </xs:complexType>
    </xs:element>
    <!--
    Definition of MAdmc Layer Elements
    -->
    <xs:element name="Security">
            <xs:annotation>
                    <xs:documentation>The security layer elements enable the description of security
                     classification and other security-related fields. These fields provide for the
                     specification of security-related attributes of the associated data assets and
                     may be used to support access control.</xs:documentation>
            </xs:annotation>
```

Figure 3 - MAdmc.xsd (*Schema file for microarray discovery metadata card*)

The logical structure of MAdmc can also be extended by employing associations among the tags. The associations can be represented in EBNF (Extended Backus Naur Form) syntax and defined in the schema file, as it was the case for the structural language in NATO (North Atlantic Treaty Organization) structured messaging systems (77). For example, repeat: an element may occur n times; alternative use: information elements such as the title, location, organization may have alternate contents, or alternative entries for different languages are accommodated; occurrence category: information elements are labelled with one of the categories such as 'Mandatory', 'Optional' or 'Conditional'. The existence of another element or presence of a value of another element may require or prohibit the use of an element based on a condition (*e.g. mutual exclusivity*). This extension can make the tags (*tokens*) in the meaning of the content within the structure of MAdmc. Although it is an optional extension, this topic could be visited upon recognition of the metadata concept. The layers (*segmentation*), repeats, nesting, and structural constraints among the mark-up tags can be designed to enhance the structure and meaning in the metadata card. The overall syntax of MAdmc is said to be the pre-established layout for the content, namely a format.

2.3 Semantic Net (FOAF Net and Result Net) – Micro format

Different parts of the metadata card can be detailed with SemNets by submitter or curator. Such a work is analogous to the one performed by domain experts on data layer in knowledge based systems. The SemNets can be generated for each GEO record or a group of related records or the whole repository depending on the contextual requirements. These SemNets accompany their related metadata cards. Knowledge representation by definition should include the intention, context, and domain knowledge by aggregating the data and the metadata. Note that the produced SemNets are machine processable and can be integrated into a related RDF store. Such an RDF store can easily be extended with new SemNets or the integration of other RDF stores. The RDF store can be coupled with any platform and can then be used for varied purposes including ontology development, database modelling, knowledge discovery operations, and any other semantic work. In that regard, semantic nets do not replace or alternate ontologies but support their development.

The encoding schemes can be used for structuring information elements such as experimenters, address, description, and summary in the metadata card. The vocabulary encoding schemes can be used ranging from Controlled Vocabularies (*such as Code lists [ISO 3166-Country codes], Classifications [ICD], Subject headings [MeSH]*), to formal notations (*such as ISO 3601[Date Time Group], ISO 639 [Language], use of specific namespace*). Syntax encoding schemes or micro formats can be used such as FOAF (Friend of a Friend) or RuleML (Rule Markup Language) Datalog. They are used for encoding related data into SemNets. The FOAF is a SemWeb language which describes relationships among people in RDF by forming ontology on its own (http://www.foaf-project.org/). The RuleML Datalog is a markup language for publishing and sharing rule bases. It is based on a deductive reasoning engine and its statements can be embedded in knowledge based systems (http://www.ruleml.org/). The RuleML Datalog is used for formalizing current natural language statements into man readable and machine-processable XML based representation. The challenge is to describe microarray related data in standard and structured way. The format is of secondary concern. This could be the subject of another study and there are already mature, successful implementation areas such as

SWRL, RIF, or structured messaging (*NATO Structured Messaging Standard, ADatP-3*) (77).

The experimenter and the result parts of the microarray record are selected for such an extension in this study. These parts are extended with the SemNets in accordance with relevant syntax to add meaning and to build semantic expressiveness. The experimenters are modelled by using the FOAF syntax, and the result part is modelled by using the RuleML Datalog syntax. Note that the position of an encoding token adds additional meaning in these syntax encoding schemes. Online tools in the public-domain mainly as suggested by W3C are used in the development of the SemNets.

The human concept in the microarray record should be structured. There are types such as human, automated; status such as novel, experienced; roles such as producer, consumer; actors such as submitter, contact, contributor, author of publication, publisher, curator, academician, supplier, funding agency representative, government official, potential experimenter, meta-analyst, verifier, system developer, reviewer etc. Such a detailed definition may hold valuable information for a potential consumer. Data sets are at different maturity levels in terms of structure and content. One’s data may be labelled as metadata or information by someone else. And today’s information may become data in the future in its lifetime. An experimenter may need to make a search in this element to make some decisions for experiment design. In that regard, the human concept in microarray record can also be represented in a standard structured form. There are mature formats such as hcard (http://www.w3.org/2006/03/hcard), vcard (http://www.w3.org/TR/vcard-rdf), or W3C’s PIM (Personal Information Management) (http://www.w3.org/2000/10/swap/pim/) to include into the FOAF model to form a coalition of complementing vocabularies.

The result/summary information has also been a frequently accessed area. This portion of the microarray record should also have a machine understandable structure and content. For that reason, we employed an encoding process for this part of the microarray record to create a SemNet. We included three elements in this encoding which are the free text statements, the encoded format in RuleML Datalog, and annotations which are all in RDF/XML notation. The details of SemNets are given in the following case study.

The properties and relationships of information resources are described within RDF graphs for SemNets (*experimenter net [in FOAF] and result net [in RuleML Datalog]*) in the study. These are associated to each or a group of related MAdmc record(s) in accordance with which specific knowledge is represented. Thus, FOAF and Result nets can be packed with the metadata card while ontology use is in place. More data are stored in RDF format to create linked data today. The RDF files (*SemNets*) can be integrated into a persistent RDF store to form connected graphs. The SemNets are data models (*RDF statements constitute a directed graph*), that are easy to create for a specific domain information, which can support both ontology development and database design. Ontology extensions can subsequently be built from these SemNets. For example, describing a person in ontology may eventually converge to a FOAF model. A new vocabulary and ontology construct can be generated from the RDF resources in the RDF store. The RDF triples for information objects may become instances for existing OWL classes or may trigger the creation of new classes for specific concepts. It is also obvious that ontology terms should be used as the tokens in a SemNet where possible as well as CV and Code lists to structure the data and to make it commonly understandable. The use of ontology ensures that commonly understood descriptions of key experimental concepts are used which in turn facilitates automated data exchange. Ontology is used for annotation, but we encode the data and the metadata with syntax systems in the SemNets. In that regard, MAdmf not only supports ontology development but also tempts structured data entry.

There is a proliferation of ontologies, and there are interoperability problems among them. Standardization efforts such as OBO Foundry and OBI focus on upper ontology development, whereas lower level ontologies remain in the realm of domain-specific ontologies such as MGED Ontology. Ontology is a conceptual model that may not map to physical data sources whereas a SemNet does. Ontologies are developed over time and are at conceptual level whereas semantic nets can be developed for each GEO record and include specific, instance data. Semantic nets in

RDF format can serve as a basis for bottom up ontology development since information resources with their relations are captured in the domain and cross domain information space. Ontology is monotonic where new statements should not falsify previous conclusions (78). Regarding microarray experiments, there are conflicting results as well as validating ones and SemNets may include such non-monotonic business rules. There are potential consumers for both the metadata card and its SemNet(s) such as RDF stores/databases that provide both reasoning and ontology modelling capabilities. Another one is a semantic platform that annotates microarray data such as THEA by Pasquier et al (28), GandrKB by Schober et al (79), and AILUN by Chen at al (80).

One can combine people, place, organization, and time information with experiment results to formulate complex inquiries by considering the SemNet and the metadata card. Moreover, the development of knowledge interoperable systems with a separate data layer can be facilitated with such a mode of operation on data. Equally, rule based systems can make use of the summary portion of a microarray record that is structured and encoded.

2.4 Queries

Some frequently asked queries can be materialized in SPARQL within the framework and be posted to shared registry (MAdmr). SPARQL is similar to SQL and is de-facto standard as being RDF Query language among others. Some queries for which the answers are not possible or difficult to obtain at the moment such as the following can then become possible when MAdmf is employed.

- List submitters who have worked on breast cancer on Tamoxifen effect on humans within X organization for which the records have been curated to GDS.
- List breast cancer records that have been published in SCI journals with citation numbers > 10 and verified (RT-PCR or confirmed with another study) and have been included in special databases.
- List all facts/hypotheses from records related to the P53 gene between 2000 and 2010.
- List the versions, states (modified, retired etc), type (comparative, collaborative, meta-analysis, validation/re-do etc) and modification details for experiments of genes BRCA1 and BRCA2 related records.
- List super GSE records and their child records that are related to experimentation on gene ATM that finds relevance on arsenite and apoptosis on breast cancer by submitters from USA in the last decade.

The metadata card and SemNets can hold data to answer these questions in a knowledge representation format. One sample query and its result are demonstrated within the Case Study section.

2.5 MAdmr (Microarray Discovery Metadata Registry)

The metadata registry will be the key element to enforce a data strategy by facilitating visibility of data assets. There are metadata registry implementations which are based on ebXML (Electronic Business using eXtensible Markup Language, http://www.ebxml.org/) such as NATO Metadata Repository and Registry (http://si.nc3a.nato.int/projects/metadata-registry), and USA Department of Defense Metadata Registry (https://metadata.dod.mil/mdr/). The submission package to this ebXML based shared space may include MAdmc, SemNet, schema file, query file, and a guidance document, Figure 4.

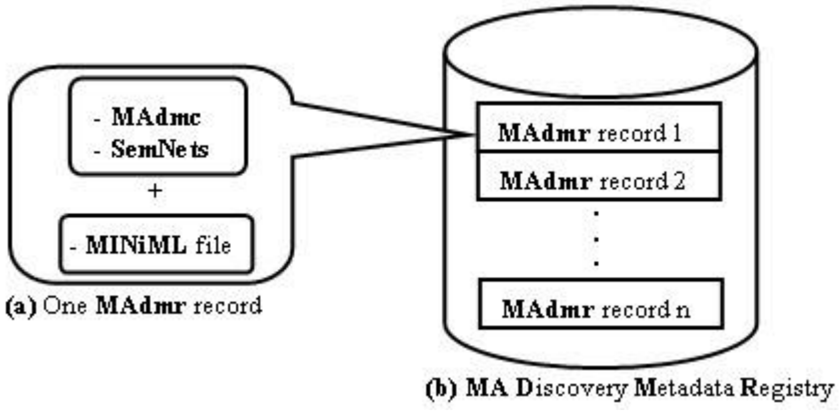


Figure 4 - The MAdmr content

Microarray repository or a metadata registry should be a reference location for the primary findings. MAdmr can be either GEO or another repository. A federated system of microarray repositories can also assume the metadata registry role to host microarray discovery data.

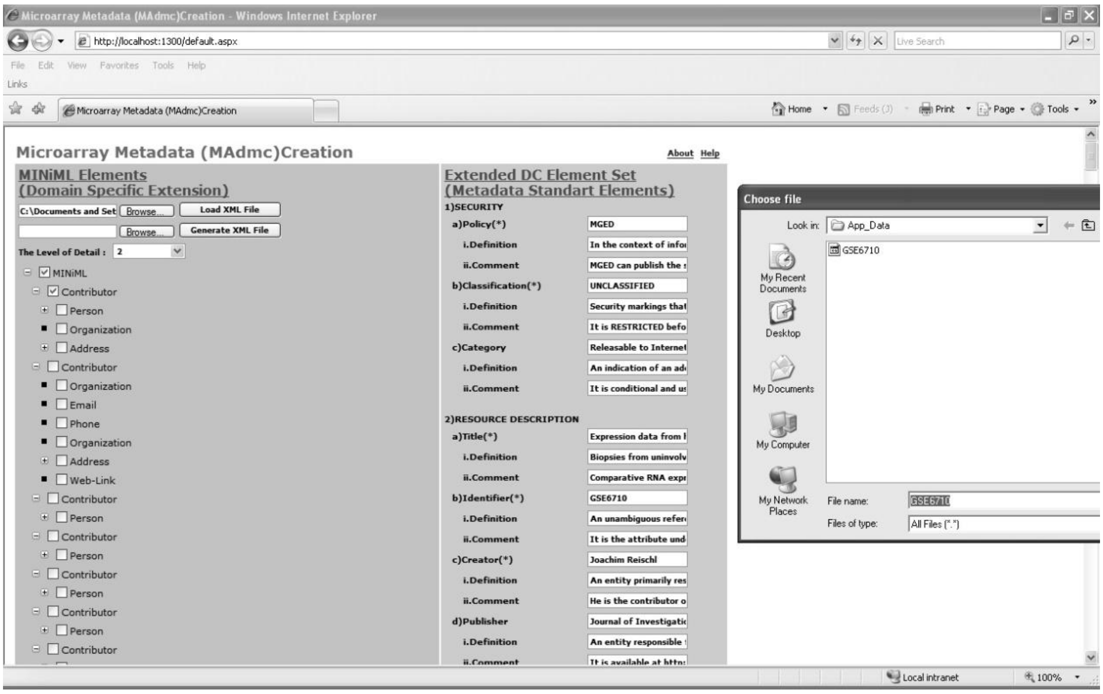
Different users (*such as submitter, reviewer, or web services program*) can subscribe to such a registry. And, producer(s) can make modifications and create new versions throughout the lifetime of the microarray records prior to retirement on the metadata registry.

2.6 The Case Study

The workflow for this case study can be given as:

- GEO has been selected as a model microarray repository
- MINiML file for Series record is examined
- Metadata card (MAdmc.xml) is created with the use of program being developed utilizing the existing MINiML.xml file
- Semantic nets are generated for biological knowledge to include sample, platform, and protocol as well as title, description, overall design, summary fields (RuleML/RDF) and experimenters (FOAF/RDF)
- SPARQL query scripts are built to extract information previously hidden, unambiguous or even unavailable
- MAdmf files are stored in one single storage area at the GEO (or alternatively, they are submitted to an ebXML based metadata registry, MAmdr)

The GEO records (*Series, Platform, and Sample*) and contact data were downloaded and loaded into MS Access Database and examined in terms of structure and semantics. There were 677 Breast Cancer experiment results (*677 GSE records, 89 GDS records*) in approximately 30000 Series records for the case study. The focus has been on MINiML file, which is a part of each record. The MINiML file (*GSExxx_family.xml*) together with its XML Schema file (*MINiML.xsd*) per record has been examined. We developed the metadata card by using the MAdmc program, Figure 5.



Microarray Metadata (MAdmc)Creation

MINiML Elements
(Domain Specific Extension)

C:\Documents and Set

The Level of Detail : 2

- ☑ MINiML
  - ☑ Contributor
    - ☐ Person
    - ☐ Organization
    - ☐ Address
  - ☐ Contributor
    - ☐ Organization
    - ☐ Email
    - ☐ Phone
    - ☐ Organization
    - ☐ Address
    - ☐ Web-Link
  - ☐ Contributor
    - ☐ Person
  - ☐ Contributor
    - ☐ Person
  - ☐ Contributor
    - ☐ Person
  - ☐ Contributor
    - ☐ Person
  - ☐ Contributor

Extended DC Element Set
(Metadata Standart Elements)

1)SECURITY
a)Policy(*) MGED
i.Definition In the context of infor
ii.Comment MGED can publish the s
b)Classification(*) UNCLASSIFIED
i.Definition Security markings that
ii.Comment It is RESTRICTED befo
c)Category Releasable to Internet
i.Definition An indication of an ad
ii.Comment It is conditional and us

2)RESOURCE DESCRIPTION
a)Title(*) Expression data from I
i.Definition Biopsies from uninvolv
ii.Comment Comparative RNA expr
b)Identifier(*) GSE6710
i.Definition An unambiguous refer
ii.Comment It is the attribute und
c)Creator(*) Joachim Reischl
i.Definition An entity primarily res
ii.Comment He is the contributor o
d)Publisher Journal of Investigatic
i.Definition An entity responsible
ii.Comment It is available at http:

Look in: App_Data
GSE6710
File name: GSE6710
Files of type: All Files (*.*)

Figure 5 - MAdmc program. An application that reads the MINiML file, accepts values for additional fields, and creates the metadata card (MAdmc.xml)

Then, two sets of SemNets were created per record(s) by using Rhodonite (*RDF Editor/SPARQL Query Module*, http://rhodonite.angelite.nl/), Protégé (*Ontology Editor and Knowledge Acquisition System*, http://protege.stanford.edu/), XML Notepad, online W3C XML Schema Validator (http://www.w3c.org/2001/03/webdata/xsv), and RDF Validation Service (http://www.w3.org/RDF/Validator/). SemNets (*RDF graphs*) in Protégé and Rhodonite are queried by using SPARQL. First SemNet was for experimenters in FOAF/RDF, and the second one was about the result section including relevant biological knowledge such as sample, platform, protocol, Tables 8a and 8b respectively. Note that the examples about these SemNets are given for proof of concept only, and the advanced ones can be developed by referring to their online resources.

Two encoded statements (*casual first order logic*) by using RuleML Datalog are given in Table 8.

Table 8 - Statements from the GEO records encoded in the RuleML Datalog

MicroRNA silences anti-proliferative genes	Free text
<rulebase> <fact> <Atom> <opr><Rel>silence</Rel></opr> <arg index=”1”><Ind>MicroRNA</Ind></arg> <arg index=”2”><Ind>anti-proliferative gene</Ind></arg> </Atom> </fact> </rulebase>	Encoded text

(a) A fact from GSE12848

RT-PCR confirms the induction of early growth response1 (Egr1) and Stratifin (Sfn) by estradiol-progesterone (EP) and RT-PCR shows that P53 is independent	Free text
<Rulebase> <Atom> <Rel>confirm</Rel> <Ind>induction</Ind> <And> <Ind>Egr1</Ind> <Ind>Sfn</Ind> </And> <Ind>EP</Ind> <Var id=1>RT-PCR</Var> </Atom> <And> <Atom> <Rel>show</Rel> <Rel>independent</Rel> <Var>P53</Var> <Var id=1>RT-PCR</Var> </Atom> </And> </Rulebase>	Encoded text

(b) A rule from GSE5483

We show an entry level encoding in Table 8 to give an insight. The statements could have been further categorized such as experimental, statistical, and computational or its status could be labelled such as verified, challenged, or withdrawn. Linking verbs, passive voice, plural endings, and tense suffixes can be removed in normalization process for a canonical format before encoding it in RuleML. The goal is to highlight the elements of MAdmf. Thus, we do not claim to present the optimal representation. We here demonstrate that the results can be formatted in a syntax encoding scheme like RuleML Datalog. This structured set of statements can then be shared and processed by automated means.

The individual statements for each of these 677 breast cancer GEO records can form a semantic net that is associated to the relevant MAdmc. There may also be global statements about meaningful findings for a specific sub-group of records or whole breast cancer records. SemNets can be in different representations such as triple notation, and graph diagram as well as XML/RDF format. We include three elements in this encoding of the SemNet: the original statements, the encoded format, and annotations. The annotation part of this package provides contextual information and may include whether

- there is a related publication?
- the results are posted somewhere else such as GO or a pathway database?
- there are any changes to results since first submission; there are other versions?
- it is a fact or hypothesis?
- it is verified or challenged?

Although the answers to above questions can be answered in individual projects at high costs provided that the related information exists, such a solution would be a proprietary one. With MAdmf platform, the relevant information has been captured and recorded into a shared space. Then, text miners, rule builder or, GO/Pathway studies can discover such information previously ambiguous, hidden or unavailable. Relevant vocabularies with namespace declarations like “MAdmc” are included into the MAdmc schema definition file to support the additional definitions. Sample Experimenter (*in FOAF format*) and Result (*in RuleML Datalog format*) SemNets are given in Table 9-12 and their graphical outputs from RDF Validator are given in Figure 6 and 7 respectively. In Result SemNet, Ontology (MGED Ontology, http://mged.sourceforge. net/ontologies/MGEDOntology.owl) and NCBI taxonomy (http://www.ncbi.nlm. nih.gov/Taxonomy/) terms are used and named entity recognition (NER) data, as human-annotated, has been included as a preprocessed text for the consumption by NER applications.

Table 9 – Semantic Net (FOAF based Experimenter SemNet for GEO Series record, GSE12848) in RDF/XML

```
<?xml version="1.0" encoding="UTF-16"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:foaf="http://xmlns.com/foaf/0.1/"
        xmlns:vCard="http://www.w3.org/2006/vcard/ns#"
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:MAdmc="http://www.ii.metu.edu.tr/MAdmc#">

<rdf:Description rdf:about="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848">
        <dc:title>p53-repressed miRNAs are Involved with E2F in a Feed Forward Loop Promoting Proliferation</dc:title>
</rdf:Description>

<vCard:ADR>
        <vCard:Street>11 Canal Road</vCard:Street>
        <vCard:Locality>Oslo</vCard:Locality>
        <vCard:Code>N-0130</vCard:Code>
        <vCard:Country>Norway</vCard:Country>
        <vCard:email rdf:resource="mailto:h.johnsen@ous-research.no"/>
</vCard:ADR>

<foaf:Organization>
        <foaf:name>Norwegian Radium Hospital</foaf:name>
        <foaf:homepage rdf:resource="http://www.rikshospitalet.no/"/>
        <foaf:phone rdf:resource="tel:+4722934421"/>
</foaf:Organization>

<foaf:Person rdf:about="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848">
        <foaf:name>Johnsen H</foaf:name>
        <foaf:title>contact</foaf:title>
</foaf:Person>

<foaf:Person rdf:ID="me">
        <foaf:name>Hilde Johnsen</foaf:name>
        <foaf:title>Dr</foaf:title>
        <foaf:mbox_sha1sum>1aa04bd93b66d95b8c303e11bb5c8a2d2cd33618</foaf:mbox_sha1sum>
        <foaf:homepage rdf:resource="http://www.rikshospitalet.no/"/>
        <foaf:phone rdf:resource="tel:+4722934421"/>
        <foaf:workInfoHomepage rdf:resource="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848"/>

        <MAdmc:Journal  rdf:parseType="Literal">
            <PMID>19034270</PMID>
        </MAdmc:Journal>

        <MAdmc:co_author>
            <foaf:Person>
               <foaf:name>Brosh R</foaf:name>
               <foaf:title>contrib1</foaf:title>
            </foaf:Person>
        </MAdmc:co_author>

        <foaf:currentProject>
            <rdf:Description rdf:about="http://biology.unc.edu/faculty/lieb/labpages/home.shtml"
                dc:title="Combining biochemical and genomic methods to study genome and chromatin organization"/>
        </foaf:currentProject>
        <foaf:pastProject>
            <rdf:Description rdf:about="https://cabig.nci.nih.gov/"
                dc:title="cancer Biomedical Informatics Grid (caBIG)"/>
        </foaf:pastProject>
        <foaf:interest>
            <rdf:Description rdf:about="http://www.w3.org/2000/01/sw/"
                dc:title="Semantic Web"/>
        </foaf:interest>

