

## Metadata Description Framework for Integration of Bioinformatics Information Resources: A Case of iBIRA

Shri Ram and N. Laxman Rao

*Jaypee University of Information Technology, Solan-173 234*

*E-mail: shriram2576@gmail.com*

*Department of Library and Information Science, Osmania University, Hyderabad-500 007*

*E-mail: naglaxman@gmail.com*

### ABSTRACT

Bioinformatics emerged as a new discipline dedicated to the answer the queries about life science using computational approaches. The basic aim of bioinformatics is to create databases, analyse data sets and managing data generated through large-scale projects such as Human Genome project (HGP). It covers a wide variety of traditional computer science domains, such as data modeling, data retrieval, data mining, data integration, data managing, data warehousing, and simulation of biological information generated through laboratory and field experiments. Due to varied form, nature, and activities in the field of bioinformatics, presenting the information in cohesive fashion is major challenge. The bioinformatics information resources are heterogeneous in nature. Integration and interoperability of information is one of the biggest challenges in this field. Bioinformatics, as an emerging field, needs attention towards metadata application for resource discovery. This paper discusses the metadata element set description framework for integration of various information resources related to the field of bioinformatics available over internet. A web-based tool iBIRA: Integrated Bioinformatics Information Resources Access has been designed and developed for integration of bioinformatics information resources. Dublin Core metadata element set has been used for description of information resources and XML schema has been used for interoperability of information resources with others. A database has been designed using structure query language (SQL)–database management system and hypertext preprocessor (PHP)–as web programming languages for integration of bioinformatics resources. The database is designated to categorise various resources into biological database, institutions, journals, patents, software tools, web-servers, etc., and the search result is presented in the form of 'tree view'. Each of these categories of the resources has been analysed for their metadata element set as per the DCMI terms.

**Keywords:** Bioinformatics, Dublin core, information resource integration, metadata description

### 1. INTRODUCTION

Internet has now become universal media for information exchange. Due to heterogeneous nature of information and activities; fragmented information creation and storage, integration and interoperability among various sources of information has become difficult<sup>1-2</sup>. There is significant attention received towards it, since early days of computer and information processing. Much of the work in the context of semantic integration, federated searching, metadata creation, integration of schemas by defining a global schema and mappings with local schema has been done in different fields. Federated database searching<sup>3</sup>, metadata generation for molecular biology databases<sup>4</sup>, computer-based patient records<sup>5-6</sup> are some of the initiatives reflecting the integration of information resources. During these processes various issues aroused pertaining to the standardisation of metadata

elements describing information resources<sup>7-8</sup>. The biggest leap in this area was the launch of Dublin Core Metadata Initiative (DCMI), a project of Online Computer Library Center, Inc. (OCLC) and National Center for Super Computing Applications (NCSA). This project was started in a workshop of OCLC and NSCA in 1995 with the objective to define a set of elements that could be used by authors to describe their own web-resources. Due to proliferation of web-based information resources and advancement of internet technologies; the libraries face challenge to catalogue all the information resources available over internet. The focused aim was to define and standardise some of the key terminology to describe the resources in standard format. The goal of the OCLC-NCSA was to set metadata standard (The term proposed to define a structured information that describe, explains, locate, or otherwise helps in retrieving, using, or managing a resource), with

initially thirteen core elements were later increased to fifteen<sup>9</sup> (Table 1).

The basic purpose of the Dublin Core Metadata Element set was to describe web-based resources. However, the concept continued to develop and find its application in many 'non-traditional' formats<sup>10</sup>. Initially, it was used to describe digital information but later, the term has broadened to include any kind of standardised descriptive information about resources, including non-digital formats<sup>11</sup>. The application of metadata description mostly applied to library catalogues, indexing tools, abstracting tools and archival activities. But, with the proliferation of internet resources, available in digital formats, it has been extended to be used for describing and embedding interoperability of information resources in other subjects (such as bioinformatics) too<sup>12</sup>. With this initial background, the purpose of this paper is to describe the metadata elements for the bioinformatics information resources available over internet, in design and development of an integrated tool, known as *iBIRA*<sup>13</sup>.

## 2. BIOINFORMATICS AND INFORMATION RESOURCE INTEGRATION

Bioinformatics term was first used for 'the study of informatics in biotic system'<sup>14</sup>. It has now emerging as a major element in contemporary biology and biomedical research<sup>15</sup>. Bioinformatics is a discipline where biology, computer science and information technology have merged to form single discipline. The purpose of bioinformatics is to enable the discovery of new biological insights with the help computers and involves the processing of data (collection, analysis, mining, management,

integration, simulation, and visualisation) originated from laboratory experiments. It involves and support computer-based massive data analysis produced by new computational technology application during experimentation in molecular biology and large-scale project such as Human Genome Sequencing projects. These experimentation processes includes the large-scale DNA sequence analysis, use of software tools for sequence analysis and sequence annotation. Later developments in proteomic techniques made this subject transition to ever expanding subject arena. The expansion of the subject can be seen in development in various kinds of databases, web-servers, software development, production of literatures, technologies in sequence analysis and comparison, etc. Furthermore, bioinformatics finds its application in medicine, targeted drug discovery, proposing new hypothesis and through these hypothesis designing new experiments for future research. Due to polymorphism nature of bioinformatics subject, it has made information a multidisciplinary in nature, covering the area of mathematics, statistics, and information and computer technology amalgamation. Bioinformatics is one of the subjects, which is ever expanding in nature. This expansion has been witnessed in the form of new form of data about nucleotide sequences and structures. The information and knowledge is being generated, communicated and used for every day task in creation of databases, development of software tools, sequence annotations, patents, web-servers, and published as literatures, etc. The fusion of information science and biology is expected to continue and expand for foreseeable future. Due to heterogeneity of the information in

**Table 1. Set of 15 DCMI terms proposed by National Information Standard Organisation (NISO)**

S. No.	DC terms	Definition
1.	Contributor	An entity responsible for making contributions to the resource
2.	Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant
3.	Creator	An entity primarily responsible for making the resource
4.	Date	A point or period of time associated with an event in the lifecycle of the resource
5.	Description	An account of the resource
6.	Format	The file format, physical medium, or dimensions of the resource
7.	Identifier	An unambiguous reference to the resource within a given context
8.	Language	A language of the resource
9.	Publisher	An entity responsible for making the resource available
10.	Relation	A related resource
11.	Rights	Information about rights held in and over the resource
12.	Source	A related resource from which the described resource is derived
13.	Subject	The topic of the resource
14.	Title	A name given to the resource
15.	Type	The nature or genre of the resource

this field the various information scientists as well as bioinformatics people started thinking about the interoperability of resources for the purpose of integration in turns facilitate information storage and retrieval.

### 2.1 Information Resource Integration

Bioinformatics subjects deal with multiple data types and formats. Interoperability and compatibility of information resources remains one of the biggest challenges for bioinformatics<sup>16</sup>. The initial purpose of bioinformatics was to manage the biological data with computers; but, after the completion of Human Genome Project in 2003, bioinformatics now aims to manipulate and ultimately discover new biological knowledge through several levels of increasing complexity and format of information resources such as genomics, proteomics, transcriptomics, metabolomics, etc. Integration of information resources can improve the functional annotations and help in the development of new biological hypothesis<sup>17</sup>. Such type of information integration requires suitable information repositories that store and integrate bioinformatics information from different data sources such a molecular biology databases, software tools, journals, institutions imparting bioinformatics education and patents. Effective integration and interoperability of bioinformatics information resources is not a new issue but one that comes up with the information scientists who are working in the organisation with bioinformatics activity facing challenges to provide and manage information successfully. Information scientists with the help of Information and Communication Technologies are trying to cope up with this problem; however working with bioinformatics information stills remains a challenge. *iBIRA*<sup>13</sup> is an effort to integrate different kinds of bioinformatics information resources.

### 3. METADATA APPROACHES FOR INFORMATION RESOURCE DESCRIPTION AND INTEGRATION

It is estimated that at present, the size of internet is more than 8.11 billion pages. This estimation is based on the minimal size of the indexed world wide webpages by popular search engines like Google, Yahoo, Bing, etc.<sup>18</sup> In this scenario finding specific information seems very difficult. However, thousands of indexes and directories exists that attempt to categorise the content of internet by various means. Information storage and retrieval (ISAR) in the present internet era is one of the major challenges in front of the user—to retrieve user-centered information. Formulation of search strategies, ISAR system analysis and design, search monitoring are some of the objectives of ISAR that are useful in successful recall of the targeted search queries. The user does not care whether a relevant document

was missed through faulty acquisition or through faulty retrieval or mismatched metadata input in the system. The users are interested in completeness of information, complacency of relevant documents available anywhere and how many are contained in the answer provided by the ISAR system?