</foaf:Person>
</rdf:RDF>
```

Table 10 - RDF triples of the data model for the semantic net in Table 9 (FOAF/RDF) (http://www.w3.org/RDF/Validator/)

No	Subject	Predicate	Object
1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://purl.org/dc/elements/1.1/title	"p53-repressed miRNAs are Involved with E2F in a Feed Forward Loop Promoting Proliferation"
2	genid:A34733	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2006/vcard/ns#ADR
3	genid:A34733	http://www.w3.org/2006/vcard/ns#Street	"11 Canal Road"
4	genid:A34733	http://www.w3.org/2006/vcard/ns#Locality	"Oslo"
5	genid:A34733	http://www.w3.org/2006/vcard/ns#Code	"N-0130"
6	genid:A34733	http://www.w3.org/2006/vcard/ns#Country	"Norway"
7	genid:A34733	http://www.w3.org/2006/vcard/ns#email	mailto:h.johnsen@ous-research.no
8	genid:A34734	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Organization
9	genid:A34734	http://xmlns.com/foaf/0.1/name	"Norwegian Radium Hospital"
10	genid:A34734	http://xmlns.com/foaf/0.1/homepage	http://www.rikshospitalet.no/
11	genid:A34734	http://xmlns.com/foaf/0.1/phone	tel:+4722934421
12	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
13	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://xmlns.com/foaf/0.1/name	"Johnsen H"
14	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://xmlns.com/foaf/0.1/title	"contact"
15	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
16	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/name	"Hilde Johnsen"
17	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/title	"Dr"
18	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/mbox_sha1sum	"1aa04bd93b66d95b8c303e11bb5c8a2d2cd33618"
19	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/homepage	http://www.rikshospitalet.no/
20	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/phone	tel:4722934421
21	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/workInfoHomepage	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848
22	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://www.ii.metu.edu.tr/MAdmc#Journal	"&lt;PMID&gt;19034270&lt;/PMID&gt;"^^http://www.w3.org/1999/02/22-rdf-syntax-ns#XMLLiteral
23	genid:A34735	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
24	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://www.ii.metu.edu.tr/MAdmc#co_author	genid:A34735

Table 10 (cont.)

25	genid:A34735	http://xmlns.com/foaf/0.1/name	"Brosh R"
26	genid:A34735	http://xmlns.com/foaf/0.1/title	"contrib1"
27	http://biology.unc.edu/faculty/lieb/labpages/home.shtml	http://purl.org/dc/elements/1.1/title	"Combining biochemical and genomic methods to study genome and chromatin organization"
28	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/currentProject	http://biology.unc.edu/faculty/lieb/labpages/home.shtml
29	https://cabig.nci.nih.gov/	http://purl.org/dc/elements/1.1/title	"cancer Biomedical Informatics Grid (caBIG)"
30	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/pastProject	https://cabig.nci.nih.gov/
31	http://www.w3.org/2000/01/sw/	http://purl.org/dc/elements/1.1/title	"Semantic Web"
32	http://www.w3.org/RDF/Validator/run/1337791387866#me	http://xmlns.com/foaf/0.1/interest	http://www.w3.org/2000/01/sw/

Table 11 – Semantic Net (Result SemNet of GEO Series record, GSE12848) in RDF/XML

```
<?xml version="1.0" encoding="UTF-16"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
            xmlns:dc="http://purl.org/dc/elements/1.1/"
            xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
            xmlns:ns="http://www.example.org/#"
            xmlns:xs="http://www.w3.org/2001/XMLSchema"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xmlns:rm="http://www.ruleml.org/0.9/xsd"
            xmlns:MAdmc="http://www.ii.metu.edu.tr/MAdmc#">

<!-- Breast Cancer GSE Records (677) as of May 2011 -->
<rdf:Description
            rdf:about="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848">
            <dc:title>p53 related Breast Cancer Record</dc:title>
            <dc:date rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-09-23</dc:date>
            <MAdmc:ner rdf:parseType="Literal">
<ENAMEX TYPE="PERSON">hilde johnsen</ENAMEX>submitted this<ENAMEX
TYPE="GENE">p53</ENAMEX>based experiment that is the part of<NUMEX
TYPE="QUANTITY">677</NUMEX>breast cancer records at<ENAMEX
TYPE="ORGANIZATION">GEO</ENAMEX>in<TIMEX TYPE="DATE">May 2011</TIMEX>
            </MAdmc:ner>
            <dc:description>A P53 related breast cancer Series record is described in this SemNet</dc:description>
            <MAdmc:mo> id=”MO_213“ category=”experimentDesignType”
value=“comparative_genome_hybridization_design“</MAdmc:mo>
            <dc:source>You can access GSE in this link</dc:source>
            <MAdmc:keyword>breast cancer, p53, E2F1, miRNA</MAdmc:keyword >
            <MAdmc:sample> category1=”title” value1=”MicMA” category2=”extracted module” value2=”genomic DNA”
id3=”NCI_9606” category3=”organism” value3=”Homo sapiens”</MAdmc:sample>
            <MAdmc:condition> id1=”MO_129” category1=”Biometrics” description1=”tumor cell percentage”
value1=”40” id2=”MO_610” category2=”BioSampleType” value2=”frozen_sample” </MAdmc:condition>
            <MAdmc:protocol> category1=”extraction protocol” value1=”fresh_frozen at -80%” category2=”label protocol”
value2=”Agilent miRNA labeling and microarray hybridization protocol v0.1” category3=”hybridization protocol”
value3=”Agilent miRNA labeling and microarray hybridization protocol v0.1” category4=”scan protocol” value4=”Agilent
microarray hybridization scanner, G25665A” </MAdmc:protocol>
</rdf:Description>

<rdf:Description rdf:nodeID="GSE12848">
            <dc:identifier>GSE_12848</dc:identifier>
            <dc:title>p53-repressed miRNAs are involved with E2F in a Feed Forward Loop Promoting
Proliferation</dc:title>
            <MAdmc:status>category=”experiment” state=”modified” status=”verified” status=”RT-PCR” rule:1:
MicroRNAs silence anti-proliferative  genes</MAdmc:status>
            <MAdmc:similar>GSE5483</MAdmc:similar>
            <MAdmc:publication>PMID=”19034270” SCI=”11” IF=”12.125”
SpecialDB=http://www.uniprot.org/uniprot/Q8TCJ2BiologicalPathway=http://www.reactome.org/</MAdmc:publication>

<!-- The first statement is encoded in RuleML here -->
            <MAdmc:RuleMLDatalog>
                    <rm:Atom>
                    <rm:Rel>silence</rm:Rel>
                    <rm:Ind>MicroRNA</rm:Ind>
                    <rm:Ind>anti-proliferative gene</rm:Ind>
                    </rm:Atom>
            </MAdmc:RuleMLDatalog>

            <MAdmc:ruleset>
                    1:MicroRNAs silence anti-proliferative genes
                    2:MicroRNAs are novel key players in the mammalian cellular proliferation network
                    3:Expression of microRNAs is down-regulated in in breast cancers harboring wild-type p53.
                    4:MicroRNAs are repressed by p53 in an E2F1-mediated manner.
            </MAdmc:ruleset>

<MAdmc:summary> source=”abstract” url=”http://www.ncbi.nlm.nih.gov/pubmed/19034270” text=”Normal cell growth is
governed by ...”</MAdmc:summary>

</rdf:Description>
</rdf:RDF>
```

Table 12 - RDF triples of the data model for the semantic net in Table 11 (*P53 gene related breast cancer record, GSE12848*) (http://www.w3.org/RDF/Validator/)

No	Subject	Predicate	Object
1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://purl.org/dc/elements/1.1/title	"p53 related Breast Cancer Record"
2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://purl.org/dc/elements/1.1/date	"2008-09-23"^^http://www.w3.org/2001/XMLSchema#date
3	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.ii.metu.edu.tr/MAdmc#ner	"&lt;ENAMEX TYPE="PERSON"&gt;hilde johnsen&lt;/ENAMEX&gt;submitted this&lt;ENAMEX TYPE="GENE"&gt;p53&lt;/ENAMEX&gt;based experiment that is the part of&lt;NUMEX TYPE="QUANTITY"&gt;677&lt;/NUMEX&gt;breast cancer records at&lt;ENAMEX TYPE="ORGANIZATION"&gt;GEO&lt;/ENAMEX&gt;in&lt;TIMEX TYPE="DATE"&gt;May 2011&lt;/TIMEX&gt;"^^http://www.w3.org/1999/02/22-rdf-syntax-ns#XMLLiteral
4	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://purl.org/dc/elements/1.1/description	"The Result of a P53 related breast cancer Series record is captured in this SemNet"
5	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.ii.metu.edu.tr/MAdmc#mo	"id=”MO_213“ category=”experimentDesignType” value=“comparative_genome_hybridization_design ““"
6	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://purl.org/dc/elements/1.1/source	"You can access GSE in this link"
7	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.ii.metu.edu.tr/MAdmc#keyword	"breast cancer, p53, E2F1, miRNA"
8	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.ii.metu.edu.tr/MAdmc#sample	"category1=”title” value1=”MicMA” category2=”extracted module” value2=”genomic DNA” id3=”NCI_9606” category3=”organism” value3=”Homo sapiens”"
9	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.ii.metu.edu.tr/MAdmc#condition	"id1=”MO_129” category1=”Biometrics” description1=”tumor cell percentage” value1=”40” id2=”MO_610” category2=”BioSampleType” value2=”frozen_sample”"

Table 12 (cont.)

10	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848	http://www.ii.metu.edu.tr/MAdmc#protocol	"category1=”extraction protocol” value1=”fresh_frozen at -80%” category2=”label protocol” value2=”Agilent miRNA labeling and microarray hybridization protocol v0.1” category3=”hybridization protocol” value3=”Agilent miRNA labeling and microarray hybridization protocol v0.1” category4=”scan protocol” value4=”Agilent microarray hybridization scanner, G25665A””"
11	genid:UGSE12848	http://purl.org/dc/elements/1.1/identifier	"GSE_12848"
12	genid:UGSE12848	http://purl.org/dc/elements/1.1/title	"p53-repressed miRNAs are involved with E2F in a Feed Forward Loop Promoting Proliferation"
13	genid:UGSE12848	http://www.ii.metu.edu.tr/MAdmc#status	"category=”experiment” state=”modified” status=”verified” status=”RT-PCR” rule:1: MicroRNAs silence anti-proliferative genes"
14	genid:UGSE12848	http://www.ii.metu.edu.tr/MAdmc#similar	"GSE5483"
15	genid:UGSE12848	http://www.ii.metu.edu.tr/MAdmc#publication	"PMID=”19034270” SCI=”11” IF=”12.125” SpecialDB=http://www.uniprot.org/uniprot/Q8TCJ2BiologicalPathway=http://www.reactome.org/"
16	genid:A153123	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.ruleml.org/0.9/xsdAtom
17	genid:UGSE12848	http://www.ii.metu.edu.tr/MAdmc#RuleMLDatalog	genid:A153123
18	genid:A153123	http://www.ruleml.org/0.9/xsdRel	"silence"
19	genid:A153123	http://www.ruleml.org/0.9/xsdInd	"MicroRNA"
20	genid:A153123	http://www.ruleml.org/0.9/xsdInd	"anti-proliferative gene"
21	genid:UGSE12848	http://www.ii.metu.edu.tr/MAdmc#ruleset	"1:MicroRNAs silence anti-proliferative genes 2:MicroRNAs are novel key players in the mammalian cellular proliferation network 3:Expression of microRNAs is down-regulated in senescent cells and in breast cancers harboring wild-type p53. 4:MicroRNAs are repressed by p53 in an E2F1-mediated manner."
22	genid:UGSE12848	http://www.ii.metu.edu.tr/MAdmc#summary	"source=”abstract” url=”http://www.ncbi.nlm.nih.gov/pubmed/19034270” text=”Normal cell growth is governed by a complicated biological system, featuring multiple levels of control, often deregulated in cancers…"

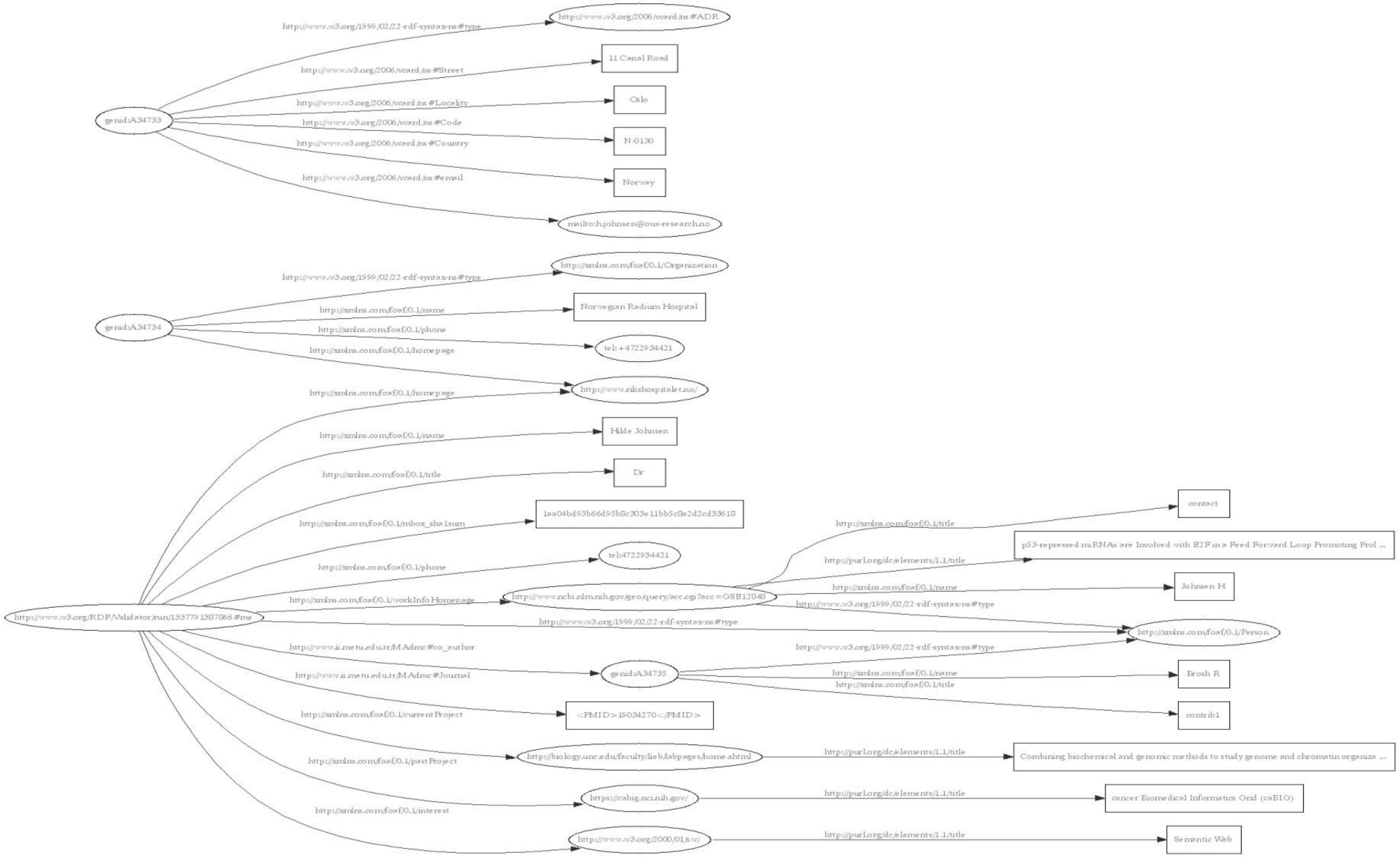

genid:A34733
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://www.w3.org/2006/vcard/ns#ADR
http://www.w3.org/2006/vcard/ns#Street
11 Canal Road
http://www.w3.org/2006/vcard/ns#Locality
Oslo
http://www.w3.org/2006/vcard/ns#Code
N-0130
http://www.w3.org/2006/vcard/ns#Country
Norway
http://www.w3.org/2006/vcard/ns#email
mailto:h.johnsen@ous-research.no
genid:A34734
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://xmlns.com/foaf/0.1/Organization
http://xmlns.com/foaf/0.1/name
Norwegian Radium Hospital
http://xmlns.com/foaf/0.1/phone
tel:+4722934421
http://xmlns.com/foaf/0.1/homepage
http://www.rikshospitalet.no/
http://www.w3.org/RDF/Validator/run/1337791387866#me
http://xmlns.com/foaf/0.1/homepage
http://xmlns.com/foaf/0.1/name
Hilde Johnsen
http://xmlns.com/foaf/0.1/title
Dr
http://xmlns.com/foaf/0.1/mbox_sha1sum
1aa04bd93b66d95b8c303e11bb5c8a2d2cd33618
http://xmlns.com/foaf/0.1/phone
tel:4722934421
http://xmlns.com/foaf/0.1/workInfoHomepage
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848
http://xmlns.com/foaf/0.1/title
contact
http://purl.org/dc/elements/1.1/title
p53-repressed miRNAs are Involved with E2F in a Feed Forward Loop Promoting Prol ...
http://xmlns.com/foaf/0.1/name
Johnsen H
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://xmlns.com/foaf/0.1/Person
http://www.ii.metu.edu.tr/MAdmc#co_author
genid:A34735
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://xmlns.com/foaf/0.1/name
Brosh R
http://xmlns.com/foaf/0.1/title
contrib1
http://www.ii.metu.edu.tr/MAdmc#Journal
<PMID>19034270</PMID>
http://xmlns.com/foaf/0.1/currentProject
http://biology.unc.edu/faculty/lieb/labpages/home.shtml
http://purl.org/dc/elements/1.1/title
Combining biochemical and genomic methods to study genome and chromatin organiza ...
http://xmlns.com/foaf/0.1/pastProject
https://cabig.nci.nih.gov/
http://purl.org/dc/elements/1.1/title
cancer Biomedical Informatics Grid (caBIG)
http://xmlns.com/foaf/0.1/interest
http://www.w3.org/2000/01/sw/
http://purl.org/dc/elements/1.1/title
Semantic Web

Figure 6 – Graphical representation of semantic net in Table 9 (FOAF/RDF) to represent the submitters as validated by the RDF Validator

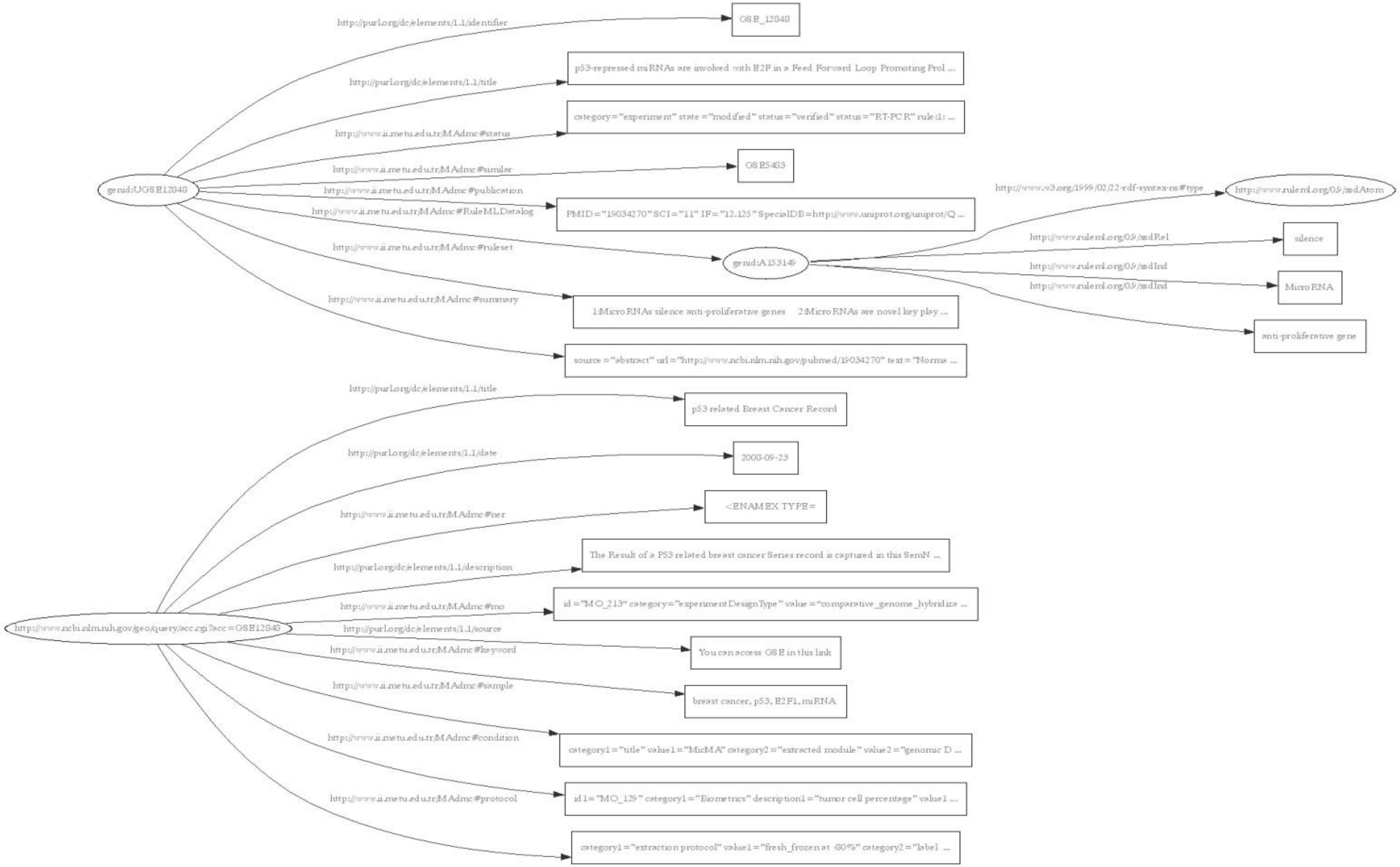
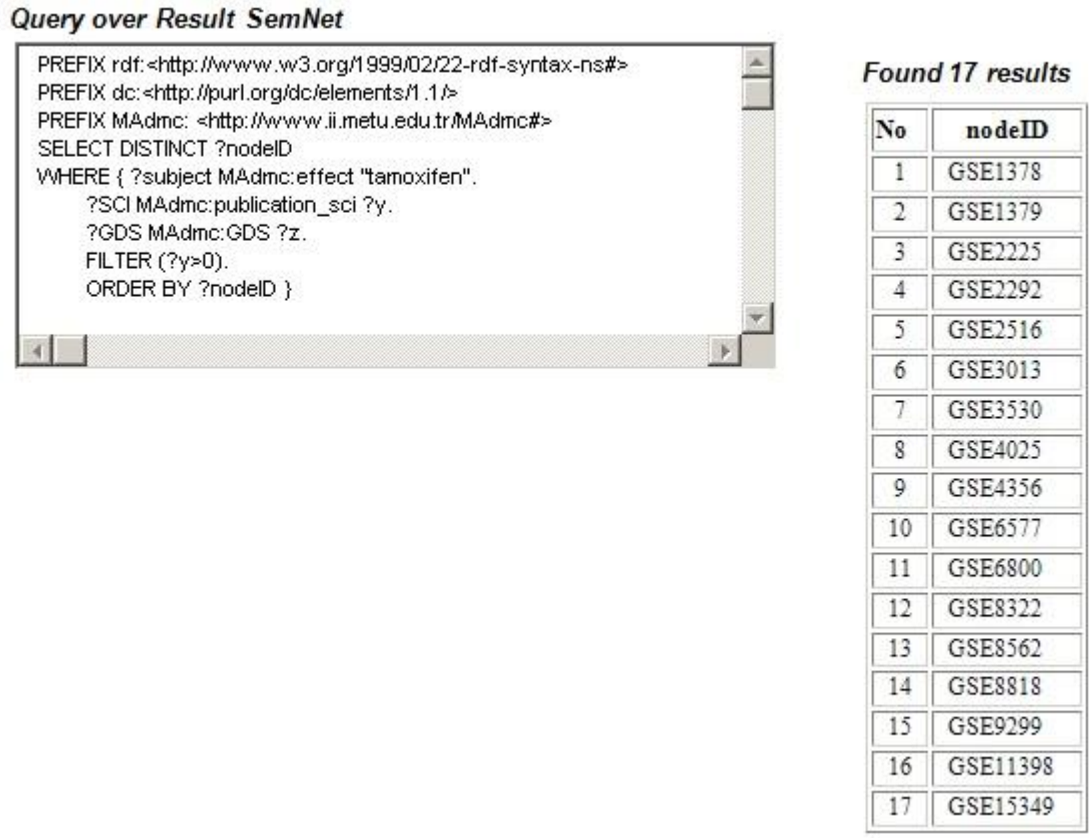

http://purl.org/dc/elements/1.1/identifier
GSE_12848
http://purl.org/dc/elements/1.1/title
p53-repressed miRNAs are involved with E2F in a Feed Forward Loop Promoting Prol ...
category="experiment" state="modified" status="verified" status="RT-PCR" rule:1: ...
http://www.ii.metu.edu.tr/MAdmc#status
http://www.ii.metu.edu.tr/MAdmc#similar
GSE5483
genid:UGSE12848
http://www.ii.metu.edu.tr/MAdmc#publication
http://www.ii.metu.edu.tr/MAdmc#RuleMLDatalog
PMID="19034270" SCI="11" IF="12.125" SpecialDB=http://www.uniprot.org/uniprot/Q ...
http://www.ii.metu.edu.tr/MAdmc#ruleset
genid:A153149
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://www.ruleml.org/0.9/xsdAtom
http://www.ruleml.org/0.9/xsdRel
silence
http://www.ruleml.org/0.9/xsdInd
http://www.ruleml.org/0.9/xsdInd
MicroRNA
anti-proliferative gene
http://www.ii.metu.edu.tr/MAdmc#summary
1:MicroRNAs silence anti-proliferative genes 2:MicroRNAs are novel key play ...
source="abstract" url="http://www.ncbi.nlm.nih.gov/pubmed/19034270" text="Norma ...
http://purl.org/dc/elements/1.1/title
p53 related Breast Cancer Record
http://purl.org/dc/elements/1.1/date
2008-09-23
http://www.ii.metu.edu.tr/MAdmc#ner
<ENAMEX TYPE=
http://purl.org/dc/elements/1.1/description
The Result of a P53 related breast cancer Series record is captured in this SemN ...
http://www.ii.metu.edu.tr/MAdmc#mo
id="MO_213" category="experimentDesignType" value="comparative_genome_hybridiza ...
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12848
http://purl.org/dc/elements/1.1/source
You can access GSE in this link
http://www.ii.metu.edu.tr/MAdmc#keyword
breast cancer, p53, E2F1, miRNA
http://www.ii.metu.edu.tr/MAdmc#sample
category1="title" value1="MicMA" category2="extracted module" value2="genomic D ...
http://www.ii.metu.edu.tr/MAdmc#condition
id1="MO_129" category1="Biometrics" description1="tumor cell percentage" value1 ...
http://www.ii.metu.edu.tr/MAdmc#protocol
category1="extraction protocol" value1="fresh_frozen at -80%" category2="label ...

Figure 7 – The graph output for the SemNet in Table 11 (*P53 gene related breast cancer record*, *GSE12848*) as validated by the RDF Validator

As acceptance and experience grow, the encoding may be more sophisticated. The Result SemNet file can be validated against “datalog.xsd” and W3C RDF validation tool. There are successful implementations on this topic such as jDREW (*A Java Deductive Reasoning Engine for the Web*, http://www.jdrew.org/) (81) with RuleML and SPARQL functionality. There are easy to follow tutorials and applications on public-domain about how to encode free text with RuleML Datalog. By doing so, we not only encode and represent the free-text result section but also open the way for further triggering chaining derivations from an already stored rule base. In fact, this is the job of a rule-based system. We here only demonstrate the capability. Rules can extend the OWL as included in the Semantic Web architecture. In that regard, for example SWRL combines RuleML (*Horn-like rules*) with OWL (*axioms*). And, the RIF mechanism allows different representations to be grouped for further use.

The metadata card and the SemNets can be queried by using the SPARQL routines (*SPARQL functionality in RDF Editor-Rhodonite or the online SPARQL tool at http://www.sparql.org/query.html*). For example, the query file *(to find the GEO Series records that inquiry Tamoxifen effect among 677 breast cancer records*) in Figure 8 can then be attached to the related SemNet file.



Query over Result SemNet

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX MAdmc: <http://www.ii.metu.edu.tr/MAdmc#>
SELECT DISTINCT ?nodeID
WHERE { ?subject MAdmc:effect "tamoxifen".
      ?SCI MAdmc:publication_sci ?y.
      ?GDS MAdmc:GDS ?z.
      FILTER (?y>0).
      ORDER BY ?nodeID }
```

Found 17 results

No	nodeID
1	GSE1378
2	GSE1379
3	GSE2225
4	GSE2292
5	GSE2516
6	GSE3013
7	GSE3530
8	GSE4025
9	GSE4356
10	GSE6577
11	GSE6800
12	GSE8322
13	GSE8562
14	GSE8818
15	GSE9299
16	GSE11398
17	GSE15349

Figure 8 - A Sample SPARQL query on Result SemNet (Online "SPARQLer RDF Query Tool" used at http://www.sparql.org/query.html)

2.7 The Performance Analysis of the Case Study

The storage medium, optimization and representation can affect a healthy comparison besides the characteristics of the data set. Query processing depends on the logical and physical organization of the data. For example if data is dynamic, interrelated, queries are unpredictable, relationships are many-to-many, or inference is needed then RDF representation performs better. And, if data is highly regular, queries access most part of the row, and the relationships are many-to-one then relational representation is more efficient. Schema for GEO is fixed whereas RDF stores are flexible for arbitrarily structured, changing, and disparate data. With these characteristics, semantically interconnected RDF datasets can form a global knowledgebase. In order not to compare apples with oranges, we made the data quite similar (semantically equal) and created a generic query without favoring any platforms. We created a table for each class and a column for each property. We wanted to see the performances when a researcher queries GEO and our RDF dataset on their representations.

We loaded 677 downloaded GSE (breast cancer) records within its 3-record relational design (*GEO design is relational*) in Oracle 11g database. We also loaded 677 RDF graphs (SemNets) which have similar structure to GSE records to Oracle 11g RDF store. We generated test data from this core dataset. The SPARQL query over RDF store and SQL query over relational database have been run for varying size of data sets in order to obtain a consolidated result. The test environment was:

- Neither store, load and inference performances nor space efficiency are measured
- No query optimizations such as materialized joins, compressed indexes are done for both platforms
- SELECT query is used
- Time performance (*computational complexity*) of the selected query set is searched, query run times are compared
- Performance metrics has been CPU time for running a query
- The hardware was: a medium sized PC with 1 CPU (Intel i3, 2.3 Ghz, 4 GB DDR3 RAM)
- The software was: Windows 7 professional 64 bit OS, Oracle 11g Spatial DB with RDF store
- OS caches are cleared, DB buffer and cache sizes in configuration file are defaulted

We obtained comparable “mili second” level computational times as shown in Table 13.

Table 13 – Time performance comparison between SPARQL querying over RDF representation and SQL querying on relational representation of semantically equivalent data set.

	40K[*] triples	**400K triples**	**1 M[**] triples**
Oracle 11g RDF Store	0.28 ms[§]/3560 QpS[§§]/	1 ms/940 QpS	2.5 ms/414 QpS
Oracle 11g Relational Database	0.25 ms/4019 QpS	0.99 ms/1009 QpS	2.6 ms/380 QpS

[*]K is 1000; [**]M is million; [§]ms is mili second; [§§]QpS is query per second.

Below 1 million RDF triples the performances are proved to be comparable in terms of time complexity between similar relational and RDF representations. Everything being equal, when the size gets larger up to billion then relational representation outperforms (82). In our case, there are 30,000 GSE records and we have at most 50 RDF statements in each SemNet which makes around 1M in total.

There is a linear correlation between the complexity of the data set and the computational time for RDF datasets (83). In our case, the metadata that is represented in a SemNet is dynamic, interrelated and complex (*hold many to many relationships*). Such a dataset is better suited to graph structure. Note also that we included hidden and previously unavailable data as well as complex data within RDF representation. Graph data structure has long been studied for computational complexity and there are efficient graph algorithms available.

The boolean expressions of GEO online analysis tool offer limited query patterns when compared with complex patterns available in RDF dataset. And, RDF store can be realigned for its structure to match frequently asked queries. And, once indexed within RDF store, an RDF dataset is computationally deterministic since each node in the forest of trees is visited once. Thus, SPARQL is guaranteed to find all requested instances. It is also common to have in-memory applications for RDF data to boost the time performance. Regarding performance of RuleML Datalog: Datalog is sound (*all statements are true*) and complete (*all true statements are provable*).

We saw that for RDF triples around a million, SELECT queries are comparable in terms of time performance between RDF and relational representations. Considering complex, dynamic, disparate nature of SemNet data, we concluded that RDF representation is the good choice. We believe that a separate study should be made to analyze their time and space consumption for common use cases.

CHAPTER 3

RESULTS AND DISCUSSION

There is exponentially rising volume of microarray data. The challenge is if an automated structured representation of microarray information is possible.

The summary part of the records on microarray repositories and related publications are not synchronized, unstructured, and in free-text format. There are statements that are incomplete and ambiguous, thus not easily comparable with others in similar studies. The results should be visible, understandable, and usable throughout their lifecycles. This is an information management principle. Once we structure (*MAdmc*) and encode the contextual data (*SemNet*), not only certain operations such as discovery and exchange become feasible, but also hidden and previously unavailable facts may be extracted from such structured and encoded data sets. This process is studied by rule encoding and processing initiatives such as RIF and SWRL. The structured entry paradigm can also be enforced besides annotation by ontology within the SemNet while encoding is implemented.

If one searches MAdmr (*over MAdmc and SemNets*), it will be more efficient (*faster and precise*) than a search on the GEO at present. It is something like sorting before an efficient search. It is the process of linking data for which the resources-properties-relationships are identified. MAdmf brings about an overhead, but if microarray repository records are designated as the reference, the benefits from such future studies will justify this start-up cost.

We extended current MINiML file and structured it into a metadata framework by introducing microarray metadata card. We then added a semantic layer by creating semantic nets to accompany the metadata card(s). Summary and Experimenter graphs are only two sample graphs. Some other semantic nets on other parts of dataset with different encoding schemes may be used. The results are expected to feed the existing knowledgebases such as gene and pathway. The experimenters submit their findings to repositories such as GEO, GO, pathway, or post at their lab web sites. There are two aspects of this mode of operation. One is that this dispersed data should look similar and the second is that hidden statements (*facts, assumptions*) need to be derived. This is the task for a rule based system but the experiment results in microarray repository should be the basis for such initiatives. The summary information in the metadata card should contain the free text findings as well as encoded version and annotational data that in order to support further processing. Note that multiple genes have been linked to e.g. breast cancer and each may have been linked to some other diseases as well. Once the link between the experimentation (*GEO records*) and the related target databases (*GO or pathway databases*) are established and upcoming results extend (*add, negate, modify*) the existing knowledgebase, then some additional information may become visible and new experiments can target these.

MAdmc, as being microarray discovery metadata card, contains structural and semantic elements to satisfy the format requirements of microarray community. With MAdmf, we not only provide a platform for a commonly understood information space but also ensure that the producer can put more findings and the consumer can understand the conveyed meaning and can uncover some hidden knowledge.

Describing data in a structured manner can be better done in a database, but microarray information space includes several microarray repositories, experimenter web sites, publications, and specialized databases. Practically, they all cannot be stored in a database or be easily federated. But, this is unlikely and there will always be different implementations that bring about exchange and interoperability problems. Note that the metadata cards and the semantic nets can also be used in a MAGE-OM/MAGE-ML based repository.

We can say that the microarray domain includes semi-structured data that can be best managed with SemWeb technology (84). SemWeb emphasizes the use of metadata standards and connected data to support data centric operations. The proposed framework, MAdmf follows SemWeb paradigm. The microarray community should adopt data centric approach because the operations mainly are data intensive. The data management is the vehicle for data centric initiatives, and an IT system is as weak as its data management. A data layer is built separately than the business logic layer in future-proof applications. MAdmf is related to the data layer. It promotes the standardization on microarray repositories. Any modelling or application development effort can then follow its use.

In this study, we examined the structure of the existing MINiML file and introduced an extended format for a metadata card. The metadata is represented in RDF and defined in a framework which is based on Dublin Core metadata element set. We created domain-specific SemNets (*RDF graphs for FOAF nets and Summary nets*), and offered to post them to a common reference point that is an ebXML based common metadata registry which provides a shared information space (*at the GEO or at a central metadata registry*). The case study shows that our approach is efficient in information identification, and knowledge discovery. Thus, in the proposed framework

- The producer can add more structured data (*express a wide range of findings*), and the consumer can get the conveyed meaning (*What has been received is limited to what has been understood*)
- Due to the possibility for more automation, backlog is reduced in curation work (*from submitted records to GEO Series or GEO Series to GEO Datasets or GEO Datasets to Array Express records*)
- Ambiguity and redundancy is reduced with standard format and additional semantics
- Data centric approach is adopted, and the quality, potential, value, and expressiveness of data are pursued where a separate data layer from business logic is maintained

- Consumers reach data otherwise unavailable (*New entries in descriptive information and semantic layer*)
- Lifecycle management (*lifetime modification and living data set*) concept is introduced including change and version management
- Visibility (*common reference point*), understandability, and usability are enforced
- Users can use W3C and the public-domain tools to extract data
- The controlled vocabularies (*Countries, Date Time Group, Names*) are used not only to annotate but also to encode the metadata and the data
- The produced metadata card and its associated SemNet(s) are extendable, integrable, queryable, and exchangeable
- Microarray records and subsequent entries (*publication, specialized databases*) can be synchronized

The creation of MAdmc has three aspects. First, content is detailed in experiment specific details such as platform, sample as well as summary and experimenter data. Second, format is materialized through the employment of data and syntax encoding schemes. The organization and structure is improved with the introduction of layers, additional metadata framework elements and attributes. Third, the process is extended with the new concepts such as lifecycle management, metadata registry use, and structured entry. In this manner, the MINiML file has been transformed into a metadata card and its semantics is extended with SemNets. Then, they can be used in any similar data store.

The main contribution of this study is to propose a discovery metadata framework with machine interpretable structured content which could support curation and analysis requests. We state that the proposed MAdmc enhanced with RDF graphs of SemNets could be used for discovery and exchange purposes among main microarray repositories. The people, experiment, and result data are linked as the proposed framework provides such a foundation. Thus, for example, a meta-analyst can get a consolidated summary of the result part of all breast cancer data sets by using a SPARQL query. The originator, the curator, the developers and other experimenters may benefit from this framework. In this dissertation, the specification is given; some key products are presented in a case study where a proof of concept is introduced. The preliminary results of this thesis have been presented in the 11th MGED-2008 (http://mged11.fbk.eu/accepted_posters), ISMB 2009 (Bio Ontologies-

Knowledge Biology, http://bio-ontologies.man.ac.uk/2009/posters.html), ECCB10-2010 (European Conference on Computational Biology, http://bioinfo.cipf.es/ node/748) conferences and in BJMG (Balkan Journal of Medical Genetics, http://www.bjmg.edu.mk/record.asp?subrecordid=1422. "Kocabas., F., Can, T., Baykal, N. (2011). Metadata Management and Semantics in Microarray Repositories. *BJMG, 14(14)*, 49-64").

The MAGE-ML and the MINiML seem to be alternative structures but actually they are not. The MINiML is rather an intermediary data structure where a MAGE-ML application can be developed onto. The creation of MAdmc (*metadata card*) and SemNet (*semantic net*) includes two different and complementary contributions to support MINiML towards a format and exchange standard. They do not replace any existing work. However, if adopted, they can be a focus for discovery, integration, and exchange. SemNets can be created for other parts of microarray records, in addition to the experimenter and summary data. Note also that this study can easily be adapted to other microarrays or high throughput repositories.

There is up to a 3% monthly increase in records at the GEO in recent years. There is a backlog of up to 15% in Series records for varying reasons. There is also a serious backlog of 80% in Dataset transformation task (*from Series record [GSE] to Dataset record [GDS]*) performed by the GEO curators, Table 14.

Table 14 - Current Data composition as of June 1st, 2012

GEO Repository	**Public**	**Unreleased**	**Total**	**Backlog**
Platforms (GPL)	10,129	532	10,661	~5.0%
Samples (GSM)	750,320	136,272	886,592	~16.0%
Series (GSE)	30,497	5,383	35,880	~15.0%
Datasets (GDS)	2,720	-	Number of experiments (*Series records/*2)	~80.0%

This is likely to increase because the amount of data and its complexity are on the rise. Related submissions, on the other hand, are difficult to follow. Standardization studies like this one that promote machine understandability and semantic interoperability are required. This study not only brings the metadata card and the semantic net concepts within a format standard approach but also introduces the importance of the lifecycle management (*modifications, versioning, retiring*), data management and structured entry concepts. Such a study will be beneficial especially for producers, curators, future experimenters, and system developers whether they employ manual or automated means. The experimental data, encoded formats, and program can be requested from the corresponding author.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTION

Microarray informatics has been a new and active research direction especially in architectural and computational aspects. This is the post-genomics era and OMICS experimentation including microarray experiments produce valuable data and the resultant data sets are annotated and put into biological context for further analysis. The experiment data is huge and stronger representation, exchange formats and standards are needed. The conduct of microarray experimentation is only the first part of the process. The second part, which is often poorly handled, is to organize present, exchange, understand and use the interpreted experimental evidence. Thus, gaps and inconsistencies as well as ambiguities in the microarray knowledgebase such as candidate theories, scientific disagreements, and open questions can be managed and resolved.

This study addresses the main problems of microarray data representation. We demonstrated that the introduction of the metadata card could support discovery and exchange operations, and SemNet would be a vehicle to capture the meaning in the microarray domain. Since domain experts create the SemNets, previously unknown details can be revealed by allowing complex queries. The proposed framework, MAdmf does not replace but complements the existing products in the microarray domain. MAdmf can be used in microarray repositories and other third party platforms. The driving philosophy behind MAdmf comes from data management, knowledge engineering, semantic web, and structured messaging paradigms.

We believe that the proposed framework (MAdmf) deserves more investigation. It is a work dedicated to the data layer of the domain. The encodings, be it structural or semantic, are in RDF/XML representation and online W3C related tools can be used to process them. MAdmc and its SemNets can be generated by domain experts with little or no IT background. And, since the work is done at data layer, subsequent modelling and application development studies are facilitated.

We assume that once such standardization efforts become adopted, the required tools and detailed guidance will follow. Following topics need further investigation: the set up of a metadata registry and guidance for how to submit a package to the metadata registry; the life cycle management of records; structured data entry; configuration model to include states such as retired, incomplete, and complete and categories (*status in each state*) such as conflicting, derived, and verified; the synchronization mechanism among various repositories over metadata information elements. Then, an association can be built up between the rule set within MAdmf in the GEO and among other repositories such as publication, GO/ pathway database. Thus, any modifications can be identified and synchronized accordingly because the results at the GEO records can augment, verify or negate each other over time. A broader performance analysis between an RDF store and its semantic equivalent, relational store should be made to address various configurations on a benchmarking platform for load, time and space efficiency at different metrics. The configurations should include native RDF store, named graph store, system that maps relational database into RDF, and SPARQL wrapper for the data store.

REFERENCES

1. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet, 29(4)*, 365-371.

2. Brazma, A. (2009). Minimum information about a microarray experiment (MIAME)-Successes, Failures, Challenges. *TheScientificWorldJO, 9*, 420-423.

3. Taylor, C.F., Field, D., Sansone, S.A., Aerts, J., Apweiler, R., Ashburner, M., et al. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol, 26(8),* 889-896.

4. MAGE Object Model, retrieved September 12, 2011 from http://www.mged.org/Workgroups/MAGE/mage-om.html.

5. Jones, A.R. & Lister, A.L. (2010). Managing Experimental Data Using FuGE. *Methods Mol Biol, 604,* 333-343.

6. Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., et al. (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol, 3(9)*/research/0046.

7. Rayner, T.F., Rocca-Serra, P., Spellman, P.T., Causton, H.C., Farne, A., Holloway, E., et al. (2006). A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics, 7,* 489.

8. Sansone, S.A., Rocca-Serra, P., Brandizi, M., Brazma, A., Field, D., Fostel, J., et al. (2008). The first RSBI (ISA-TAB) workshop: “Can a simple format work for complex studies?”. *OMICS: A Journal of Integrative Biology, 12 (2),* 143-149.

9. Whetzel, P.L., Parkinson, H., Causton, H.C., Fan, L., Fostel, J., Fragoso, G., et al. (2006). The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics, 22(7),* 866-873.

10. Whetzel, P.L., Brinkman, R.R., Causton, H.C., Fan, L., Field, D., Fostel, J., et al. (2006). Development of FuGO: An Ontology for Functional Genomics Investigations. *OMICS: A Journal of Integrative Biology, 10(2),* 199-204.

11. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol, 25(11),* 1251-1255.

12. OBI, Ontology for Biomedical Investigations, retrieved September 12, 2011 from http://obi-ontology.org/.

13. GO, Gene Ontology, retrieved September 12, 2011 from http://www.geneontology.org/.

14. Ball, C.A., Brazma, A., Causton, H.C., Chervitz, S., Edgar, R., Hingamp, P., et al. (2004). Submission of Microarray Data to Public Repositories. *PLoS Biol, 2(9),* E317.