This scenario in bioinformatics is more prominent and pertinent. These are due to the nature of bioinformatics information in respect to (i) data intensive nature of bioinformatics (ii) rate and pace of data generation (iii) multiple scale of data—size, time and order of complexity, and (iv) integration of the data in from one product to another product for generating hypothesis for future research. Such issues have been discussed in some of the earlier works pertaining to the bioinformatics resource integration<sup>19</sup>.

This paper discusses about constructing metadata element sets for the bioinformatics information resources integration in the project *iBIRA*<sup>20</sup>. The aim of the *iBIRA* project is to develop a database for information resources available over web, in order to comprehend their accessibility at single platform. During the study, based on the nature of resources, it is found that the bioinformatics resources primarily available in six broad categories. These categories are:

- (i) Molecular biology databases
- (ii) Software tools
- (iii) Patents
- (iv) Web-servers
- (v) Journals, and
- (vi) Educational and R&D institutions

*iBIRA* segregates all these information resource and integrates them at single platform. In the process to build a base for the project, defining metadata element sets was one of the important aspects that had been taken into consideration. To incorporate information into the local database of *iBIRA*, a metadata integration framework schema was developed as per DC metadata element set. The bioinformatics information sources are being maintained by many autonomous sources<sup>21</sup>. The resources available over internet are often being changed for its schema, content, and addition of new results as and when it appears. Due to volume, diversity and dynamic characteristics of bioinformatics data sources, it is essential to create predefined schema with the help of available standards such as DCMI<sup>22</sup>.

### 4. METADATA DESCRIPTION AND INFORMATION RESOURCE INTEGRATION USING XML

XML is a markup language that allows identification of the structure of a document. XML sets a rule that can be used to control the creation of markup

language and identify document content so that information can be formatted consistently throughout the document. The XML helps in exchange of data with each other by sharing XML files. It provides a tree-based structure for organising information storage. Information is represented as text, with markup describing the hierarchy of character data, container-like element, and attributes of those elements.

The markup language system XML is being used in the description of metadata elements for the purpose of integrating information in to the iBIRA. Earlier the application of XML in presenting information was not popular. But now a days, the publishers are providing provisions for exporting the data into XML format, used by different bibliographic management software such as EndNote, PubMed, etc. PubMed is supplying not just metadata, but full article text in XML. The migration of the data in the form of XML to HTML files for respective category of information resources in being described here in the following sections ahead.

The detailed XML schema for the description of the bioinformatics information resources has been depicted in Fig. 1. Each metadata element has been encoded into XML tag which helps in interoperability with some of the standard database such as NCBI PubMed.

#### 4.1 Molecular Biology Databases

Molecular biology databases are developed on the principle of database management system. As a part of every day tasks, database development is one of the primary activities in bioinformatics. The data generated through experiments carried out by the scientists in laboratory are managed through databases. These activities have given rise to emergence of databases for the management and access of primary sources of information, and also introduce some secondary databases that systematically group the data. The databases are being developed in the

area of nucleic acids, proteins, genomics, proteomics, plants, organelles, immunology as categorised by Nucleic Acid Research journals and published as an annual database issue in the month of January every year<sup>23</sup>, Bioinformatics Link Directory<sup>24</sup>, and BioMed Central Databases Collection summary<sup>25</sup>.

Database interoperability is an important area of bioinformatics research, particularly molecular biology<sup>26</sup>. Molecular Biology databases are the result of extensive research in the area of molecular biology. The Institutions such as European Bioinformatics Institute (EBI), European Molecular Biology Laboratory (EMBL), National Center for Biotechnological Information (NCBI), and DNA Databank of Japan (DBJ) laboratories are pioneering institutions, working for management of data in the area of nucleic acids, protein, analysis of sequences for genomics, and proteomics. In last 10-15 years, the scientists working in this area; collecting biological data and setting up centralised database management system for easy access and long term availability. However, with the exponential growth of experimental data, owing to rapid biotechnological advances and high-throughput technologies, as well as advent of internet as new means of data exchange, the world is dramatically changed. The enormous amount of data now produced on daily basis requires more sophisticated management system and solutions and the availability of the internet as modern infrastructure for the scientific exchange has created new demands for data accessibility and standardisation of the information parameters.