15. Ball, C.A., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., et al. (2004). Microarray Data Standards: An Open Letter. *Environ Health Perspect, 112(12),* A666-A667.

16. Ball, C.A., Sherlock, G., Parkinson, H., Rocca-Sera, P., Brooksbank, C., Causton, H.C., et al. (2002). Standards for microarray data. *Science, 298(5593),* 539.

17. Field, D. & Sansone, S.A. (2006). A special issue on data standards. *OMICS: A Journal of Integrative Biology, 10(2),* 84-93.

18. Ball, C.A. & Brazma, A. (2006). MGED standards: work in progress. *OMICS: A Journal of Integrative Biology, 10,* 138-144.

19. Stoeckert, C.J. Jr., Causton, H.C. & Ball, C.A. (2002). Microarray databases: standards and ontologies, *Nat Genet, (32)*, 469-472.

20. Brazma, A., Krestyaninova, M. & Sarkans, U. (2006). Standards for systems biology. *Nat Rev Genet, 7(8)*, 593-605.

21. Taylor, C.F. (2007). Progress in standards for reporting omics data. *Curr Opin Drug Discov Devel 10(3)*, 254-263.

22. Field, D., Sansone, S.A., Collis, A., Booth, T., Dukes, P., Gregurick, S.K., et al. (2009). Omics Data Sharing. *Science, 326(5950)*: 234-236.

23. Bizer, C., Heath, T. & Berners-Lee, T. (2009). Linked Data - The Story So Far. *Int J Semantic Web Inf Syst (IJSWIS) 5(3)*, 1-22.

24. Brors, B. (2005). Microarray annotation and biological information on function. *Methods Inf Med, 44(3)*, 468-472.

25. Aalai, E., Gleghorn, C., Webb, A. & Glover, S.W. (2009). Accessing public health information. *Health Info Libr J, 26(1)*, 56-62.

26. Schober, D., Leser, U., Zenke, M. & Reich, J. (2005). GandrKB-ontological framework for gene annotation. *Structural Bioinformatics*, 2785-2786.

27. Chen, R., Li, L. & Butte, A.J. (2007). AILUN: Reannotating gene expression data automatically. *Nat Methods, 4(11)*, 879.

28. Pasquier, C., Girardot, F., Jevardat de Fombelle, K. & Christen, R. (2004). THEA: Ontology driven analysis of microarray data. *Bioinformatics 20(16)*, 2636-2643.

29. a. Edgar, R., Domrachev, M. & Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res, 30(1),* 207-210.

29. b. Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., et al. (2005). NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res, 33,* D562-D566.

29. c. Barrett, T. & Edgar, R. (2006). Gene Expression Omnibus (GEO): Microarray data storage, submission, retrieval, and analysis. *Methods Enzymol, 411,* 352-369.

29. d. Barrett, T. & Edgar, R. (2006). Mining Microarray Data at NCBI's Gene Expression Omnibus (GEO). *Methods Mol Biol, 338,* 175-190.

29. e. Edgar, R. & Barrett, T. (2006). NCBI GEO standards and services for microarray data. *Nat Biotechnol, 24(12)*, 1471-1472.

29. f. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2007). NCBI GEO: mining tens of millions of expression profiles—database and tools update (MINiML). *Nucleic Acids Res, 35(suppl 1),* D760-D765.

29. g. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2009). NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res, 37(suppl 1),* D885-D890.

29. h. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., et al. (2011). NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res, 39,* D1005-D1010.

30. a. Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., et al. (2007). ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res, 35,* D747-D750.

30. b.Parkinson, H., Kapushesky, M., Kolesnikov, N., Rustici, G., Shojatalab, M., Abeygunawardena, N., et al. (2009). ArrayExpress update-from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res, 37(suppl 1),* D868-D872.

31. a. Ikeo, K., Ishi-i, J., Tamura, T., Gojobori, T. & Tateno, Y. (2003). CIBEX: center for information biology gene expression database. *C R Biol, 326(10-11),* 1079-1082.

31. b. Sugawara, H., Ogasawara, O., Okubo, K., Gojobori, T. & Tateno, Y. (2008). DDBJ with new system and face. *Nucleic Acids Res, 36 (suppl 1),* D22-D24.

32. Bioconductor (open source software for bioinformatics), retrieved September 12, 2011 from http://www.bioconductor.org.

33. Zhu, Y., Davis, S., Stephens, R., Meltzer, P.S. & Chen, Y. (2008). GEOmetadb: powerful alternative search engine for the GEO. *Bioinformatics, 24(23),* 2798-2800.

34. Davis, S. & Meltzer, P.S. (2007). GEOquery: a bridge between the GEO and BioConductor. *Bioinformatics, 23(14),* 1846-1847.

35. Neuweger, H., Albaum, S.P., Dondrup, M., Persicke, M., Watt, T., Niehaus, K., et al. (2008). MeltDB: a software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics, 24(23),* 2726-2732.

36. Bekel, T., Henckel, K., Küster, H., Meyer, F., Mittard-Runte, V., Neuweger, H., et al. (2009). The Sequence Analysis and Management System – SAMS-2.0: Data management and sequence analysis adapted to changing requirements from traditional sanger sequencing to ultrafast sequencing technologies. *J of Biotechnol, 140(1-2),* 3-12.

37. Dondrup, M., Albaum, S.P., Griebel, T., Henckel, K., Jünemann, S., Kahlke, T., et al. (2009). EMMA 2: a MAGE-compliant system for the collaborative analysis and integration of microarray data. *BMC Bioinformatics, 10,* 50.

38. Maurer, M., Molidor, R., Sturn, A., Hartler, J., Hackl, H., Stocker, G., et al. (2005). MARS: microarray analysis, retrieval, and storage system. *BMC Bioinformatics, 6,* 101.

39. Te Pas, M.F., Hulsegge, I., Coster, A., Pool, M.H., Heuven, H.H. & Janss, L.L. (2007). Biochemical pathways analysis of microarray results: regulation of myogenesis in pigs. *BMC Dev Biol, 7,* 66.

40. Rainer, J., Sanchez-Cabo, F., Stocker, G., Sturn, A. & Trajanoski, Z. (2006). CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res, 34(Web Server)***:** W498-W503.

41. Yang, Y.H., Paquet, A. & Dudoit, S. (2007). marray: Exploratory analysis for two-color spotted microarray data (Bioconductor package), retrieved September 12, 2011 from http://www.bioconductor.org/packages/2.5/bioc/html/marray.html.

42. Kupershmidt, I., Su, Q.J., Grewal, A., Sundaresh, S., Halperin, I., Flynn, J., et al. (2010). Ontology-based meta-analysis of global collections of high-throughput public data. *PloS on 5(9),* e13066.

43. Larsson, O. & Sandberg, R. (2006). Lack of correct data format and comparability limits future integrative microarray research. *Nat Biotechnol 24,* 1322-1323.

44. The Reporting Structure for Biological Investigation (RSBI) working group, retrieved September 12, 2011 from http://www.mged.org/Workgroups/rsbi/rsbidetail.html.

45. Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C, et al. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol, 24(9),* 1151-1161.

46. Lors, R.K., van Ginneken, A.M. & van der Lei, J. (2005). OpenSDE: a strategy for expressive and flexible structured data entry. *Int J Med Inform, 74(6),* 481-490.

47. PATIKA, Pathway Analysis Tool for Integration and Knowledge Acquisition, retrieved September 12, 2011 from http://www.patika.org.

48. REACTOME, a curated knowledgebase of biological pathways, retrieved September 12, 2011 from http:// www.reactome.org.

49. Brazma, A. (2001). On The Importance of Standardisation in Life Sciences. *Bioinformatics, 17(2),* 113-114.

50. Spellman, P.T. (2005). Status Report On MAGE. *Bioinformatics, 21(17),* 3459-3460.

51. Quackenbush, J. (2006). Standardizing the standards. *Mol Syst Biol, 2*: 2006.0010.

52. Wang, X., Gorlitsky, R. & Almeida, J.S. (2005). From XML to RDF: how semantic web technologies will change the design of ‘omic' standards. *Nat Biotechnol 23:* 1099–1103.

53. Sarkans, U., Parkinson, H., Lara, G.G., Oezcimen, A., Sharma, A., Abeygunawardena, N., et al. (2005). The ArrayExpress gene expression database: a software engineering and implementation perspective. *Bioinformatics 21(8),* 1495-1501.

54. Seibel, P.N., Krüger, J., Hartmeier, S., Schwarzer, K., Löwenthal, K., Mersch, H., et al. (2006). XML schemas for common bioinformatic data types and their application in workflow systems. *BMC Bioinformatics,* 7:490.

55. Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., et al. (2003). ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res 31(1),* 68–71.

56. Ontology-Based Annotator Web Service. C. Jonquet, M. A. Musen, & N. H. Shah. Technical Report, Published 2008, http://www.bioontology.org/annotator-service.

57. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine Fall 1996*, 37-54.

58. Tan, P-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining, *Pearson Addison Wesley (May, 2005)*, ISBN: 0321321367.

59. Kumar, A.V.S. (2012). Advanced data mining techniques. *Session paper, The Second International Conference on Digital Information and Communication Technology and its Applications (DICTAP2012)*, http://www.sdiwc.net/thi/files/session_proposal.doc.

60. Frawley, W.J., Piatetsky-Shapiro, G., & Matheus, C.J. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine, 13(3)*, 57-70.

61. Pyle, D. (1999). Data Preparation for Data Mining Morgan. *Kaufmann Publishers*, ISBN 1-55860-529-0.

62. Jiang, D., Pei, J., Ramanathan, M., Lin, C., Tang, C., & Zhang, A. (2006). Mining gene–sample–time microarray data: a coherent gene cluster discovery approach. *Knowl Inf Syst 13*, 305-335.

63. Gottlieb, B., Lorraine E. Chalifour, L.E., Mitmaker, B., Sheiner, N., Obrand, D., et al. M. (2009). BAK1 gene variation and abdominal aortic aneurysms. *Hum Mutat, 30(7)*, 1043–1047.

64. Branca, M.A., Goodman, N., and Venkatesh, T.V. (2001). Bioinformatics: Getting Results in the Era of High-Throughput Genomics". *Cambridge Healthtech Institute Report 9*, May 2001.

65. Parkinson, H. EBI Presentation, at http://www.ict.ox.ac.uk/odit/projects/digitalrepository/docs/workshop/Helen_Parkinson-RDMW0608.ppt).

66. Pham, T.D., Wells, C., & Crane, D.I. (2006). Analysis of Microarray Gene Expression Data. *Curr Bioinform, 1*, 37-53.

67. Hernandez-Boussard, T., Woon, M., Klein, T.E. & Altman, R. (2006). Integrating Large-Scale Genotype and Phenotype Data. *OMICS: A Journal of Integrative Biology, 10(4),* 545-554.

68. Jurisica, I. (2005). Knowledge Discovery in High-Throughput Biological Domains: Introduction to Computational Biology. *Tutorial presentation at The Tenth International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC)* , University of Regina , Regina, SA, Canada, September 1, 2005.

69. Mirnics, K., Levitt P, & Lewis, D.A. (2006). Critical appraisal of DNA microarrays in psychiatric genomics. *Biol Psychiatry, 15;60(2)*:163-176.

70. Krishnamurthy, R., Naughton, J.F., & Shanmugasundaram, J. (2001). Dealing with (un)structuredness in XML data and quearies using relational databases. *DB seminar at Wise university*, USA.

71. Crampton, D. (2004). How to deal with structured and unstructured data?, *Techworld 2004*.

72. SPARQLer, an online RDF Query platform on the public domain, retrieved September 12, 2011 from http://www.sparql.org/query.html.

73. Miller, H., Norton, C.N. & Sarkar, I.N. (2009). Genbank and PubMed: How connected are they? *BMC Res Notes, 2,* 101.

74. Duewer, D.L., Jones, W.D., Reid, L.H. & Salit, M. (2009). Learning from microarray interlaboratory studies: measures of precision for gene expression. *BMC Genomics, 10*, 153.

75. Garde, S., Chen, R., Leslie, H., Beale, T., McNicoll, I. & Heard, S. (2009). Archetype-based knowledge management for semantic interoperability of electronic health records. *Stud Health Technol Inform, 150,* 1007-1011.

76. Dogac, A., Gokce, B. Laleci, B.G., Kabak, Y., Unal, S., Heard, S., et al. (2006). Exploiting ebXML registry semantic constructs for handling archetype metadata in healthcare informatics. *IJMSO, 1(1)*, 21-36.

77. ADatP-3(A) - NATO Message Text Formatting System (FORMETS) Concept of FORMETS (CONFORMETS), (can be requested from corresponding author).

78. Esposito M. (2007). An Ontological and Non-monotonic Rule-Based Approach to Label Medical Images. Proceedings of the *Third International IEEE (Institute of Electrical and Electronics Engineers) Conference on Signal-Image Technologies and Internet-Based System (SITIS),* Shanghai, China, 16-18 December 2007;pp 603-611.

79. Schober, D., Leser, U., Zenke, M. & Reich, J. (2005). GandrKB-ontological framework for gene annotation. *Structural Bioinformatics*, 2785-2786.

80. Chen, R., Li, L., & Butte, A.J. (2007). AILUN: Reannotating gene expression data automatically. *Nat Methods*, *4(11)*, 879.

81. jDREW, A Java Deductive Reasoning Engine for the Web (SPARQL, RuleML support) , retrieved September 12, 2011 from http://www.jdrew.org/.

82. Berlin SPARQL Benchmark Platform, retrieved August 28, 2012 from http://www4.wiwiss.fu-berlin.de/bizer/BerlinSPARQLBenchmark/.

83. RDF Benchmarking for toxicology semantic data, retrieved August 28, 2012 from http://opentox.ntua.gr/wiki/RDF_benchmarking.

84. Angles, R. & Gutierrez, C. (2005). Querying RDF Data from a Graph Database Perspective. *Lect Notes in Comput Sc, Volume 3532*, 93-107.

APPENDICES

A: A BRIEF INTRODUCTION TO GEO AND MINiML

Overview

GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/) is a public functional genomics data repository supporting MIAME-compliant data submissions. GEO is the largest microarray repository. There is a data exchange file (MINiML) for each microarray experiment at GEO where its definition is published in an XML schema. We have chosen GEO as subject repository in our study. Our study proposes a metadata card and semantic net in place of MINiML file in a metadata framework architecture. Thus, the structure and semantics can be aligned with the needs of the producer and the consumer.

GEO freely distributes high-throughput OMICS data submitted by the scientific community. Besides microarray, other high throughput technologies are: SNP (Single Nucleotide Polymorphism) Arrays, SAGE (Serial Analysis of Gene Expression), MPSS (Massively Parallel Signature Sequencing), Next-generation sequence data, Protein arrays, and Mass Spectrometry data. The journals require accession numbers for microarray data before acceptance of a paper for publication. Thus, data should be deposited in a repository like GEO before its paper is submitted to a journal for publication. The benefits for submitter are long term archiving, integration with other tools and inclusion of link to the project web site. Note that MIAME compliance is determined by the content provided, not by the submission format. Also note that journal publication is not a requirement for data submission to GEO. The submitter may update his/her submission at a later date. However, only GEO staff can make deletions (http://www.ncbi.nlm.nih.gov/geo/info/faq.html).

GEO Architecture

The submitters load their gene expression data in four sections. The first three also reflect the basic record types (*Platform (e.g., array), Sample (e.g., hybridization), and Series (experiment)*) within the primary database. They are assigned a unique GEO accession number during the submission. The fourth section of the submission is supplementary data which are original microarray scan images and raw quantification data. The basic record types within GEO, as depicted in Figure 9, are listed subsequently.

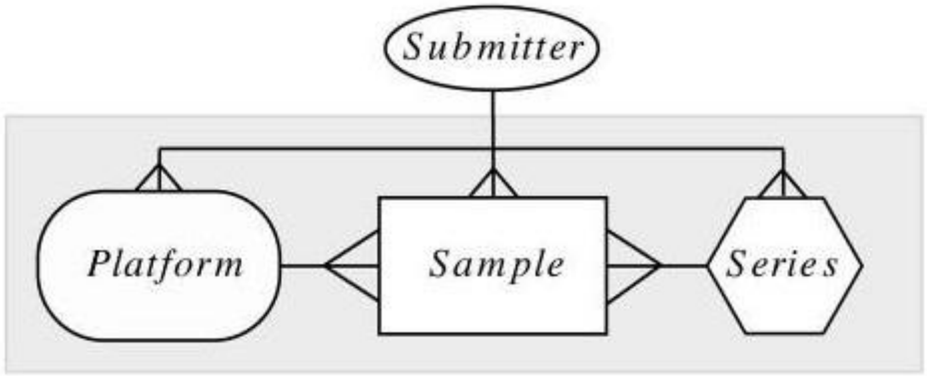
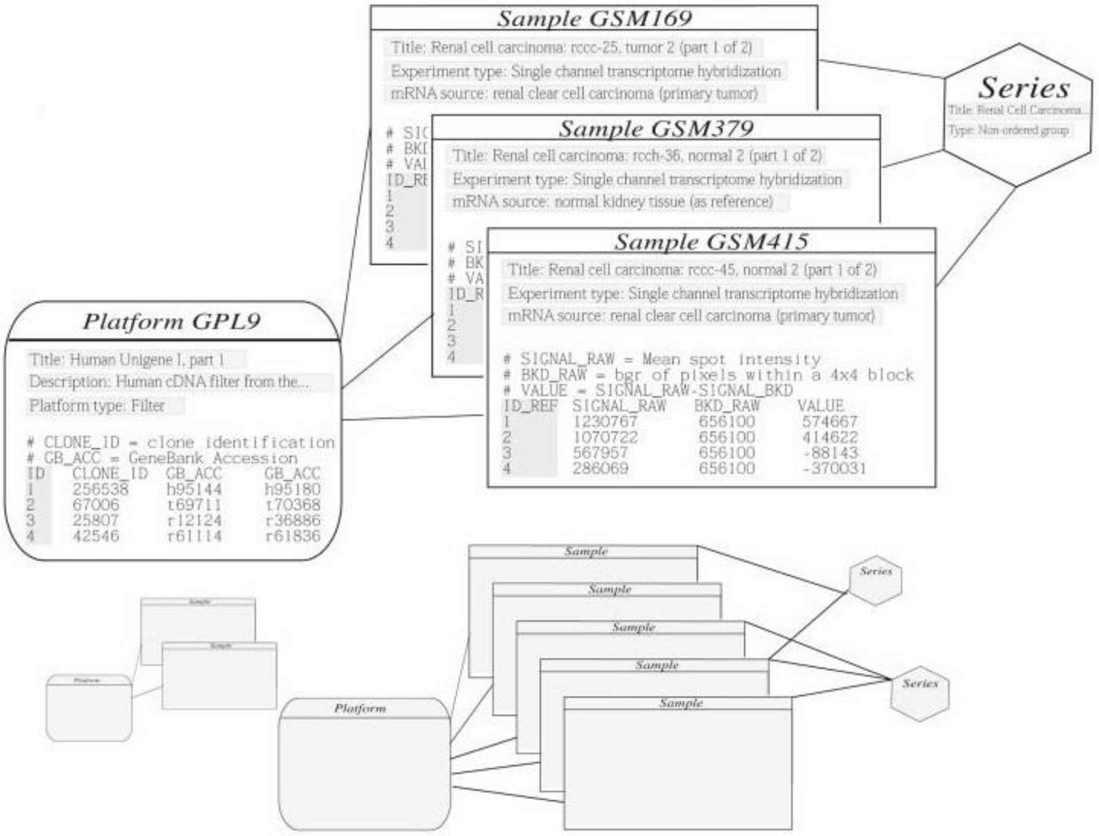


Figure 9.a - GEO schema (*The entity-relationship diagram for the GEO database to show the relationship between submitter, platform, sample, and series records*)


Sample GSM169
Title: Renal cell carcinoma: rccc-25, tumor 2 (part 1 of 2)
Experiment type: Single channel transcriptome hybridization
mRNA source: renal clear cell carcinoma (primary tumor)
Sample GSM379
Title: Renal cell carcinoma: rcch-36, normal 2 (part 1 of 2)
Experiment type: Single channel transcriptome hybridization
mRNA source: normal kidney tissue (as reference)
Sample GSM415
Title: Renal cell carcinoma: rccc-45, normal 2 (part 1 of 2)
Experiment type: Single channel transcriptome hybridization
mRNA source: renal clear cell carcinoma (primary tumor)
Series
Title: Renal Cell Carcinoma...
Type: Non-ordered group
Platform GPL9
Title: Human Unigene I, part 1
Description: Human cDNA filter from the...
Platform type: Filter
Sample
Platform
Series

SIGNAL_RAW = Mean spot intensity
BKD_RAW = bgr of pixels within a 4x4 block
VALUE = SIGNAL_RAW-SIGNAL_BKD

ID_REF	SIGNAL_RAW	BKD_RAW	VALUE
1	1230767	656100	574667
2	1070722	656100	414622
3	567957	656100	-88143
4	286069	656100	-370031

CLONE_ID = clone identification
GB_ACC = GeneBank Accession

ID	CLONE_ID	GB_ACC	GB_ACC
1	256538	h95144	h95180
2	67006	t69711	t70368
3	25807	r12124	r36886
4	42546	r61114	r61836

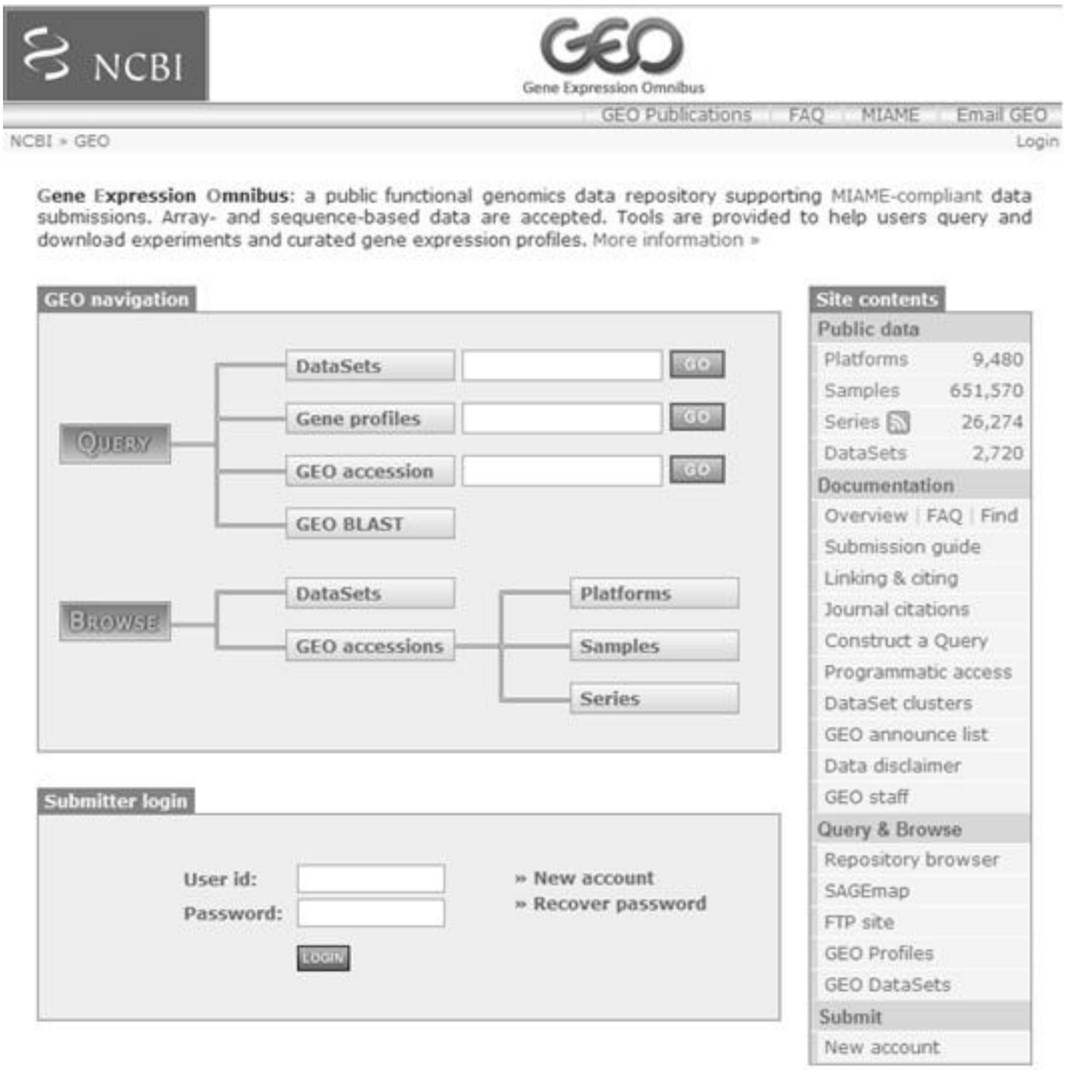
Figure 9.b. - An actual example of three samples referencing one platform and contained in a single series (*taken from 29.a*)

Figure 9 – GEO Model

Platform: It is the text description of the array. A Platform may reference many Samples that have been submitted by multiple submitters. Sample: It is the text description of a biological sample. Original raw data is contained here. A Sample entity must reference only one Platform and may be included in multiple Series entities. Series: A Series record links a group of related Samples that are a part of the experiment and provides the description of whole study. Series records may also contain tables describing extracted data, summary conclusions, or analyses (http://www.ncbi.nlm.nih.gov/geo/info/overview.html). Selected data sets are assembled into biologically and statistically meaningful and comparable GEO DataSets (GDSxxx) by GEO curators. DataSet records provide a coherent summary about an experiment and form the basis of GEO's analysis tools

]

(http://www.ncbi.nlm.nih.gov/geo/info/GEOHandoutFinal.pdf). Samples within a DataSet refer to the same platform. There is considerable amount of backlog in DataSet creation, so not all Series have corresponding DataSet record(s). Both Series and DataSets are searchable using the GEO DataSets interface. However, only DataSets are processed by GEO's analysis tools such as gene expression profile chart data set clusters, Figure 9. A Profile consists of the expression measurements for an individual gene across all Samples in a DataSet (http://www.ncbi.nlm.nih.gov/geo/info/overview.html).



Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »

Public data	
Platforms	9,480
Samples	651,570
Series	26,274
DataSets	2,720

Figure 10 - GEO website (http://www.ncbi.nlm.nih.gov/geo/) on the 16th of November 2011

The query workflow is given in Figure 11.

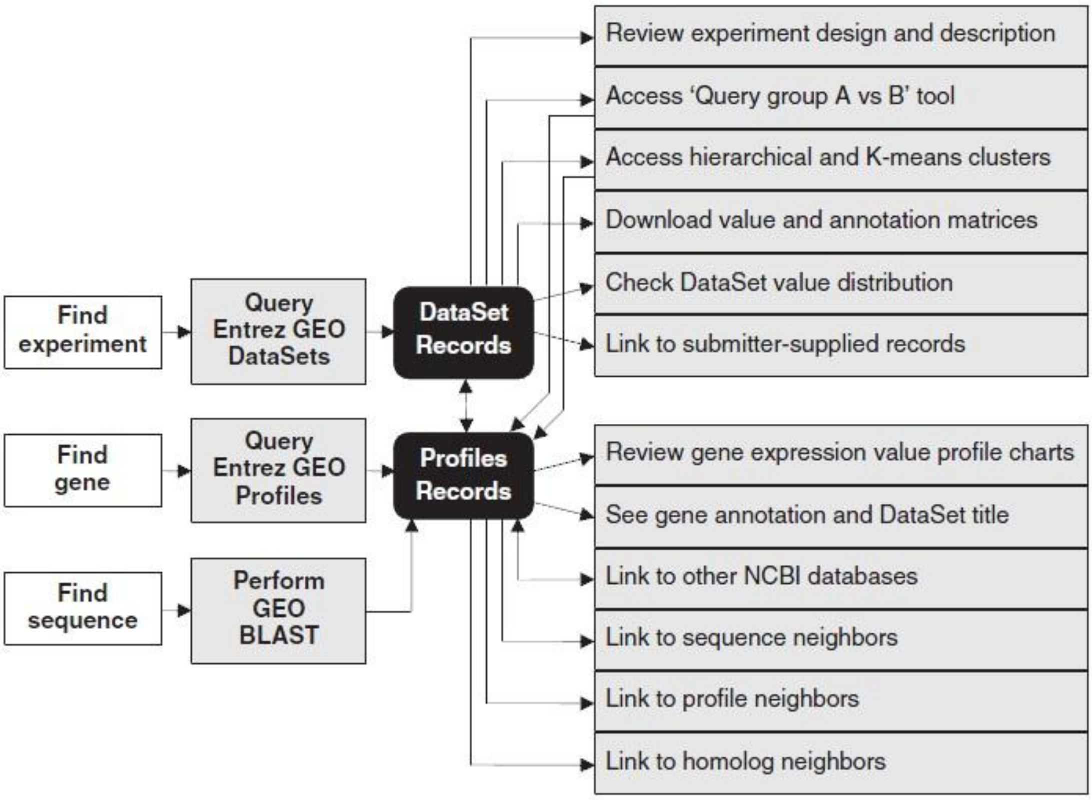

Find experiment
Query Entrez GEO DataSets
DataSet Records
Find gene
Query Entrez GEO Profiles
Profiles Records
Find sequence
Perform GEO BLAST
Review experiment design and description
Access 'Query group A vs B' tool
Access hierarchical and K-means clusters
Download value and annotation matrices
Check DataSet value distribution
Link to submitter-supplied records
Review gene expression value profile charts
See gene annotation and DataSet title
Link to other NCBI databases
Link to sequence neighbors
Link to profile neighbors
Link to homolog neighbors

Figure 11- Schematic overview of query workflow in GEO repository (*taken from 29.f*)

The Retrieval of GEO Data

GEO records (*original GEO records and curated DataSets*) can be viewed and downloaded in several ways as listed subsequently:

- **GEO Navigation:** Query wizard allow to build a query to access required DataSet information and Browse wizard allows to access DataSets and submitted platform, sample, and series records (http://www.ncbi.nlm.nih.gov/geo/).
- **FTP download:** All GEO records and raw data files are freely available for bulk download from the **FTP site** (ftp://ftp.ncbi.nih.gov/pub/geo/).
- **Links on the Series/DataSet records:** Links to experiment family downloads in various formats and supplementary files are provided.

- **Accession Display Bar:** It is found at the top of each GEO record and can be used to download or view complete or partial records, or related Platform, Sample and Series records by entering a valid GEO accession number. The records can be displayed in HTML, plain text or MINiML (XML) format. An alternative way to this is to construct a URL to retrieve data (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi).
- **Programmatic access:** GEO metadata can be programmatically accessed and retrieved using a suite of programs called the Entrez Programming Utilities (E-Utils) (http://www.ncbi.nlm.nih.gov/geo/info/geo_paccess.html). BioConductor users may use GEOquery package which parses GEO SOFT files for integration with BioConductor 'R' analysis resources.
- **Entrez GEO DataSets query downloads:** All original records can be searched and retrieved using the Entrez GEO DataSets interface. As with other NCBI Entrez databases (*e.g., PubMed*) a simple boolean phrase may be entered and restricted to any number of fields, enabling effective query (http://www.ncbi.nlm.nih.gov/gds/).

We believe that these existing browsing and searching functionality will be extended with the introduction of metadata framework to include more meaningful inquiries over metadata cards and semantic nets.

Submissions

There are four ways in which data may be deposited with GEO. Web form deposit: Interactive web forms. Spreadsheets: Excel spreadsheet templates for batch deposit. SOFT (Simple Omnibus Format in Text): A plain text format designed for batch submission. MINiML: An XML format designed for batch submission. Deciding which method to use depends on the amount of data one has to submit, the format in which the data currently exist, and the applications used. No matter what the deposit method, final GEO records will look the same and contain equivalent information.

For almost all data submissions, the experimenter will be asked to provide the following five step information (Table 15) in line with MIAME guidelines.

Table 15 - The elements of GEO submission

1	**Text description of the array and text tab-delimited table of the array template**	Platform record. Platform submission is not necessary if the array is already in GEO. In this case, the experimenter needs to reference the Platform accession number (GPLxxx).
2	**Text description of the biological sample**	Sample record. It includes sample and protocol data.
3	**Text tab-delimited table of processed hybridization result**	Sample record. For example a CHP file for Affymetrix data.
4	**Processed sequence data**	Sample record. External raw data files, e.g., Affymetrix quantification files or GenePix scan image files.
5	**Text description of the overall experiment**	Series record. It also includes TAR archive of original raw data.

Some Statistics on GEO

There are 58 RT-PCR experiments to be used in validation for microarray experiments in GEO as of 14 December 2011. There are 13 different types of platforms over 100 different organisms; 5 different samples (*RNA*, *Genomic, Mixed, SAGE*, *Protein*). The number of experiments almost is doubled every two year since its inception in 2001. 50% of all Series records have been submitted in the last 2 years. If we consider that the first submission was made in 2001, there is a serious increase in the number and volume of the submissions, Table 16.

Table 16 – Number of experiments until 2011

Year	No of Series records	Increase in No	Increase in %
2001	13	-	
2002	105	92	-
2003	628	523	135
2004	1,475	847	94
2005	2,868	1,393	63
2006	4,682	1,814	58
2007	7,393	2,711	44
2008	10,672	3,279	40
2009	14,986	4,314	37
2010	20,561	5,575	37
2011	27,000	6,439	32

We see that the numbers of the records in GEO repository have doubled since December 2008 in three years time, Table 17. However, the percentage of backlogs has not changed since then.

Table 17– GEO records in 2008 and 2011

GEO Repository	**Public**	**Unreleased**	**Total**	**Backlog**
Platforms (GPL**)**	5,312	536	5,848	~9.0%
Samples (GSM**)**	266,690	57,608	324,298	~18.0%
Series (GSE**)**	10,396	2,105	12,501	~17.0%

17.a - GEO records in December 2008

GEO Repository	**Public**	**Unreleased**	**Total**	**Backlog**
Platforms (GPL)	9,516	550	10,066	~6.0%
Samples (GSM)	656,855	135,777	792,632	~18.0%
Series (GSE)	26,464	5,142	31,606	~16.0%

17.b - GEO records in December 2011

At present, there are around 15000 publications that cite deposit of data in GEO and more than 1200 publications that cite GEO data as evidence to support independent studies as being third party usage citation.

MINiML (MIAME Notation in Markup Language,
(http://www.ncbi.nlm.nih.gov/geo/info/MINiML.html**)**

MINiML is a data exchange format designed for microarray data and high-throughput experiment results. It captures all components of the MIAME checklist, as well as any additional information that the submitter wants to provide. GEO accepts data submissions and retrievals in MINiML format. MAGE-ML (Microarray Gene Expression - Markup Language) is another data exchange format. However, MINiML has a stand alone XML Schema definition and MAGE-ML uses DTD (Data Type Definition) syntax generated automatically from an object model (MAGE-OM). MAGE-ML can structure data in a variety of ways and is mostly suitable when using the MAGE-OM as object model in an underlying database. MINiML can be seen as a metadata card for the experiment. However, it is not in line with metadata standards and its semantics is not accessible. This is where our study in this dissertation addresses.

MINiML Elements and Their Content

Table 18 displays guidelines and constraints for the content of MINiML elements which pertain to Series data (*experiment metadata*). Note that there are also platform and sample parts of the MINiML file. The reader should refer to its web site and the schema definition, for more information (http://www.ncbi.nlm.nih.gov/geo/info/MINiML.html).

Table 18 – MINiML content (taken from http://www.ncbi.nlm.nih.gov/geo/info/MINiML.html)

<Series>			
Element name	**Occurrence**	**Allowed values**	**Content guidelines for submitters**
Title	required	string of length 1-120 characters that must be unique	Provide a unique title that describes the overall study.
Summary	required	any	Summarize the goals and objectives of this study. The abstract from the associated publication may be suitable.
Type	required	any	Enter keyword(s) that generally describe the type of study. Examples include: time course, dose response, comparative genomic hybridization, cell type comparison, disease state analysis, stress response etc.
Overall-Design	required	any	Provide a brief description of the experimental design. Indicate how many Samples are analyzed, if replicates are included, are there control and/or reference Samples etc.
Pubmed-ID	optional and unbounded	an integer	Specify a valid PubMed identifier (PMID) that references a published article describing this study. It can be added later once the data are published.
Web-Link	optional and unbounded	valid URL	Specify a Web link that directs users to supplementary information about the study.
Contributor-Ref	optional and unbounded	any	List all people associated with this study.
Sample-Ref	required and unbounded	valid Sample identifiers	Reference the Sample that makes up this experiment.
Variable Factor Description Sample-Ref	optional and unbounded	Dose, time, tissue, gender, development stage, age, cell type, infection, metabolism, stress, temperature, specimen, disease state, protocol, genotype, species etc.	Describe the variable and repeat type(s) investigated in this study.
Repeats Factor Sample-Ref	optional and unbounded	Biological replicate, technical replicate.	

B: PROGRAMS USED IN THE CASE STUDY

Overview

The files used in the case study of this dissertation are:

- **MAdmc.xml:** It is the microarray discovery metadata card. It is an XML file as validated against its schema file.
- **MAdmc.xsd:** It is the schema file for MAdmc to contain the definitions, vocabularies, and patterns to support the metadata card.
- **Experimenter.rdf:** It is the semantic net (SemNet) in FOAF/RDF format to model the human element of microarray record.
- **Result.rdf:** It is the SemNet in RuleML Datalog/RDF format to model the result/summary section.
- **MAdmc.rq:** It is the query file in SPARQL syntax that runs on SemNets in RDF/XML.

The MAdmc.xml (*a sample metadata card*) and its schema file are given below. These files are to set an example. They are not intended to be optimal. This may be the subject of another study.

MAdmc.xml file is created by MAdmc program (*A web application program developed in MS Visual Studio including aspx packages – about: information about the program, default: the main program, help: help about the program*). Together with its schema file, it forms the metadata card. Semantic Nets (*experimenter.rdf and result.rdf*) can be created in any RDF Editor. We used Prodigé but there are several public editors. One can write RDF statements even in XML Notepad. Then, SPARQL file is created (*in SPARQler or Prodigé*) to take advantage of the content in metadata card and SemNet. The produced rq file is run against especially on SemNet to extract domain knowledge.

MAdmc.xml file

```
<MAdmc>
- <Extended_DC_Elements>
- <Security>
  <Policy Definition="In the context of information exchange, a security policy is a set of rules for protecting information against unauthorized disclosure, while maintaining authorised access, and preventing loss or unauthorized modification" Comment="MGED can publish the guidelines and the individual repositories can label the data sets">MGED</Policy>
  <Classification Definition="Security markings that indicate the sensitivity level of the information" Comment="It is RESTRICTED before being published">UNCLASSIFIED</Classification>
  <Category Definition="An indication of an additional, specific sensitivity, or a dissemination control, or an informational marking on which no automated access control is performed" Comment="It is conditional and used where applicable">RELEASABLE TO INTERNET</Category>
  </Security>
- <Resource_Description>
  <Title>-</Title>
  <Identifier Definition="An unambiguous reference to the resource within a given context" Comment="It is the attribute under element series in MINiML">GSE6710</Identifier>
  <Creator Definition="An entity primarily responsible for making the content of the resource" Comment="He is the contributor number 3 and is the first author for this study in its publication">Joachim Reischl</Creator>
  <Publisher Definition="An entity responsible for making the resource available" Comment="It is available at http://www.nature.com/jid/index.html">Journal of Investigation Dermatology. 2007 Jan;127(1):163-9. Epub 2006 Jul 20</Publisher>
  <Contributor>-</Contributor>
  <Date>-</Date>
  <Rights Definition="Information about rights held over the resource" Comment="It is iaw the publishing authority">Intellectual Property Rights belong to Journal of Investigation Dermatology</Rights>
  <Language Definition="A language of the resource" Comment="The controlled vocabulary, RFC 4646 is used">en-uk</Language>
  <Type Definition="The nature or genre of the content of the resource." Comment="For MAdmc, this element is limited to the various types covered by MAdmc, i.e. MAdmc instances, format specifications or any other document">MAdmc instance</Type>
  <Source Definition="References to assets or resources from which the tagged data asset is derived" Comment="The study is contained in this record of GEO repository">GSE6710, GEO</Source>
  <Relation Definition="A reference to a related resource" Comment="This is the GEO Accession string ‚Series GSE6710, at” http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6710”">Series GSE6710</Relation>
  </Resource_Description>
- <Format_Description>
  <Format Definition="The physical or digital manifestation of the resource" Comment="It is iaw “http://www.iana.org/assignments/media-types/text/”">Text/MIME</Format>
  <Format_Specification Definition="The source information is further detailed" Comment="The standards and proprietary formats are listed">MINiML, XSD, DC, ISO 11179, NS, FOAF, RIF</Format_Specification>
  <Version Definition="It details the the described items with versioning information." Comment="MGED can publish the guidelines for a versioning scheme so that the records can go through modifications throughout their lifetimes">1.0</Version>
  </Format_Description>
- <Content_Description>
  <Subject Definition="A topic of the content of the resource" Comment="“Psoriasis vulgaris” and the role of “Wnt pathway” are the subject of this study">Wnt5a in psoriatic plaques</Subject>
  <Description Definition="An account of the content of the resource." Comment="This is the abstract from PMID: 16858420">Psoriasis vulgaris is characterized by hyperproliferation and incomplete terminal differentiation of epidermal keratinocytes. Despite the established role of Wnt pathways in the regulation of stem cell proliferation and differentiation, they have not yet been associated with the pathophysiology of psoriasis. Biopsies from uninvolved and from lesional skin of 20 patients with plaque-type psoriasis were taken. The biopsies were used for microarray RNA expression profiling. Based on paired samples from 13 patients, 179 genes that were more than 2-fold differentially expressed in lesional skin were defined. This list included 16 genes with known or possible association to the canonical Wnt/beta-catenin or the non-canonical Wnt/Ca2+ pathway. These findings were confirmed by quantitative reverse transcription-PCR experiments. It was concluded that Wnt5a and other Wnt pathway genes are differentially expressed in psoriatic plaques. Their functional contribution to the pathophysiology of psoriasis needs to be elaborated.</Description>
  <Coverage Definition="The extent or scope of the content of this resource" Comment="The study is conducted to cover organism, homo sapiens">Homo sapiens</Coverage>
  </Content_Description>
  </Extended_DC_Elements>
</MAdmc>
```