Interoperability of the information parameters in case of databases is one of the key issues for consideration in today's internet world. The databases grow in multifaceted arena. As discussed above that the databases are the result of extensive research in the area of molecular biology. Success of utilisation of these databases depends on the clarity in information architecture, user-friendly interface,

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <biranet_ibira xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
- <ibira>
  <master_list>5064</master_list>
  <nar_sr_no>1476</nar_sr_no>
  <main_clusters>Databases</main_clusters>
  <main_category>Protein sequence databases</main_category>
  <sub_category>Protein properties</sub_category>
  <Name_of_Resource>TopFIND</Name_of_Resource>
  <url_of_resource>http://clipserve.clip.ubc.ca/topfind</url_of_resource>
  <brief_description>Protein N- and C-termini and protease processing</brief_description>
  <name_of_the_organization>University of British Columbia, Vancouver, BC, Canada V6T 1Z3</name_of_the_organization>
  <url_of_the_organization>http://ubc.ca/</url_of_the_organization>
  <address_of_the_organization>University of British Columbia, Vancouver, BC, Canada V6T 1Z3</address_of_the_organization>
  <Country>Canada</Country>
  <Year>2011</Year>
  <contact_person>Philipp F. Lange; Pitter F. Huesgen; Christopher M. Overall</contact_person>
  <contact_email>ris.overall@ubc.ca</contact_email>
  <language>English</language>
  <associated_institutes>Centre for Blood Research, Department of Oral Biological and Medical Sciences, Department of
  Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada V6T
  1Z3</associated_institutes>
  <associated_country>Canada</associated_country>
</ibira>
</biranet_ibira>
```

Figure 1. XML schema for metadata description of bioinformatics resources.

accurate information, authority, and specificity. Thus the design parameter is one of the key concerns in case of information interoperability. The DC metadata element set has been used to describe bioinformatics database resources as given in Table 2.

#### 4.2 Web Server Applications

Web server refers to either the hardware (the computer) or the software (the computer application) that helps to deliver content that can be accessed through the internet. Web server functions as a delivery agent of webpages to a client through

internet. The contents (document, image, style sheet and scripts embedded in a webpage) are delivered through web server<sup>27</sup>. Web servers are invaluable resources for the scientific community and opens access bioinformatics tool are available online to guide and enable research in any number of life science domains. Since Year 2005, in partnership with Nucleic Acid Research, Bioinformatics Links Directory<sup>24</sup> has collected and organised all the web servers and published as a link directory. The information available through these web servers are communicated based on logical subject categories of common tasks.

**Table 2. Comparative metadata element description for different categories of resources**

S. No.	DC terms	Definition	Database	Web servers	Software tools	Journals	Institutions
1.	Contributor	An entity responsible for making contributions to the resource	Creator	Creator	Developer	Authors	N/A
2.	Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant	Subject area	Subject area	Primary function	Subject	Research area
3.	Creator	An entity primarily responsible for making the resource	Creator	Creator	Developer	Editor	Head of the institution
4.	Date	A point or period of time associated with an event in the lifecycle of the resource	Date of launch	Date of launch	Date of product launch	Publication date	Start of Bioinformatics Inf. at the institute
5.	Description	An account of the resource	Description of database	Description of web server	Description of software tool	Description of the journal	Description of institution
6.	Format	The file format, physical medium, or dimensions of the resource	Type and format of database	Type and format of web server	Type of software platform	N/A	N/A
7.	Identifier	An unambiguous reference to the resource within a given context	URL as an identifier	URL as an identifier	Name of the software	ISSN	N/A
8.	Language	A language of the resource	Language in which the database is produced	Language in which the web server developed	N/A	Language of publication	N/A
9.	Publisher	An entity responsible for making the resource available	The institution/ person	The institution/ developer	The developing company/ institution	Publisher of journal	N/A
10.	Relation	A related resource	Related database	Related web servers	N/A	Related journals	Related institutions
11.	Rights	Information about rights held in and over the resource	The Licensing terms of use of Database	The licensing terms of user of web server	Licensing terms of use of software	Copyright information	N/A
12.	Source	A related resource from which the described resource is derived	Database source	Web server source	Company name	Journal	Institutions
13.	Subject	The topic of the resource	Major subject coverage	Major subject coverage	Primary function	Subject coverage	Area of research
14.	Title	A name given to the resource	Name of the Database	Name of the web server	Name of software	Name of the journal	Name of the institution
15.	Type	The nature or genre of the resource	N/A	N/A	N/A	N/A	N/A

While working with *iBIRA* it has been noticed that there is variation in level of descriptions available in Nucleic Acid Research and Bioinformatics Links Directory. For this reason it has been proposed to standardise the level of description of such resources, compliant with the DCMI terms to ensure the interoperability of the resources for future works (Table 2).

#### **4.3 Journals Publishing Bioinformatics Research**

Journals are primary source of information for publishing research articles. Bioinformatics is also serving its people with huge amount of research potential due to newness of the subject and multifaceted activities in data handling. The key journals specialised in publishing bioinformatics research includes Bioinformatics, Nucleic Acid Research, Genomics, Proteomics, Medical Informatics, to mention few. The *iBIRA* have a description of about 150 journals references, whose domain specially covers the field of bioinformatics. The various information scientists discussed the metadata as well as bibliographic schemes for indexing journals. As for as the article is concern, the XML schema is used to describe journals and articles. The XML schema is compliant with that of MEDLINE XML schema for interoperability of the articles in to *iBIRA*. The DC metadata description for journals is given in Table 2.

#### **4.4 Patents in Bioinformatics**

In the area of bioinformatics research, huge amount is being invested for setting up laboratory for large-scale analysis of data generated through sequencing projects such as Human Genome Project (HGP), creation of database, and development of software tool, etc. Yet the intellectual property protection available in return for those investments has, until recently, seemed elusive. The bioinformatics sector first came to the general public's attention as a result of the HGP. In reality, bioinformatics had been a discipline in all over a decade for these projects. Although the business models have begun to diverge, and the companies started off as essentially database providers. They start intended themselves to recover their investment by charging subscription fees for access to their data. As the volume of data produced grew exponentially, companies specialising in mining the data (such as Lion Biosciences) started emerging. Development of algorithms and software becomes the crucial activities and companies started using the software for analysis and manipulation of data related to proteomic, clinical, and medical data. Perhaps as a result of their proximity to the patent-conscious biotechnology community, bioinformatics companies are increasingly looking towards patent protection in an attempt to capitalise on their investments. Furthermore, the protection

rights lies with the companies, which are involved with the inventions related to drug-screening methods, database rights, confidential information and copyright. The government agencies are very keen to protect the intellectual output through setting up different policies and procedures. The European Patent Office (EPO), United State Patent and Trademark Office (USPTO) and Indian Patent Office are very active in protecting the intellectual output in European region, USA and India respectively.