MAdmc.xsd file

```
   <?xml version="1.0" ?>
- <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="urn:edu:metu:MAdmc:1:0" targetNamespace="urn:edu:metu:MAdmc:1:0"
      elementFormDefault="unqualified" xml:lang="en-GB">
        - <!--    This schema specifies the Dublin Code element set part of the Microarray Discovery Metadata Card (MAdmc). The MAdmc specifies a set of information
          fields as included in DC Element Set that are to be used to describe all items belonging to the Microarray experiments in MINiML format.   -->
        - <!--    Root element of MAdmc Discovery Metadata Specification    -->
- <xs:element name="MAdmc_DC">
- <xs:complexType>
- <xs:sequence>
  <xs:element ref="Security" />
  <xs:element ref="ResourceDescription" />
  <xs:element ref="FormatDescription" />
  <xs:element ref="ContentDescription" />
    </xs:sequence>
    </xs:complexType>
    </xs:element>
        - <!--

                Definition of MAdmc_DC Layer Elements

    -->
- <xs:element name="Security">
- <xs:annotation>
  <xs:documentation>The security layer elements enable the description of security classification and other security-related fields. These fields provide for the specification
      of security-related attributes of the associated data assets and may be used to support access control.</xs:documentation>
    </xs:annotation>
- <xs:complexType>
- <xs:sequence>
  <xs:element ref="Policy" />
  <xs:element ref="Classification" />
  <xs:element ref="Category" minOccurs="0" maxOccurs="unbounded" />
    </xs:sequence>
    </xs:complexType>
    </xs:element>
- <xs:element name="ResourceDescription">
- <xs:annotation>
  <xs:documentation>The elements of the resource description layer are used to describe aspects of a data asset that support maintenance, administration, and the
      derivation of the data asset.</xs:documentation>
    </xs:annotation>
- <xs:complexType>
- <xs:sequence>
  <xs:element ref="Title" maxOccurs="unbounded" />
  <xs:element ref="Identifier" maxOccurs="unbounded" />
  <xs:element ref="Creator" maxOccurs="unbounded" />
  <xs:element ref="Publisher" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Contributor" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Date" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Rights" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Language" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Type" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Source" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Relation" minOccurs="0" maxOccurs="unbounded" />
    </xs:sequence>
    </xs:complexType>
    </xs:element>
- <xs:element name="FormatDescription">
- <xs:annotation>
  <xs:documentation>The elements of the format description layer are used to provide the description of physical attributes of the asset and include elements such as the
      mime type, source of format definition and version.</xs:documentation>
    </xs:annotation>
- <xs:complexType>
- <xs:sequence>
  <xs:element ref="Format" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="FormatSpecification" maxOccurs="unbounded" />
  <xs:element ref="Version" maxOccurs="unbounded" />
    </xs:sequence>
    </xs:complexType>
    </xs:element>
- <xs:element name="ContentDescription">
- <xs:annotation>
  <xs:documentation>The elements of the content description layer are used to provide the description of concepts and additional contextual aspects of the data asset being
      tagged and include such elements as subject, description, and coverage. These elements are intended to capture asset-level information that describes the content
      and/or context. An additional purpose of the content elements is to aid in precision discovery and to offer a level of description that is better than simple
      indexing.</xs:documentation>
    </xs:annotation>
- <xs:complexType>
- <xs:sequence>
  <xs:element ref="Subject" maxOccurs="unbounded" />
  <xs:element ref="Description" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="Coverage" minOccurs="0" maxOccurs="unbounded" />
    </xs:sequence>
    </xs:complexType>
```

```
     </xs:element>
           - <!--

                         Definition of MAdmc_DC Elements

     -->
- <xs:element name="Category">
- <xs:annotation>
  <xs:documentation>An indication of an additional, specific sensitivity, or a dissemination control, or an informational marking on which no automated access control is
        performed.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Classification">
- <xs:annotation>
  <xs:documentation>Security markings that indicate the sensitivity level of the information.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Contributor">
- <xs:annotation>
  <xs:documentation>An entity responsible for making contributions to the content of the resource.Examples of a Contributor include a person, an organization, or a
        service. Typically, the name of a Contributor should be used to indicate the entity.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Coverage">
- <xs:annotation>
  <xs:documentation>The extent or scope of the content of the resource.The spatial or temporal topic of the resource, the spatial applicability of the resource, or the
        jurisdiction under which the resource is relevant. Spatial topic and spatial applicability may be a named place or a location specified by its geographic
        coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the
        resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named
        places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Creator">
- <xs:annotation>
  <xs:documentation>An entity primarily responsible for making the content of the resource. Examples of a Creator include a person, an organization, or a service.
        Typically, the name of a Creator should be used to indicate the entity.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Date">
- <xs:annotation>
  <xs:documentation>A calendar date associated with an event in the life cycle of the resource. Date may be used to express temporal information at any level of
        granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Description">
- <xs:annotation>
  <xs:documentation>An account of the content of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation,
        or a free-text account of the resource.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Format">
- <xs:annotation>
  <xs:documentation>The physical or digital manifestation of the resource.The file format, physical medium, or dimensions of the resource. Examples of dimensions
        include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="FormatSpecification">
- <xs:annotation>
  <xs:documentation>Format Specification shall be used to further detail the source information about the described items.</xs:documentation>
     </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
     </xs:simpleType>
     </xs:element>
- <xs:element name="Identifier">
- <xs:annotation>
```

```
 <xs:documentation>An unambiguous reference to the resource within a given context. An internal, external, and/or universal identification number for a data asset or
        resource. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Language">
- <xs:annotation>
 <xs:documentation>A language of the intellectual content of the resource.A language of the resource. Recommended best practice is to use a controlled vocabulary such
        as RFC 4646 [RFC4646].</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Policy">
- <xs:annotation>
 <xs:documentation>In the context of information exchange, a security policy is a set of rules for protecting information against unauthorized disclosure, while
        maintaining authorised access, and preventing loss or unauthorized modification.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Publisher">
- <xs:annotation>
 <xs:documentation>An entity responsible for making the resource available.Examples of a Publisher include a person, an organization, or a service. Typically, the name
        of a Publisher should be used to indicate the entity.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Relation">
- <xs:annotation>
 <xs:documentation>A reference to a related resource.Recommended best practice is to identify the related resource by means of a string conforming to a formal
        identification system.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Rights">
- <xs:annotation>
 <xs:documentation>Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated
        with the resource, including intellectual property rights.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Source">
- <xs:annotation>
 <xs:documentation>References to assets or resources from which the tagged data asset is derived. The described resource may be derived from the related resource in
        whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification
        system.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Subject">
- <xs:annotation>
 <xs:documentation>A topic of the content of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended
        best practice is to use a controlled vocabulary. To describe the spatial or temporal topic of the resource, use the Coverage element.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Title">
- <xs:annotation>
 <xs:documentation>A name given to the resource. Typically, a Title will be a name by which the resource is formally known.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
 <xs:restriction base="xs:string" />
    </xs:simpleType>
    </xs:element>
- <xs:element name="Type">
- <xs:annotation>
 <xs:documentation>The nature or genre of the content of the resource. Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary
        [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.For MAdmc_DC this element is limited to
        the various types covered by MAdmc, i.e. MAdmc instances, format specifications or any other document.</xs:documentation>
    </xs:annotation>
- <xs:simpleType>
- <xs:restriction base="xs:string">
 <xs:enumeration value="MAdmc Instance" />
 <xs:enumeration value="Document" />
 <xs:enumeration value="Format Specification" />
    </xs:restriction>
```

```
       </xs:simpleType>
       </xs:element>
- <xs:element name="Version">
- <xs:annotation>
  <xs:documentation>Version shall be used to further detail the version information about the described items.</xs:documentation>
       </xs:annotation>
- <xs:simpleType>
  <xs:restriction base="xs:string" />
       </xs:simpleType>
       </xs:element>
       </xs:schema>
       </xml>
```

MAdmc Program

Below is the project file (MS Visual studio) for MAdmc Program (.aspx) (*Meta file and the project file*).

```
<Project xmlns="http://schemas.microsoft.com/developer/msbuild/2003">
  <ProjectExtensions>
    <VisualStudio>
      <FlavorProperties GUID="{349c5851-65df-11da-9384-00065b846f21}">
        <WebProjectProperties>
          <StartPageUrl>Default.aspx</StartPageUrl>
          <StartAction>SpecificPage</StartAction>
          <AspNetDebugging>True</AspNetDebugging>
          <NativeDebugging>False</NativeDebugging>
          <SQLDebugging>False</SQLDebugging>
          <PublishCopyOption>RunFiles</PublishCopyOption>
          <PublishTargetLocation>
          </PublishTargetLocation>
          <PublishDeleteAllFiles>False</PublishDeleteAllFiles>
          <PublishCopyAppData>True</PublishCopyAppData>
          <ExternalProgram>
          </ExternalProgram>
          <StartExternalURL>
          </StartExternalURL>
          <StartCmdLineArguments>
          </StartCmdLineArguments>
          <StartWorkingDirectory>
          </StartWorkingDirectory>
          <EnableENC>False</EnableENC>
          <AlwaysStartWebServerOnDebug>True</AlwaysStartWebServerOnDebug>
        </WebProjectProperties>
      </FlavorProperties>
    </VisualStudio>
  </ProjectExtensions>
</Project>
```

```
<Project ToolsVersion="3.5" DefaultTargets="Build" xmlns="http://schemas.microsoft.com/developer/msbuild/2003">
  <PropertyGroup>
    <Configuration Condition=" '$(Configuration)' == " ">Debug</Configuration>
    <Platform Condition=" '$(Platform)' == " ">AnyCPU</Platform>
    <ProductVersion>9.0.21022</ProductVersion>
    <SchemaVersion>2.0</SchemaVersion>
    <ProjectGuid>{1057FBFF-7DEE-47D5-863B-4817A63C25D2}</ProjectGuid>
    <ProjectTypeGuids>{349c5851-65df-11da-9384-00065b846f21};{fae04ec0-301f-11d3-bf4b-
00c04f79efbc}</ProjectTypeGuids>
    <OutputType>Library</OutputType>
    <AppDesignerFolder>Properties</AppDesignerFolder>
    <RootNamespace>MAdmc_Program</RootNamespace>
    <AssemblyName>MAdmc_Program</AssemblyName>
    <TargetFrameworkVersion>v3.5</TargetFrameworkVersion>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Debug|AnyCPU' ">
    <DebugSymbols>true</DebugSymbols>
    <DebugType>full</DebugType>
    <Optimize>false</Optimize>
    <OutputPath>bin\</OutputPath>
    <DefineConstants>DEBUG;TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Release|AnyCPU' ">
    <DebugType>pdbonly</DebugType>
    <Optimize>true</Optimize>
    <OutputPath>bin\</OutputPath>
    <DefineConstants>TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <ItemGroup>
    <Reference Include="System" />
    <Reference Include="System.Data" />
    <Reference Include="System.Core">
      <RequiredTargetFramework>3.5</RequiredTargetFramework>
    </Reference>
    <Reference Include="System.Data.DataSetExtensions">
      <RequiredTargetFramework>3.5</RequiredTargetFramework>
    </Reference>
    <Reference Include="System.Web.Extensions">
      <RequiredTargetFramework>3.5</RequiredTargetFramework>
    </Reference>
    <Reference Include="System.Xml.Linq">
      <RequiredTargetFramework>3.5</RequiredTargetFramework>
    </Reference>
    <Reference Include="System.Drawing" />
    <Reference Include="System.Web" />
    <Reference Include="System.Xml" />
    <Reference Include="System.Configuration" />
    <Reference Include="System.Web.Services" />
    <Reference Include="System.EnterpriseServices" />
    <Reference Include="System.Web.Mobile" />
  </ItemGroup>
  <ItemGroup>
    <Content Include="about.aspx" />
    <Content Include="App_Data\test.xml" />
    <Content Include="Default.aspx" />
    <Content Include="help.aspx" />
    <Content Include="Web.config" />
  </ItemGroup>
  <ItemGroup>
    <Compile Include="about.aspx.cs">
      <DependentUpon>about.aspx</DependentUpon>
      <SubType>ASPXCodeBehind</SubType>
    </Compile>
    <Compile Include="about.aspx.designer.cs">
      <DependentUpon>about.aspx</DependentUpon>
    </Compile>
    <Compile Include="Default.aspx.cs">
      <SubType>ASPXCodeBehind</SubType>
      <DependentUpon>Default.aspx</DependentUpon>
    </Compile>
    <Compile Include="Default.aspx.designer.cs">
      <DependentUpon>Default.aspx</DependentUpon>
    </Compile>
    <Compile Include="help.aspx.cs">
      <DependentUpon>help.aspx</DependentUpon>
      <SubType>ASPXCodeBehind</SubType>
    </Compile>
    <Compile Include="help.aspx.designer.cs">
      <DependentUpon>help.aspx</DependentUpon>
    </Compile>
    <Compile Include="Properties\AssemblyInfo.cs" />
  </ItemGroup>
  <Import Project="$(MSBuildBinPath)\Microsoft.CSharp.targets" />
  <Import
Project="$(MSBuildExtensionsPath)\Microsoft\VisualStudio\v9.0\WebApplications\Microsoft.WebApplication.targets" />
  <!-- To modify your build process, add your task inside one of the targets below and uncomment it.
```

```
       Other similar extension points exist, see Microsoft.Common.targets.
  <Target Name="BeforeBuild">
  </Target>
  <Target Name="AfterBuild">
  </Target>
  -->
  <ProjectExtensions>
    <VisualStudio>
      <FlavorProperties GUID="{349c5851-65df-11da-9384-00065b846f21}">
        <WebProjectProperties>
          <UseIIS>False</UseIIS>
          <AutoAssignPort>True</AutoAssignPort>
          <DevelopmentServerPort>1300</DevelopmentServerPort>
          <DevelopmentServerVPath>/</DevelopmentServerVPath>
          <IISUrl>
          </IISUrl>
          <NTLMAuthentication>False</NTLMAuthentication>
          <SaveServerSettingsInUserFile>False</SaveServerSettingsInUserFile>
        </WebProjectProperties>
      </FlavorProperties>
    </VisualStudio>
  </ProjectExtensions>
</Project>
```

Default

```
//-----------------------------------------------------------------------------
//    This is the .aspx code for for MAdmc Program (default.aspx).
// by Fahri KOCABAS
//-----------------------------------------------------------------------------
using System;
using System.Collections;
using System.Configuration;
using System.Data;
using System.Linq;
using System.Web;
using System.Web.Security;
using System.Web.UI;
using System.Web.UI.HtmlControls;
using System.Web.UI.WebControls;
using System.Web.UI.WebControls.WebParts;
using System.IO;
using System.Xml;

namespace MAdmc_Program
{
    public partial class _Default : System.Web.UI.Page
    {
        protected void Page_Load(object sender, EventArgs e)
        {
            if (IsPostBack == false)
            {
                loadedXMLFile.Value = "";
            }
        }
        protected void Button1_Click(object sender, EventArgs e)
        {
            XmlDataSource XmlDataSource1 = new XmlDataSource();
            XmlDataSource1.DataFile = XMLToLoad.PostedFile.FileName;
            TreeView1.ExpandDepth = 1;
            TreeView1.DataSource = XmlDataSource1;
            TreeView1.DataBind();
            loadedXMLFile.Value = XMLToLoad.PostedFile.FileName;
            fillTreeView();

            int maxDepth = 20;

            LevelList.Items.Add(new ListItem("All", "-1"));
            for(int i=0;i<maxDepth;i++){
                LevelList.Items.Add(new ListItem(Convert.ToString(i), Convert.ToString(i)));
            }
            LevelList.SelectedIndex = 2;
        }
        private void collapseNodes(TreeNode parentNode, int depth){
            if (depth == -1)
            {
                parentNode.ExpandAll();
                return;
            }
            if (parentNode.Depth >= Convert.ToUInt32(LevelList.SelectedValue) )
            {
                parentNode.Collapse();
            }
            else
            {
                foreach (TreeNode node in parentNode.ChildNodes)
                {
                    collapseNodes(node, Convert.ToInt32(LevelList.SelectedValue) );
                }
            }
        }
        protected void LevelList_SelectedIndexChanged(object sender, EventArgs e)
        {
                foreach (TreeNode node in TreeView1.Nodes)
            {
                node.ExpandAll();
                collapseNodes(node, Convert.ToInt32(LevelList.SelectedValue) );
            }
        }
        private void checkNodes(TreeNode treeNode)
        {
            treeNode.Checked = true;

            if (treeNode.Parent != null)
            {
                checkNodes(treeNode.Parent);
            }
        }

        private void unCheckNodes(TreeNode treeNode)
        {
            treeNode.Checked = false;

            foreach (TreeNode node in treeNode.ChildNodes)
            {
                unCheckNodes(node);
            }
        }

        protected void TreeView1_TreeNodeCheckChanged(object sender, TreeNodeEventArgs e)
        {
            if (e.Node.Checked == true)
            {
                checkNodes(e.Node);
            }
            else
            {
                unCheckNodes(e.Node);
            }

        }

        protected TreeNode getNextTreeNode(TreeNode treeNode, bool toUpwards)
        {
            bool isFound = false;

            if ((treeNode.ChildNodes.Count > 0)&&(toUpwards==false))
            {
                IEnumerator iterator = treeNode.ChildNodes.GetEnumerator();
                iterator.MoveNext();

                return (TreeNode)iterator.Current;

            }
            else
            {
                if (treeNode.Parent != null)
                {
                    IEnumerator iterator = treeNode.Parent.ChildNodes.GetEnumerator();
                    while ((isFound == false) && (iterator.MoveNext()))
                    {
                        if ((TreeNode)iterator.Current == treeNode)
                        {
                            isFound = true;
                        }
                    }
                    if (iterator.MoveNext())
                    {
                        return (TreeNode)iterator.Current;
                    }
                    else
                    {
                        if (treeNode.Parent != null)
                        {
                            return getNextTreeNode(treeNode.Parent, true);
                        }
                        else
                        {
                            return null;
                        }
                    }
                }
                else
                {
                    return null;
                }
            }
        }

        protected void fillTreeView(){
            if (File.Exists(loadedXMLFile.Value) == true)
            {
                bool fileBegin = true;
                TreeNode root = null;
                TreeNode currNode = null;

                foreach (TreeNode treeNode in TreeView1.Nodes)
                {
                    root = treeNode;
                }

                currNode = root;

                XmlTextReader xmlTextReader = new XmlTextReader(loadedXMLFile.Value);
                xmlTextReader.WhitespaceHandling = WhitespaceHandling.None;

                while (xmlTextReader.Read())
                {
                    switch (xmlTextReader.NodeType)
                    {
                        case XmlNodeType.Element:

                            if (fileBegin != true)
                            {
                                currNode = getNextTreeNode(currNode, false);
                            }
                            else
                            {
                                fileBegin = false;
                            }

                            if (xmlTextReader.IsEmptyElement == true)
                            {
                                while (xmlTextReader.MoveToNextAttribute())
                                {
                                    currNode.Value += xmlTextReader.Name+';'+xmlTextReader.Value+";";
                                }
                            }
                            else
                            {
                                while (xmlTextReader.MoveToNextAttribute())
                                {
                                    currNode.Value += xmlTextReader.Name + ';' + xmlTextReader.Value + ";";
                                }
                            }
                        break;

                    case XmlNodeType.Text:
                        currNode.ToolTip = xmlTextReader.Value;
                        break;

                    case XmlNodeType.EndElement:

                        break;
                    }
                }
            }
        }

        protected void writeXML(XmlTextWriter textWriter)
        {
            TreeNode root = null;

            int prevdepth = -1;

            foreach (TreeNode treeNode in TreeView1.Nodes)
            {
                if (treeNode.Checked == true)
                {
                    root = treeNode;
                }
            }

            TreeNode currNode = root;

            while (currNode != null)
            {

                if (currNode.Checked == true)
                {
                    if (currNode.Depth < prevdepth)
                    {
                        for (int i = 0; i < prevdepth - currNode.Depth;i++ )
                            textWriter.WriteEndElement();
                    }

                    textWriter.WriteStartElement(currNode.Text);

                    if(currNode.Value!="")
                    {
                        string[] attrList = currNode.Value.Split(new char[]{';'});
                        for (int i = 0; i < attrList.Length / 2; i++)
                        {
                            textWriter.WriteAttributeString(attrList[2 * i], attrList[2 * i + 1]);
                        }
                    }

                    if (currNode.ToolTip == "")
                    {
                        if (currNode.ChildNodes.Count == 0)
                        {
                            textWriter.WriteEndElement();
                        }
                        else
                        {
                            bool isAllChildsUnchecked = true;
                            foreach (TreeNode childNode in currNode.ChildNodes)
                            {
                                if (childNode.Checked == true)
                                {
                                    isAllChildsUnchecked = false;
                                }
                            }

                            if (isAllChildsUnchecked == true)
                            {
                                textWriter.WriteEndElement();
                            }
                        }
                    }
                    else
                    {
                        textWriter.WriteString(currNode.ToolTip);
                        textWriter.WriteEndElement();
                    }
                    prevdepth = currNode.Depth;
                }

                currNode = getNextTreeNode(currNode, false);
            }
            for (int i = 0; i < prevdepth; i++)
            {
                textWriter.WriteEndElement();
            }
        }

        protected void Button4_Click(object sender, EventArgs e)
        {
            XmlTextWriter textWriter = new XmlTextWriter(XMLToGenerate.PostedFile.FileName,null);

            textWriter.WriteStartDocument();

            textWriter.WriteStartElement("MAdmc");

            textWriter.WriteStartElement("MINiML_Elements");

            writeXML(textWriter);

            textWriter.WriteEndElement();

            textWriter.WriteStartElement("Extended_DC_Elements");

            textWriter.WriteStartElement("Security");

            //Policy
            textWriter.WriteStartElement("Policy");
            if (PolicydefTB.Text != "")
            {
                textWriter.WriteStartAttribute("Definition");
                textWriter.WriteString(PolicydefTB.Text);
                textWriter.WriteEndAttribute();
            }
            if (PolicyComTB.Text != "")
            {
                textWriter.WriteStartAttribute("Comment");
                textWriter.WriteString(PolicyComTB.Text);
                textWriter.WriteEndAttribute();
            }
            textWriter.WriteString(PolicyTB.Text);
            textWriter.WriteEndElement();

            //Classification
            textWriter.WriteStartElement("Classification");
            if (ClassificationDefTB.Text != "")
            {
                textWriter.WriteStartAttribute("Definition");
                textWriter.WriteString(ClassificationDefTB.Text);
                textWriter.WriteEndAttribute();
            }
            if (ClassificationComTB.Text != "")
```

```
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(ClassificationComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(ClassificationTB.Text);
textWriter.WriteEndElement();

//Category
if (CategoryTB.Text != "")
{
   textWriter.WriteStartElement("Category");
   if (CategoryDefTB.Text != "")
   {
      textWriter.WriteStartAttribute("Definition");
      textWriter.WriteString(CategoryDefTB.Text);
      textWriter.WriteEndAttribute();
   }
   if (CategoryComTB.Text != "")
   {
      textWriter.WriteStartAttribute("Comment");
      textWriter.WriteString(CategoryComTB.Text);
      textWriter.WriteEndAttribute();
   }
   textWriter.WriteString(CategoryTB.Text);
   textWriter.WriteEndElement();
}

textWriter.WriteEndElement();

textWriter.WriteStartElement("Resource_Description");

//Title
textWriter.WriteStartElement("Title");
if (TitleDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(TitleDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (TitleComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(TitleComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(TitleTB.Text);
textWriter.WriteEndElement();


//Identifier
textWriter.WriteStartElement("Identifier");
if (IdentifierDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(IdentifierDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (IdentifierComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(IdentifierComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(IdentifierTB.Text);
textWriter.WriteEndElement();
```

```
//Creator
textWriter.WriteStartElement("Creator");
if (CreatorDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(CreatorDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (CreatorComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(CreatorComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(CreatorTB.Text);
textWriter.WriteEndElement();


//Publisher
if (PublisherTB.Text != "")
{
   textWriter.WriteStartElement("Publisher");
   if (PublisherDefTB.Text != "")
   {
      textWriter.WriteStartAttribute("Definition");
      textWriter.WriteString(PublisherDefTB.Text);
      textWriter.WriteEndAttribute();
   }
   if (PublisherComTB.Text != "")
   {
      textWriter.WriteStartAttribute("Comment");
      textWriter.WriteString(PublisherComTB.Text);
      textWriter.WriteEndAttribute();
   }
   textWriter.WriteString(PublisherTB.Text);
   textWriter.WriteEndElement();
}


//Contributor
if (ContributorTB.Text != "")
{
   textWriter.WriteStartElement("Contributor");
   if (ContributorDefTB.Text != "")
   {
      textWriter.WriteStartAttribute("Definition");
      textWriter.WriteString(ContributorDefTB.Text);
      textWriter.WriteEndAttribute();
   }
   if (ContributorComTB.Text != "")
   {
      textWriter.WriteStartAttribute("Comment");
      textWriter.WriteString(ContributorComTB.Text);
      textWriter.WriteEndAttribute();
   }
   textWriter.WriteString(ContributorTB.Text);
   textWriter.WriteEndElement();
}


//Date
textWriter.WriteStartElement("Date");
if (DateDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(DateDefTB.Text);
   textWriter.WriteEndAttribute();
}
```

```
if (DateComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(DateComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(DateTB.Text);
textWriter.WriteEndElement();


//Rights
if (RightsTB.Text != "")
{
   textWriter.WriteStartElement("Rights");
   if (RightsDefTB.Text != "")
   {
      textWriter.WriteStartAttribute("Definition");
      textWriter.WriteString(RightsDefTB.Text);
      textWriter.WriteEndAttribute();
   }
   if (RightsComTB.Text != "")
   {
      textWriter.WriteStartAttribute("Comment");
      textWriter.WriteString(RightsComTB.Text);
      textWriter.WriteEndAttribute();
   }
   textWriter.WriteString(RightsTB.Text);
   textWriter.WriteEndElement();
}


//Language
textWriter.WriteStartElement("Language");
if (LanguageDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(LanguageDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (LanguageComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(LanguageComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(LanguageTB.Text);
textWriter.WriteEndElement();

//Type
textWriter.WriteStartElement("Type");
if (TypeDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(TypeDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (TypeComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(TypeComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(TypeTB.Text);
textWriter.WriteEndElement();

//Source
if (SourceTB.Text != "")
{
   textWriter.WriteStartElement("Source");

   if (SourceDefTB.Text != "")
   {
      textWriter.WriteStartAttribute("Definition");
      textWriter.WriteString(SourceDefTB.Text);
      textWriter.WriteEndAttribute();
   }
   if (SourceComTB.Text != "")
   {
      textWriter.WriteStartAttribute("Comment");
      textWriter.WriteString(SourceComTB.Text);
      textWriter.WriteEndAttribute();
   }
   textWriter.WriteString(SourceTB.Text);
   textWriter.WriteEndElement();
}

//Relation
if (RelationTB.Text != "")
{
   textWriter.WriteStartElement("Relation");
   if (RelationDefTB.Text != "")
   {
      textWriter.WriteStartAttribute("Definition");
      textWriter.WriteString(RelationDefTB.Text);
      textWriter.WriteEndAttribute();
   }
   if (RelationComTB.Text != "")
   {
      textWriter.WriteStartAttribute("Comment");
      textWriter.WriteString(RelationComTB.Text);
      textWriter.WriteEndAttribute();
   }
   textWriter.WriteString(RelationTB.Text);
   textWriter.WriteEndElement();
}

textWriter.WriteEndElement();

textWriter.WriteStartElement("Format_Description");
//Format
textWriter.WriteStartElement("Format");
if (FormatDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(FormatDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (FormatComTB.Text != "")
{
   textWriter.WriteStartAttribute("Comment");
   textWriter.WriteString(FormatComTB.Text);
   textWriter.WriteEndAttribute();
}
textWriter.WriteString(FormatTB.Text);
textWriter.WriteEndElement();

//Format Specification
textWriter.WriteStartElement("Format_Specification");
if (FormatSpecDefTB.Text != "")
{
   textWriter.WriteStartAttribute("Definition");
   textWriter.WriteString(FormatSpecDefTB.Text);
   textWriter.WriteEndAttribute();
}
if (FormatSpecComTB.Text != "")
{
    textWriter.WriteStartAttribute("Comment");
    textWriter.WriteString(FormatSpecComTB.Text);
    textWriter.WriteEndAttribute();
}
textWriter.WriteString(FormatSpecTB.Text);
textWriter.WriteEndElement();

//Version
textWriter.WriteStartElement("Version");
if (VersionDefTB.Text != "")
{
    textWriter.WriteStartAttribute("Definition");
    textWriter.WriteString(VersionDefTB.Text);
    textWriter.WriteEndAttribute();
}
if (VersionComTB.Text != "")
{
    textWriter.WriteStartAttribute("Comment");
    textWriter.WriteString(VersionComTB.Text);
    textWriter.WriteEndAttribute();
}
textWriter.WriteString(VersionTB.Text);
textWriter.WriteEndElement();

textWriter.WriteEndElement();

textWriter.WriteStartElement("Content_Description");
//Subject
textWriter.WriteStartElement("Subject");
if (SubjectDefTB.Text != "")
{
    textWriter.WriteStartAttribute("Definition");
    textWriter.WriteString(SubjectDefTB.Text);
    textWriter.WriteEndAttribute();
}
if (SubjectComTB.Text != "")
{
    textWriter.WriteStartAttribute("Comment");
    textWriter.WriteString(SubjectComTB.Text);
    textWriter.WriteEndAttribute();
}
textWriter.WriteString(SubjectTB.Text);
textWriter.WriteEndElement();

//Description
textWriter.WriteStartElement("Description");
if (DescDefTB.Text != "")
{
    textWriter.WriteStartAttribute("Definition");
    textWriter.WriteString(DescDefTB.Text);
    textWriter.WriteEndAttribute();
}
if (DescComTB.Text != "")
{
    textWriter.WriteStartAttribute("Comment");
    textWriter.WriteString(DescComTB.Text);
    textWriter.WriteEndAttribute();
}
textWriter.WriteString(DescTB.Text);
textWriter.WriteEndElement();

//Coverage
textWriter.WriteStartElement("Coverage");
if (CoverageDefTB.Text != "")
{
    textWriter.WriteStartAttribute("Definition");
    textWriter.WriteString(CoverageDefTB.Text);
    textWriter.WriteEndAttribute();
```

```
            }
            if (CoverageComTB.Text != "")
            {
                textWriter.WriteStartAttribute("Comment");
                textWriter.WriteString(CoverageComTB.Text);
                textWriter.WriteEndAttribute();
            }
            textWriter.WriteString(CoverageTB.Text);
            textWriter.WriteEndElement();

            textWriter.WriteEndElement();
            textWriter.WriteEndElement();

            textWriter.WriteEndDocument();
            textWriter.Close();
        }

    }
}
```