#### **4.5 Education Institutions for Bioinformatics**

In the present scenario industry and academia both agree that there are not enough scientists adequately trained in bioinformatics or computational biology. This sentiment stems principally from the difficulties in finding employees, graduate students and post-docs with appropriate skills for joining research and/or development teams in this field. The recent drain of academics into industry threatens to reduce ability to provide the training needed to meet the demand of the job markets. Bioinformatics being multidisciplinary field involved various subjects in its daily activities. Because of multidisciplinary curriculum, bioinformatics has addresses inclusion of multidisciplinary people for teaching and research. Further, it addresses the issue of educating new generation of student specialised in bioinformatics techniques and training old biology, computer science and mathematics student to gain more insight of the subject. Altman<sup>28</sup> analysed the curriculum for bioinformatics. According to Altman analysis, it is found that there are two models for training in bioinformatics. In the first model, post-doctoral fellows with core training in a technical field (such as computer science) or in a sub-discipline of biology receive specialised training in computational biology to become a 'computer scientist who specialises in biology' or a 'biologist who specialises in computer science'. While a valuable strategy, the post-doctoral model suffers because it is an expensive way (both in time and resources) to train individuals—learning the 'other' field is in many cases like going back to graduate school. In the second model, the graduate students are trained primarily in bioinformatics or computational biology. Thus, to strengthen the educational background, specialised training centers and educational institutions, are coming up with specialised training program in graduate, post graduate, doctoral and post doctoral level. Further, to train existing graduate and postgraduate student, short-term courses like diploma and certificate programmes are also being run simultaneously to educate with bioinformatics activities. This aspect of bioinformatics has been considered as one of the vital activities in *iBIRA* to integrate the information simultaneously along with other resources of bioinformatics.

#### 4.6 Software Tools in Bioinformatics

There is an increasing need for the development of data analysis software that provides bioinformatics functionalities to biologists without requiring prior knowledge of programming and scripting languages and it is essential that the software tools should fulfill biologists' needs<sup>29</sup>. At the same time, these biologist-friendly tools need to be useable on different operating systems, and they need to provide a natural language description of the results produced in order to state assumptions made in the analysis. As a result of continued demand, new tool are proliferating with the automated facility for sequencing, genotyping and studying gene expression. It indicates that the data generation will be a continued process and generated at ever increasing rates. This quiet revolution in the biological sciences has been enabled by our ability to collect, manage, analyse, and integrate large quantities of data and powerful computational tools are a paramount requirement. Many of these tools are not user friendly and can be platform dependent. The internet allows to access bio-computing and bioinformatics tools from the world wide web (www)

at the click of a button from personal computer. The information available about the software over internet has demanded the standardisation of the basic information in order to enhance the usability and specificity in bioinformatics analysis. Table 2 provides the details of metadata information for software incorporated in the process of development of iBIRA.

#### 5. OBSERVATIONS

The metadata is the key concern for every kind of information exchange in case of web-based resources. Interoperability can only be ensured, when the information description framework is as per the international standard. Bioinformatics is a subject where heterogeneity is one of the key issues where the information resources are available in varied format and media as well as the resources are available in heterogeneous categories. The standardising metadata description element could be one of the solutions in integration of bioinformatics resources. There are varied format and schema used by different kind of activities in integration of resources. Here we have used XML as medium of exchange of information in the development of iBIRA.

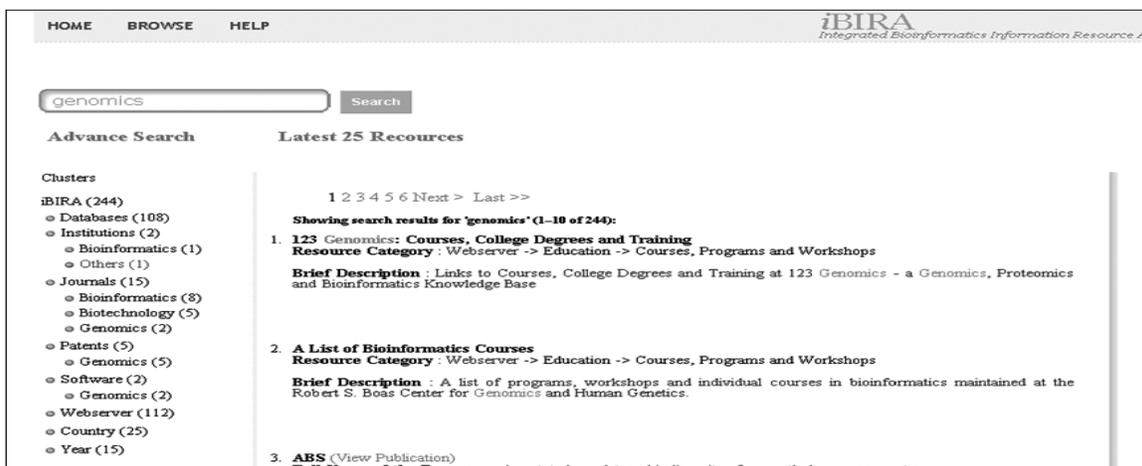


Figure 2. Resource tree of search result on iBIRA search page.

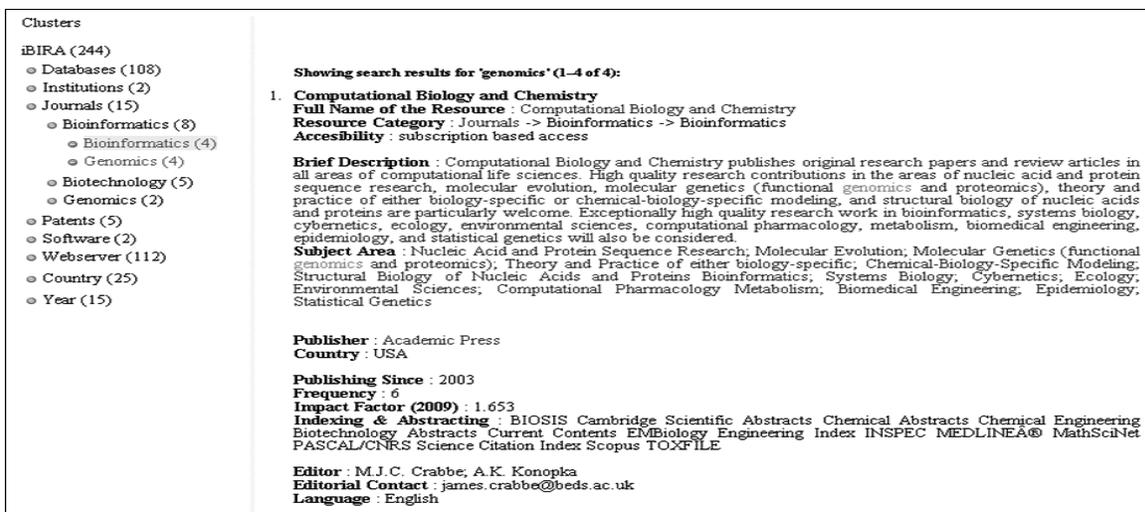


Figure 3. Detailed resource description (journal) on iBIRA.

XML has the potential to facilitate many features, which add value to the migration of electronic journals, database, patents, etc., as may be the case of bioinformatics resources. There are enormous potential for the XML to enable the repurposing of components such as e-journals, databases, patents and many components; powerful XML tools such as XSLT (Extensible Stylesheet Language Transformation) is of central importance.

The resources which have been extracted and database for the information resources has been shown in the Fig. 2 and 3. The Search facility of iBIRA facilitates development of a tree structure of the information distributed in different broad categories. The respective broad categories further navigate with subcategory and land up the option for choosing information of users own interest areas under different broad category (Fig. 2). When a user navigates to individual category, it provides the detailed information as per metadata description framework (Fig. 3).

## 6. CONCLUSIONS

Bioinformatics is one of the subjects where heterogeneity is prominent in respect to availability of information, resource format and resource availability as well as heterogeneity in tasks undertaken by bioinformatics scientists for solving biology related problem. During such activity huge amount of data releases and management of data is one of the important tasks in bioinformatics resulting in development of various databases. Simultaneously scientists are also adopting different computational methods for managing the data and information resources leading to origin of web servers. The research works are being published in journals and conferences paving the path for future research and education for novice bioinformatics community. These aspects of bioinformatics are so fast and active that the searching, finding, locating and accessing the bioinformatics information resources become difficult for this community. iBIRA is an effort to integrate bioinformatics information resources at single platform to provide and facilitate with tailor made information in the area of bioinformatics. The resources in the area of bioinformatics are so heterogeneous, the standardisation of information is becomes essential. Metadata is one aspect which has been found useful in standardisation of resource description.

This aspect used in the development of iBIRA open the future path for further research in this area in order to develop a standard so that interoperability of information may be sought and uniformity may be achieved.

## REFERENCES

1. Lee, J. & Wernerfelt, B. Integrating information from disparate contexts: A theory of semantic interoperability. MIT University USA, 1996. PhD Thesis.
2. Lakshmanan, L.V.S. & Sadri, F. Information integration and the semantic web. *Bull. IEEE Comp. Soc. Tech. Commit. Data Engg.*, 2003.
3. Sheth, A. & Larson, J. Federated database systems for managing distributed heterogeneous and autonomous databases. *ACM Comput. Surveys*, 1990, **22**(3), 183-236.
4. Eccles, J.R. & Saldanha, J.W. Metadata-based generation and management of knowledgebases from molecular biological databases. *Comput. Methods Prog. Biomed*, 1990, **32**(2), 115-23.
5. Kirklin, J.W. & Vicinanza, S.S. Metadata and computer-based patient records. *Annals Thorac. Surg.*, 1999, **68**(3), S23-24.