Default.aspx

```
//-----------------------------------------------------------------------------------
//     This is the stylesheet as used by MAdmc program (default.aspx) .
// by Fahri KOCABAS
//-----------------------------------------------------------------------------------

<%@ Page Language="C#" AutoEventWireup="true" CodeBehind="Default.aspx.cs" Inherits="MAdmc_Program._Default"
%>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" >
<head runat="server">
    <title>Microarray Metadata (MAdmc)Creation</title>
    <style type="text/css">
        .style1
        {
            width: 300px;
        }
        .style3
        {
            text-decoration: underline;
            font-weight: bold;
            font-family: Verdana;
            font-size: medium;
            color: #993300;
        }
        .style4
        {
            font-size: medium;
        }
        .style5
        {
            text-decoration: underline;
            color: #993300;
        }
        .style6
        {
            font-family: Verdana;
            font-size: xx-small;
            font-weight: bold;
        }
    </style>
</head>
<body>
    <form id="form1" runat="server" enctype="multipart/form-data">
    <table>
    <tr>
        <td>
            <asp:Label ID="Label1" runat="server" style="text-align: center"
                Text="Microarray Metadata (MAdmc)Creation" Font-Bold="True"
                Font-Names="Verdana" Font-Size="Larger" ForeColor="#FF3300"></asp:Label>
        </td>
        <td style="text-align: right">
            <asp:LinkButton ID="LinkButton2" runat="server" CausesValidation="False"
                Font-Bold="True" Font-Names="Verdana" Font-Size="XX-Small"
                PostBackUrl="~/about.aspx">About</asp:LinkButton>&nbsp;
            <asp:LinkButton ID="LinkButton1" runat="server" CausesValidation="False"
                Font-Bold="True" Font-Names="Verdana" Font-Size="XX-Small"
                PostBackUrl="~/help.aspx">Help</asp:LinkButton>
            &nbsp;
            </td>
```

```
</tr>
<tr>
    <td valign="top" bgcolor="#99CCFF">
        <table>
            <tr>
                <td valign="top" style="text-align: left" colspan="2" >
                    <span class="style3">MINiML Elements</span><br class="style3" />
                    <span class="style3">(Domain Specific Extension)</span></td>
            </tr>
            <tr>
                <td valign="top" style="text-align: left" >
                    <asp:FileUpload ID="XMLToLoad" runat="server" Font-Names="Verdana"
                        Font-Size="XX-Small" Font-Bold="True" />
                </td>
                <td valign="top" class="style1">
                    <asp:Button ID="Button1" runat="server" Text="Load XML File"
                        OnClick="Button1_Click" Font-Names="Verdana" Font-Size="XX-Small"
                        Font-Bold="True" Width="120px" CausesValidation="False" />
                    <asp:HiddenField ID="loadedXMLFile" runat="server" />
                </td>
            </tr>
            <tr>
                <td valign="middle" style="text-align: left">
                    <asp:FileUpload ID="XMLToGenerate" runat="server" Font-Names="Verdana"
                        Font-Size="XX-Small" Font-Bold="True" />
                  </td>
                  <td>
                    <asp:Button ID="Button4" runat="server" Font-Bold="True" Font-Names="Verdana"
                        Font-Size="XX-Small" Text="Generate XML File" Width="120px"
                          onclick="Button4_Click" />
                </td>
            </tr>
            <tr>
                <td colspan="2" style="text-align: left" valign="middle">
                    <span class="style6">The Level of Detail&nbsp;:</span>
                    &nbsp;<asp:ListBox ID="LevelList" runat="server" Rows="1" AutoPostBack="True"
                        OnSelectedIndexChanged="LevelList_SelectedIndexChanged" Font-Bold="True"
                        Font-Names="Verdana" Font-Size="XX-Small" Width="85px">
                    </asp:ListBox></td>
            </tr>
            <tr>
                <td colspan="2" valign="top">
                    <asp:TreeView ID="TreeView1" runat="server" ImageSet="Simple" NodeIndent="10"
                        ShowCheckBoxes="All"
                        ontreenodecheckchanged="TreeView1_TreeNodeCheckChanged">
                        <ParentNodeStyle Font-Bold="False" />
                        <HoverNodeStyle Font-Underline="True" ForeColor="#DD5555" />
                        <SelectedNodeStyle Font-Underline="True" ForeColor="#DD5555" HorizontalPadding="0px"
                            VerticalPadding="0px" />
                        <NodeStyle Font-Names="Verdana" Font-Size="8pt" ForeColor="Black" HorizontalPadding="0px"
                            NodeSpacing="0px" VerticalPadding="0px" />
                    </asp:TreeView>
                </td>
            </tr>
        </table>
    </td>
    <td valign="top" bgcolor="#CCCCCC">
         <table style="font-family: Verdana; font-size: x-small; font-weight: 700">
         <tr>
            <td colspan="4" class="style4"><span class="style5">Extended DC Element Set </span>
                <br class="style5" />
                <span class="style5">(Metadata Standart Elements)</span></td>
         </tr>
         <tr>
            <td colspan="4">1)SECURITY</td>
         </tr>
         <tr>
            <td>&nbsp;</td>
         <td colspan="2">a)Policy(*)<asp:RequiredFieldValidator ID="RequiredFieldValidator1"
             runat="server" ControlToValidate="PolicyTB"
             ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
          </td><td>
               <asp:TextBox ID="PolicyTB" runat="server" Font-Bold="True" Font-Names="Verdana"
                 Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td><td>&nbsp;</td>
         <td>i.Definition</td><td>
               <asp:TextBox ID="PolicydefTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
         <td>ii.Comment</td><td>
               <asp:TextBox ID="PolicyComTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td colspan="2">b)Classification(*)<asp:RequiredFieldValidator
             ID="RequiredFieldValidator2" runat="server"
             ControlToValidate="ClassificationTB"
ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
               </td><td>
               <asp:TextBox ID="ClassificationTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
         <td>i.Definition</td><td>
               <asp:TextBox ID="ClassificationDefTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
         <td>ii.Comment</td><td>
               <asp:TextBox ID="ClassificationComTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td colspan="2">c)Category</td><td>
               <asp:TextBox ID="CategoryTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
         <td>i.Definition<td>
               <asp:TextBox ID="CategoryDefTB" runat="server" Font-Bold="True"
                 Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
               </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
```

```
  <td>ii.Comment</td><td>
          <asp:TextBox ID="CategoryComTB" runat="server" Font-Bold="True"
             Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>
<tr>
  <td colspan="4">&nbsp;</td>
</tr>
<tr>
  <td colspan="4">2)RESOURCE DESCRIPTION</td>
</tr>
<tr>
<td>&nbsp;</td>
  <td colspan="2">a)Title(*)<asp:RequiredFieldValidator ID="RequiredFieldValidator3"
       runat="server" ControlToValidate="TitleTB"
       ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
          <asp:TextBox ID="TitleTB" runat="server" Font-Bold="True" Font-Names="Verdana"
             Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition<td>
          <asp:TextBox ID="TitleDefTB" runat="server" Font-Bold="True"
             Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
          <asp:TextBox ID="TitleComTB" runat="server" Font-Bold="True"
             Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>

<tr>
  <td>&nbsp;</td>
  <td colspan="2">b)Identifier(*)<asp:RequiredFieldValidator
       ID="RequiredFieldValidator4" runat="server" ControlToValidate="IdentifierTB"
       ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
          <asp:TextBox ID="IdentifierTB" runat="server" Font-Bold="True"
             Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition</td><td>
          <asp:TextBox ID="IdentifierDefTB" runat="server" Font-Bold="True"
             Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
          <asp:TextBox ID="IdentifierComTB" runat="server" Font-Bold="True"
             Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
          </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td colspan="2">c)Creator(*)<asp:RequiredFieldValidator
       ID="RequiredFieldValidator5" runat="server" ControlToValidate="CreatorTB"
```

```
        ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
            <asp:TextBox ID="CreatorTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td>&nbsp;</td>
   <td>i.Definition</td><td>
            <asp:TextBox ID="CreatorDefTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td>&nbsp;</td>
   <td>ii.Comment</td><td>
            <asp:TextBox ID="CreatorComTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td colspan="2">d)Publisher</td><td>
            <asp:TextBox ID="PublisherTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td>&nbsp;</td>
   <td>i.Definition</td><td>
            <asp:TextBox ID="PublisherDefTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td>&nbsp;</td>
   <td>ii.Comment</td><td>
            <asp:TextBox ID="PublisherComTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td colspan="2">e)Contributor</td><td>
            <asp:TextBox ID="ContributorTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td>&nbsp;</td>
   <td>i.Definition<td>
            <asp:TextBox ID="ContributorDefTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
   <td>&nbsp;</td>
   <td>&nbsp;</td>
   <td>ii.Comment</td><td>
            <asp:TextBox ID="ContributorComTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
```

```
        <tr>
          <td>&nbsp;</td>
          <td colspan="2">f)Date(*)<asp:RequiredFieldValidator ID="RequiredFieldValidator6"
              runat="server" ControlToValidate="DateTB"
ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
           </td><td>
                <asp:TextBox ID="DateTB" runat="server" Font-Bold="True" Font-Names="Verdana"
                  Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>i.Definition<td>
                <asp:TextBox ID="DateDefTB" runat="server" Font-Bold="True"
                  Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>ii.Comment</td><td>
                <asp:TextBox ID="DateComTB" runat="server" Font-Bold="True"
                  Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>

        <tr>
          <td>&nbsp;</td>
          <td colspan="2">g)Rights</td><td>
                <asp:TextBox ID="RightsTB" runat="server" Font-Bold="True" Font-Names="Verdana"
                  Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>i.Definition<td>
                <asp:TextBox ID="RightsDefTB" runat="server" Font-Bold="True"
                  Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>ii.Comment</td><td>
                <asp:TextBox ID="RightsComTB" runat="server" Font-Bold="True"
                  Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>

        <tr>
          <td>&nbsp;</td>
          <td colspan="2">h)Language(*)<asp:RequiredFieldValidator
              ID="RequiredFieldValidator7" runat="server" ControlToValidate="LanguageTB"
              ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
           </td><td>
                <asp:TextBox ID="LanguageTB" runat="server" Font-Bold="True"
                  Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>i.Definition<td>
                <asp:TextBox ID="LanguageDefTB" runat="server" Font-Bold="True"
                  Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>ii.Comment</td><td>
                  <asp:TextBox ID="LanguageComTB" runat="server" Font-Bold="True"
                     Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td colspan="2">i)Type(*)<asp:RequiredFieldValidator ID="RequiredFieldValidator8"
                runat="server" ControlToValidate="TypeTB"
ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
           </td><td>
                  <asp:TextBox ID="TypeTB" runat="server" Font-Bold="True" Font-Names="Verdana"
                     Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>i.Definition<td>
                  <asp:TextBox ID="TypeDefTB" runat="server" Font-Bold="True"
                     Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>ii.Comment</td><td>
                  <asp:TextBox ID="TypeComTB" runat="server" Font-Bold="True"
                     Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>

        <tr>
          <td>&nbsp;</td>
          <td colspan="2">j)Source</td><td>
                  <asp:TextBox ID="SourceTB" runat="server" Font-Bold="True" Font-Names="Verdana"
                     Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>i.Definition<td>
                  <asp:TextBox ID="SourceDefTB" runat="server" Font-Bold="True"
                     Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>
        <tr>
          <td>&nbsp;</td>
          <td>&nbsp;</td>
          <td>ii.Comment</td><td>
                  <asp:TextBox ID="SourceComTB" runat="server" Font-Bold="True"
                     Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>

        <tr>
          <td>&nbsp;</td>
          <td colspan="2">k)Relation</td><td>
                  <asp:TextBox ID="RelationTB" runat="server" Font-Bold="True"
                     Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                  </td>
        </tr>
        <tr>
```

```
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition<td>
            <asp:TextBox ID="RelationDefTB" runat="server" Font-Bold="True"
              Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
            <asp:TextBox ID="RelationComTB" runat="server" Font-Bold="True"
              Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td colspan="4">&nbsp;</td>
</tr>
<tr>
  <td colspan="4">3)FORMAT DESCRIPTION     <tr>
  <td>&nbsp;</td>
  <td colspan="2">a)Format(*)<asp:RequiredFieldValidator ID="RequiredFieldValidator9"
        runat="server" ControlToValidate="FormatTB"
        ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
            <asp:TextBox ID="FormatTB" runat="server" Font-Bold="True" Font-Names="Verdana"
              Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition<td>
            <asp:TextBox ID="FormatDefTB" runat="server" Font-Bold="True"
              Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
            <asp:TextBox ID="FormatComTB" runat="server" Font-Bold="True"
              Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>

<tr>
  <td>&nbsp;</td>
  <td colspan="2">b)Format Specification(*)<asp:RequiredFieldValidator
        ID="RequiredFieldValidator10" runat="server" ControlToValidate="FormatSpecTB"
        ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
            <asp:TextBox ID="FormatSpecTB" runat="server" Font-Bold="True"
              Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition<td>
            <asp:TextBox ID="FormatSpecDefTB" runat="server" Font-Bold="True"
              Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
```

```
            <asp:TextBox ID="FormatSpecComTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>

<tr>
  <td>&nbsp;</td>
  <td colspan="2">c)Version(*)<asp:RequiredFieldValidator
        ID="RequiredFieldValidator11" runat="server" ControlToValidate="VersionTB"
        ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
            <asp:TextBox ID="VersionTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition</td><td>
            <asp:TextBox ID="VersionDefTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
            <asp:TextBox ID="VersionComTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td colspan="4">&nbsp;</td>
</tr>
<tr>
  <td colspan="4">4)CONTENT DESCRIPTION   </tr>
<tr>
  <td>&nbsp;</td>
  <td colspan="2">a)Subject(*)<asp:RequiredFieldValidator
        ID="RequiredFieldValidator12" runat="server" ControlToValidate="SubjectTB"
        ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
   </td><td>
            <asp:TextBox ID="SubjectTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>i.Definition</td><td>
            <asp:TextBox ID="SubjectDefTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>
<tr>
  <td>&nbsp;</td>
  <td>&nbsp;</td>
  <td>ii.Comment</td><td>
            <asp:TextBox ID="SubjectComTB" runat="server" Font-Bold="True"
               Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
            </td>
</tr>

<tr>
  <td>&nbsp;</td>
  <td colspan="2">b)Description(*)<asp:RequiredFieldValidator
        ID="RequiredFieldValidator13" runat="server" ControlToValidate="DescTB"
        ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
```

```
          </td>
         <td>
                   <asp:TextBox ID="DescTB" runat="server" Font-Bold="True"
                      Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                   </td>
       </tr>            <td>&nbsp;</td>
         <td>&nbsp;</td><td>
  i.Definition</td><td>
                   <asp:TextBox ID="DescDefTB" runat="server" Font-Bold="True"
                      Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                   </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
         <td>ii.Comment</td><td>
                   <asp:TextBox ID="DescComTB" runat="server" Font-Bold="True"
                      Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                   </td>
       </tr>

       <tr>
         <td>&nbsp;</td>
         <td colspan="2">c)Coverage(*)<asp:RequiredFieldValidator
                ID="RequiredFieldValidator14" runat="server" ControlToValidate="CoverageTB"
                ErrorMessage="RequiredFieldValidator">*</asp:RequiredFieldValidator>
          </td>
         <td>
                   <asp:TextBox ID="CoverageTB" runat="server" Font-Bold="True"
                      Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                   </td>
       </tr>
       <tr>
                <td>&nbsp;</td>
         <td>&nbsp;</td><td>i.Definition</td><td>
                   <asp:TextBox ID="CoverageDefTB" runat="server" Font-Bold="True"
                      Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                   </td>
       </tr>
       <tr>
         <td>&nbsp;</td>
         <td>&nbsp;</td>
         <td>ii.Comment</td><td>
                   <asp:TextBox ID="CoverageComTB" runat="server" Font-Bold="True"
                      Font-Names="Verdana" Font-Size="XX-Small"></asp:TextBox>
                   </td>
       </tr>

       </table>

</td></tr></table>
   </form>
</body>
</html>
```

Default.aspx.designer

```
//------------------------------------------------------------------------------
// <auto-generated>
//     Runtime Version:2.0.50727.1433
// by Fahri KOCABAS
//     This is the auto generated designer code for MAdmc program as called by default page from .aspx server.
// </auto-generated>
//------------------------------------------------------------------------------


namespace MAdmc_Program {


    public partial class _Default {

        /// <summary>
        /// form1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.HtmlControls.HtmlForm form1;

        /// <summary>
        /// Label1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.Label Label1;

        /// <summary>
        /// LinkButton2 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.LinkButton LinkButton2;

        /// <summary>
        /// LinkButton1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.LinkButton LinkButton1;

        /// <summary>
        /// XMLToLoad control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.FileUpload XMLToLoad;

        /// <summary>
        /// Button1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.Button Button1;

        /// <summary>
        /// loadedXMLFile control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.HiddenField loadedXMLFile;

        /// <summary>
        /// XMLToGenerate control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.FileUpload XMLToGenerate;

        /// <summary>
        /// Button4 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.Button Button4;

        /// <summary>
        /// LevelList control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.ListBox LevelList;

        /// <summary>
        /// TreeView1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.TreeView TreeView1;

        /// <summary>
        /// RequiredFieldValidator1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator1;

        /// <summary>
        /// PolicyTB control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
```

```
protected global::System.Web.UI.WebControls.TextBox PolicyTB;

/// <summary>
/// PolicydefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox PolicydefTB;

/// <summary>
/// PolicyComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox PolicyComTB;

/// <summary>
/// RequiredFieldValidator2 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator2;

/// <summary>
/// ClassificationTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox ClassificationTB;

/// <summary>
/// ClassificationDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox ClassificationDefTB;

/// <summary>
/// ClassificationComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox ClassificationComTB;

/// <summary>
/// CategoryTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox CategoryTB;

/// <summary>
/// CategoryDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox CategoryDefTB;

/// <summary>
/// CategoryComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox CategoryComTB;

/// <summary>
/// RequiredFieldValidator3 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator3;

/// <summary>
/// TitleTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox TitleTB;

/// <summary>
/// TitleDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox TitleDefTB;

/// <summary>
/// TitleComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox TitleComTB;

/// <summary>
/// RequiredFieldValidator4 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator4;

/// <summary>
/// IdentifierTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox IdentifierTB;
```

```
/// <summary>
/// IdentifierDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox IdentifierDefTB;

/// <summary>
/// IdentifierComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox IdentifierComTB;

/// <summary>
/// RequiredFieldValidator5 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator5;

/// <summary>
/// CreatorTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox CreatorTB;

/// <summary>
/// CreatorDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox CreatorDefTB;

/// <summary>
/// CreatorComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox CreatorComTB;

/// <summary>
/// PublisherTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox PublisherTB;

/// <summary>
/// PublisherDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox PublisherDefTB;

/// <summary>
/// PublisherComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox PublisherComTB;

/// <summary>
/// ContributorTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox ContributorTB;

/// <summary>
/// ContributorDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox ContributorDefTB;

/// <summary>
/// ContributorComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox ContributorComTB;

/// <summary>
/// RequiredFieldValidator6 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator6;

/// <summary>
/// DateTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox DateTB;

/// <summary>
/// DateDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox DateDefTB;
```

```
/// <summary>
/// DateComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox DateComTB;

/// <summary>
/// RightsTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox RightsTB;

/// <summary>
/// RightsDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox RightsDefTB;

/// <summary>
/// RightsComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox RightsComTB;

/// <summary>
/// RequiredFieldValidator7 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator7;

/// <summary>
/// LanguageTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox LanguageTB;

/// <summary>
/// LanguageDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox LanguageDefTB;

/// <summary>
/// LanguageComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox LanguageComTB;

/// <summary>
/// RequiredFieldValidator8 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator8;

/// <summary>
/// TypeTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox TypeTB;

/// <summary>
/// TypeDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox TypeDefTB;

/// <summary>
/// TypeComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox TypeComTB;

/// <summary>
/// SourceTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox SourceTB;

/// <summary>
/// SourceDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox SourceDefTB;

/// <summary>
/// SourceComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox SourceComTB;

/// <summary>
/// RelationTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox RelationTB;

/// <summary>
/// RelationDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox RelationDefTB;

/// <summary>
/// RelationComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox RelationComTB;

/// <summary>
/// RequiredFieldValidator9 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator9;

/// <summary>
/// FormatTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox FormatTB;

/// <summary>
/// FormatDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox FormatDefTB;

/// <summary>
/// FormatComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox FormatComTB;

/// <summary>
/// RequiredFieldValidator10 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator10;

/// <summary>
/// FormatSpecTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox FormatSpecTB;

/// <summary>
/// FormatSpecDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox FormatSpecDefTB;

/// <summary>
/// FormatSpecComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox FormatSpecComTB;

/// <summary>
/// RequiredFieldValidator11 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator11;

/// <summary>
/// VersionTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox VersionTB;

/// <summary>
/// VersionDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox VersionDefTB;

/// <summary>
/// VersionComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox VersionComTB;

/// <summary>
/// RequiredFieldValidator12 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator12;

/// <summary>
/// SubjectTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox SubjectTB;

/// <summary>
/// SubjectDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox SubjectDefTB;

/// <summary>
/// SubjectComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox SubjectComTB;

/// <summary>
/// RequiredFieldValidator13 control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator13;

/// <summary>
/// DescTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox DescTB;

/// <summary>
/// DescDefTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
protected global::System.Web.UI.WebControls.TextBox DescDefTB;

/// <summary>
/// DescComTB control.
/// </summary>
/// <remarks>
/// Auto-generated field.
/// To modify move field declaration from designer file to code-behind file.
/// </remarks>
```

```
        protected global::System.Web.UI.WebControls.TextBox DescComTB;

        /// <summary>
        /// RequiredFieldValidator14 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.RequiredFieldValidator RequiredFieldValidator14;

        /// <summary>
        /// CoverageTB control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.TextBox CoverageTB;

        /// <summary>
        /// CoverageDefTB control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.TextBox CoverageDefTB;

        /// <summary>
        /// CoverageComTB control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.WebControls.TextBox CoverageComTB;
    }
}
```

About

```
//------------------------------------------------------------------------------
//     This is the stylesheet as used by about.aspx .
// by Fahri KOCABAS
//------------------------------------------------------------------------------



<%@ Page Language="C#" AutoEventWireup="true" CodeBehind="about.aspx.cs" Inherits="MAdmc_Program.about" %>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" >
<head runat="server">
    <title>Untitled Page</title>
    <style type="text/css">

 p.MsoNormal
                {margin-bottom:.0001pt;
                font-size:12.0pt;
                font-family:"Times New Roman","serif";
                        margin-left: 0cm;
            margin-right: 0cm;
            margin-top: 0cm;    }
a:link
                {color:#333333;
                text-decoration:underline;
                text-underline:single;    }
        .style1
        {
            font-weight: bold;      }
    </style>
</head>
<body>
    <form id="form1" runat="server">
    <div>

        <p class="MsoNormal"
            style="font-family: Verdana; font-size: x-small; font-weight: bold">
            <span lang="EN-US">The program MAdmc version 1.0 </span>
        </p>
        <p class="MsoNormal"
            style="font-family: Verdana; font-size: x-small; font-weight: bold">
            <span lang="EN-US">Microarray Metacard Definition Tool</span></p>
        <p class="MsoNormal"
            style="font-family: Verdana; font-size: x-small; font-weight: bold">
            <span lang="EN-US">METU Informatics Institute</span></p>
        <p class="MsoNormal" style="font-family: Verdana; font-size: x-small">
            <span class="style1" lang="DE" style="mso-ansi-language:DE">E-mail: </span>
            <span lang="EN-US"><a href="mailto:fahri@ii.metu.edu.tr"><span class="style1"
                lang="DE" style="mso-ansi-language:DE">fahri@ii.metu.edu.tr</span></a></span><span
                class="style1" lang="DE" style="mso-ansi-language:DE">;
            f.kocabas@hq.nato.int<o:p></o:p></span></p>
        <p class="MsoNormal">
            <span lang="DE" style="mso-ansi-language:DE"><o:p>&nbsp;</o:p></span></p>

    </div>
    </form>
</body>
</html>
```

about.aspx

```
//------------------------------------------------------------------------------
//     This is the .aspx code for about file for MAdmc Program.
// by Fahri KOCABAS
//------------------------------------------------------------------------------



using System;
using System.Collections;
using System.Configuration;
using System.Data;
using System.Linq;
using System.Web;
using System.Web.Security;
using System.Web.UI;
using System.Web.UI.HtmlControls;
using System.Web.UI.WebControls;
using System.Web.UI.WebControls.WebParts;
using System.Xml.Linq;

namespace MAdmc_Program
{
    public partial class about : System.Web.UI.Page
    {
        protected void Page_Load(object sender, EventArgs e)
        {

        }
    }
}
```

about.aspx.designer

```
//------------------------------------------------------------------------------
// <auto-generated>
//     Runtime Version:2.0.50727.1433
// by Fahri KOCABAS
//     This is the auto generated designer code for about file for MAdmc program.
// </auto-generated>
//------------------------------------------------------------------------------


namespace MAdmc_Program {


    public partial class about {

        /// <summary>
        /// form1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.HtmlControls.HtmlForm form1;
    }
}
```

Help

```
//------------------------------------------------------------------------------
//     This is the stylesheet as used by help.aspx .
// by Fahri KOCABAS
//------------------------------------------------------------------------------

<%@ Page Language="C#" AutoEventWireup="true" CodeBehind="help.aspx.cs" Inherits="MAdmc_Program.help" %>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" >
<head runat="server">
    <title>Untitled Page</title>
    <style type="text/css">

 p.MsoNormal
                        {margin-bottom:.0001pt;
                        font-size:12.0pt;
                        font-family:"Times New Roman","serif";
                                margin-left: 0cm;
            margin-right: 0cm;
            margin-top: 0cm;
        }
ul
                        {margin-bottom:0cm;}
 li.MsoNormal
                        {margin-bottom:.0001pt;
                        font-size:12.0pt;
                        font-family:"Times New Roman","serif";
                                margin-left: 0cm;
            margin-right: 0cm;
            margin-top: 0cm;
        }
h3
                        {margin-right:0cm;
                        margin-left:0cm;
                        font-size:12.5pt;
                        font-family:"Verdana","sans-serif";
                        color:black;
                        }
pre
                        {margin-bottom:.0001pt;
                        font-size:10.0pt;
                        font-family:Courier;
                        color:black;
                                margin-left: 0cm;
            margin-right: 0cm;
            margin-top: 0cm;
        }
a:link
                        {color:#333333;
                        text-decoration:underline;
                        text-underline:single;
        }
    </style>
</head>
<body style="font-family: Verdana">
    <form id="form1" runat="server">
    <div>

        <p class="MsoNormal" style="margin-left:36.0pt;text-indent:-18.0pt;mso-list:l0 level4 lfo1;
tab-stops:list 36.0pt">
            </p>
        <ul style="margin-top:0cm" type="disc">
            <li class="MsoNormal" style="mso-list:l1 level1 lfo2;tab-stops:list 36.0pt">
                <span lang="EN-US">The program MAdmc reads the
            MINiML xml file that is in GEO, NCBI Microarray Repository and guides the user
            to form the metacard for the experiment for upload in XML format. There is a
            schema file to support these instances.</span></li>
            <li class="MsoNormal" style="mso-list:l1 level1 lfo2;tab-stops:list 36.0pt">
                <span lang="EN-US">There are 15 elements in Dublin Core Metadata Element Set
                (http://dublincore.org/documents/1999/07/02/dces/) and there are 10 attributes
                to attach any data element as specified by ISO 11179
                (http://metadata-standards.org/11179/) – Standart for the description of data
                element).</span></li>
            <li class="MsoNormal" style="mso-list:l1 level1 lfo2;tab-stops:list 36.0pt">
                <span lang="EN-US">The definitions, data types, and constraints, ie content
                model is in XML schema (XSD) file.</span></li>
            <li class="MsoNormal" style="mso-list:l1 level1 lfo2;tab-stops:list 36.0pt">
                <span lang="EN-US">The 15 Elements of Dublin Core Metadata Elements set</span></li>
        </ul>
        <h3 style="text-align:justify">
```

```
      <strong>
      <span lang="EN-US" style="font-size:8.0pt;
font-family:&quot;Verdana&quot;,&quot;sans-serif&quot;">Element: Title</span></strong><span
        lang="EN-US" style="font-size:8.0pt"><o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Title<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Title<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>A name given to the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Typically, a Title will be a name by which the resource is
<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>formally
known.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="creator"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Creator<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Creator<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Creator<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>An entity primarily responsible for making the content
of<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Examples of a Creator include a person, an
organisation,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>or a service.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>Typically, the name of a Creator should be used to<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>indicate the entity.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="subject"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Subject<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Subject and Keywords<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Subject<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>The topic of the content of the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Typically, a Subject will be expressed as
keywords,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>key phrases or classification codes that describe a topic<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>of the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>Recommended best practice is to select a value from a<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>controlled vocabulary or formal classification scheme.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="description"></a><span lang="FR"
        style="font-size:8.0pt;mso-ansi-language:FR">Element: Description<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="FR" style="font-size:8.0pt;
mso-ansi-language:FR"><span style="mso-spacerun:yes">  </span>Name:<span
      style="mso-spacerun:yes">        </span>Description<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="FR" style="font-size:8.0pt;mso-ansi-language:
FR"><span style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Description<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="FR" style="font-size:8.0pt;mso-ansi-language:
FR"><span style="mso-spacerun:yes">  </span></span><span lang="EN-US"
      style="font-size:8.0pt">Definition:<span style="mso-spacerun:yes">  </span>An account of the content of the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Description may include but is not limited to: an
abstract,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>table of contents, reference to a graphical representation<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>of content or a free-text account of the content.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="publisher"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Publisher<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Publisher<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Publisher<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>An entity responsible for making the resource
available<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Examples of a Publisher include a person, an
organisation,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">               </span>or a service.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
```

```
      style="mso-spacerun:yes">              </span>Typically, the name of a Publisher should be used to<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>indicate the entity.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="contributor"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Contributor<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Contributor<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Contributor<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>An entity responsible for making contributions to
the<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>content of the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Examples of a Contributor include a person, an
organisation,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>or a service.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Typically, the name of a Contributor should be used to<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>indicate the entity.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="date"></a><span lang="EN-US" style="font-size:8.0pt">Element: Date<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Date<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Date<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>A date associated with an event in the life cycle of
the<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Typically, Date will be associated with the creation
or<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>availability of the resource.<span
      style="mso-spacerun:yes">  </span>Recommended best practice<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>for encoding the date value is defined in a profile of<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>ISO 8601 [<a
      href="http://dublincore.org/documents/1999/07/02/dces/#w3cdtf#w3cdtf">W3CDTF</a>] and follows the YYYY-MM-DD format.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="type"></a><span lang="EN-US" style="font-size:8.0pt">Element: Type<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Resource Type <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;</span>Identifier:<span style="mso-spacerun:yes">  </span>Type<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>The nature or genre of the content of the
resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Type includes terms describing general categories,
functions,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>genres, or aggregation levels for content. Recommended best<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>practice is to select a value from a controlled vocabulary<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">   &nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>(for example, the working draft list of
Dublin Core Types <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>[<a
      href="http://dublincore.org/documents/1999/07/02/dces/#dct1#dct1">DCT1</a>]). To describe the physical or digital manifestation<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>of the resource, use the FORMAT element.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="format"></a><span lang="EN-US" style="font-size:8.0pt">Element: Format<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Format<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Format<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>The physical or digital manifestation of the
resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Typically, Format may include the media-type or dimensions
of<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>the resource. Format may be used to determine the software, <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>hardware or
other equipment needed to display or operate the <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>resource.
Examples of dimensions include size and duration.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
```

```
      style="mso-spacerun:yes">              </span>Recommended best practice is to select a value from a<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>controlled vocabulary (for example, the list of Internet Media<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Types [<a
      href="http://dublincore.org/documents/1999/07/02/dces/#mime#mime">MIME</a>] defining computer media formats).<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="identifier"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Identifier<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Resource Identifier<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Identifier<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>An unambiguous reference to the resource within a given
context.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Recommended best practice is to identify the resource by
means<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>of a string or number conforming to a formal identification<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>system. <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>Example formal
identification systems include the Uniform<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Resource Identifier (URI) (including the Uniform Resource<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Locator (URL)), the Digital Object Identifier (DOI) and the<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>International Standard Book Number (ISBN).<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="source"></a><span lang="EN-US" style="font-size:8.0pt">Element: Source<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Source<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Source<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>A Reference to a resource from which the present
resource<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>is derived.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>The present resource may be derived from the Source
resource<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>in whole or in part.<span
      style="mso-spacerun:yes">  </span>Recommended best practice is to reference <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>the resource by
means of a string or number conforming to a <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>formal
identification system.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="language"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Language<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Language<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Language<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>A language of the intellectual content of the
resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Recommended best practice for the values of the
Language<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>element is defined by RFC 1766 [<a
      href="http://dublincore.org/documents/1999/07/02/dces/#rfc1766#rfc1766">RFC1766</a>] which includes<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>a two-letter Language Code (taken from the ISO 639<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>standard [<a
      href="http://dublincore.org/documents/1999/07/02/dces/#iso639#iso639">ISO639</a>]), followed optionally, by a two-letter<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Country Code (taken from the ISO 3166 standard [<a
      href="http://dublincore.org/documents/1999/07/02/dces/#iso3166#iso3166">ISO3166</a>]). <o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</span>For example,
&#39;en&#39; for English, &#39;fr&#39; for French, or<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>&#39;en-uk&#39; for English used in the <st1:country-region
w:st="on"><st1:place w:st="on">United Kingdom</st1:place></st1:country-region>.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="relation"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Relation<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Relation<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
```

```
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Relation<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>A reference to a related resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Recommended best practice is to reference the resource by
means<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>of a string or number conforming to a formal identification<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>system.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="coverage"></a><span lang="EN-US" style="font-size:8.0pt">Element:
      Coverage<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">        </span>Coverage<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Identifier:<span style="mso-spacerun:yes">  </span>Coverage<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition:<span style="mso-spacerun:yes">  </span>The extent or scope of the content of the
resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">     </span>Coverage will typically include spatial location (a place
name<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>or geographic coordinates), temporal period (a period label,<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>date, or date range) or jurisdiction (such as a named<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>administrative entity).<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Recommended best practice is to select a value from a<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>controlled vocabulary (for example, the Thesaurus of Geographic<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>Names [TGN]) and that, where appropriate, named places or time<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>periods be used in preference to numeric identifiers such as<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">              </span>sets of coordinates or date ranges.<o:p></o:p></span></pre>
    <h3 style="text-align:justify">
      <a name="rights"></a><span lang="EN-US" style="font-size:8.0pt">Element: Rights<o:p></o:p></span></h3>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Name:<span style="mso-spacerun:yes">       </span>Rights Management<span
      style="mso-spacerun:yes">   </span><o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">&nbsp;&nbsp;</span>Identifier: Rights<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Definition: Information about rights held in and over the resource.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">  </span>Comment:<span style="mso-spacerun:yes">    </span>Typically, a Rights element will contain a
rights<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>management statement for the resource, or reference<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>a service providing such information. Rights information<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>often encompasses Intellectual Property Rights (IPR),<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>Copyright, and various Property Rights.<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>If the Rights element is absent, no assumptions can be made<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>about the status of these and other rights with respect to<o:p></o:p></span></pre>
    <pre style="text-align:justify"><span lang="EN-US" style="font-size:8.0pt"><span
      style="mso-spacerun:yes">             </span>the resource.<o:p></o:p></span></pre>
    <p class="MsoNormal">
      <span lang="EN-US"><o:p>&nbsp;</o:p></span></p>
    <ul style="margin-top:0cm" type="disc">
      <li class="MsoNormal" style="mso-list:l1 level1 lfo2;tab-stops:list 36.0pt">
        <span lang="EN-US">ISO 11179 Attribute List</span></li>
    </ul>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Name</span></b><span lang="EN-US"
        style="color:black"> - The label assigned to the data element <o:p></o:p>
      </span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Identifier</span></b><span
        lang="EN-US" style="color:black"> - The unique identifier assigned to the
      data element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Version</span></b><span lang="EN-US"
        style="color:black"> - The version of the data element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Registration Authority</span></b><span
        lang="EN-US" style="color:black"> - The entity authorised to register the
      data element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
```

```
      <b><span lang="EN-US" style="color:black">Language</span></b><span lang="EN-US"
        style="color:black"> - The language in which the data element is specified <o:p></o:p>
      </span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Definition</span></b><span
        lang="EN-US" style="color:black"> - A statement that clearly represents the
      concept and essential nature of the data element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Obligation</span></b><span
        lang="EN-US" style="color:black"> - Indicates if the data element is
      required to always or sometimes be present (contain a value) <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Datatype</span></b><span lang="EN-US"
        style="color:black"> - Indicates the type of data that can be represented in
      the value of the data element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Maximum Occurrence</span></b><span
        lang="EN-US" style="color:black"> - Indicates any limit to the repeatability
      of the data element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <b><span lang="EN-US" style="color:black">Comment</span></b><span lang="EN-US"
        style="color:black"> - A remark concerning the application of the data
      element <o:p></o:p></span>
    </p>
    <p class="MsoNormal">
      <span lang="EN-US" style="color:black"><o:p>&nbsp;</o:p></span></p>
    <ul style="margin-top:0cm" type="disc">
      <li class="MsoNormal" style="mso-list:l1 level1 lfo2;tab-stops:list 36.0pt">
        <span lang="EN-US">There are two common attributes selected to further define
        the DC elements which are Definition (Semantic concept) and Comment (Data
        representation)</span></li>
    </ul>

  </div>
  </form>
</body>
</html>
```

help.aspx

```
//------------------------------------------------------------------------------
//     This is the .aspx code for help file for MAdmc Program.
// by Fahri KOCABAS
//------------------------------------------------------------------------------



using System;
using System.Collections;
using System.Configuration;
using System.Data;
using System.Linq;
using System.Web;
using System.Web.Security;
using System.Web.UI;
using System.Web.UI.HtmlControls;
using System.Web.UI.WebControls;
using System.Web.UI.WebControls.WebParts;
using System.Xml.Linq;

namespace MAdmc_Program
{
    public partial class help : System.Web.UI.Page
    {
        protected void Page_Load(object sender, EventArgs e)
        {

        }
    }
}
```

help.aspx.designer

```
//------------------------------------------------------------------------------
// <auto-generated>
//     Runtime Version:2.0.50727.1433
// by Fahri KOCABAS
//     This is the auto generated designer code for help file for MAdmc program.
// </auto-generated>
//------------------------------------------------------------------------------

namespace MAdmc_Program {


    public partial class help {

        /// <summary>
        /// form1 control.
        /// </summary>
        /// <remarks>
        /// Auto-generated field.
        /// To modify move field declaration from designer file to code-behind file.
        /// </remarks>
        protected global::System.Web.UI.HtmlControls.HtmlForm form1;
    }
}
```

VITA

Fahri Kocabaş was born in Çatalzeytin, Kastamonu on May 6, 1964. He received his B.S. degree in Electrics (Telecommunication) from Military Academy in 1986 with honor degree (2nd place). He worked as signal officer until 1991. Then, he completed 1 year course at METU Computer Engineering Department and re-branched to IT. Between 1992 and 1995, he taught IT classes at Army Computer Science School in Ankara while he was pursuing his graduate program at Computer Engineering Department at METU. He earned his MSC at METU Computer Engineering in 1995 with a thesis dedicated to deductive databases. He worked as CIS planning officer at Turkish Military Representation Office, NATO HQ, Brussels between 1995-1998. He earned his MBA degree from Bowie State University in 1998. Then, he was assigned to General Staff HQ CIS Directorate where he served until his resignation in 2005. In this period, he taught classes at METU and he was involved in large projects like strategic C2IS. He participated two summer courses on Bioinformatics: one is NATO Bioinformatics Summer Course in Croatia in 2003 where he contributed a book chapter. The other is Bioinformatics II Graduate Summer School in Istanbul as organized in 2004 by Yeditepe, Ege and Berlin Technical Universities. He also took molecular biology and bioinformatics classes. In 2005, he conducted research on microarray technology at Mayo Clinic, USA. Then, he started his career at NATO HQ as IT staff. He has been teaching online classes at Maryland University since 2009. His main areas of interest are knowledge management and discovery, bioinformatics, semantic interoperability, and semantic web.