6. Schiffman, R.N.; *et al.* A design model for computer-based guideline implementation based on information management services. *J. Amer. Med. Inf. Assoc.*, 1999, **12**(6), 99-103.
7. Magda El-Sherbini. Metadata and the future of cataloging. *Library Review*, 2001, **50**(1), 16-27.
8. Iannella R. & Waugh, A. Metadata: Enabling the internet. *In CAUSE97 Conference*, April 1997, Melbourne, .
9. NISO. National Information Standard Organisation, Understanding Metadata. NISO Press, 2004.
10. El-Sherbini, Magda & Klim, G. Metadata and cataloging practices, *Electronic Library*, 2004, **22**(3), 238-48.
11. Miller, P. Metadata for the masses. [www.ariadne.ac.uk/issue5/metadata-masses/](http://www.ariadne.ac.uk/issue5/metadata-masses/).
12. Li; *et al.* A methodology for semantic integration of metadata in bioinformatics data sources. *In 43<sup>rd</sup> ACM Southeast Conference*. 18-20 March 2005, Kennesaw, GA, USA.
13. iBIRA. <http://ibiranet.in>.
14. Hogeweg, P. & Hesper, B. Interactive instruction on population interactions. *Comp. Biol. Med.*, 1978, **8**(4), 319-27.
15. Hogeweg, P. The roots of bioinformatics in theoretical biology. *PLoS Comput. Biol.*, 2011, **7**(3), e1002021.
16. Denn, S.O. & MacMullen, W.J. The ambiguous bioinformatics domain: A conceptual map of information science applications for molecular biology. *In ASIS&T 2002: Proceedings of the 65<sup>th</sup> Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 18-21 November 2002, Philadelphia, PA.
17. Ge, H.; Walhout, A.J.M. & Vidal, M. Integrating OMIC information a bridge between genomics and system biology. *Trends in Genetics*, 2002, **19**(10), 551-60.
18. WorldWideSize.Com, 2012.
19. Haas, L.; *et al.* DiscoveryLink: A system for integrated access to life sciences data sources. *IBM Sys. J.*, 2001, **40**(2), 489-511.

20. Ram, Shri & Rao, N.L. *iBIRA – integrated bioinformatics information resourceome access: Organising the bioinformatics resource. Refer. Serv. Rev.*, 2012, **40**(2), 326-43.
21. Navathe, S. & Patil, U. Genomic and proteomic databases and applications: A challenge for database technology. *In DASFAA 2004: Proc. 9<sup>th</sup> International Conference on Database Systems for Advanced Applications (DASFAA 2004)*, March 2004, Jeju Island, Korea.
22. Hernandez, T. & Kambhampati, S. Integration of biological sources: Current systems and challenges ahead. *SIGMOD Record*, 2004, **33**(3), 51-60.
23. Nucleic acid research databases. <http://nar.oxfordjournals.org/>.
24. Bioinformatics link directory, Canada. <http://www.bioinformatics.ca/>.
25. BioMed Central database. <http://databases.biomedcentral.com/>
26. Robbins, R. Report of the invitational DOE workshop on genome informatics. Universal Academy Press, Tokyo, Japan, 1993.
27. Web server, [http://en.wikipedia.org/wiki/Web\\_server/](http://en.wikipedia.org/wiki/Web_server/).
28. Altman, R.B. A curriculum for bioinformatics: The time is ripe. *Bioinformatics*, 1998, **47**(7), 549-50.
29. Kumar, S. & Dudley, J. Bioinformatics software for biologists in the genomic era. *Bioinformatics*, 2007, **23**(14), 1713-17.

### About the Authors

**Mr Shri Ram** is working as Deputy Librarian at Jaypee University of Information Technology, Solan since 2002. Earlier, he has worked at Indian Council of Medical Research, New Delhi and CEMCA India in various capacities. He has completed Master's in Anthropology and MLIS. He is pursuing his PhD (LIS) from Osmania University, Hyderabad. Recently, he has been awarded with Commonwealth Professional Fellowship at University of East London, UK. He has published more than 25 papers in different peer reviewed international and national journals and conferences.

**Prof. N. Laxman Rao** after serving more than 40 years of teaching and research in the area of library and information science (LIS), has recently superannuated as UGC Emeritus Fellow, Department of LIS, Osmania University, Hyderabad. He has more than 70 research publications, more than 150 conference deliberations, and five edited books to his credit. He is member of various advisory boards for selection, recruitment, and other government committees for accreditation of library and information services.