Project no. **MOBI-CT-2004-510971**

Project acronym **BulRMCNet**

Project title: **Bulgarian Network of Research Mobility Centres**

Instrument **SIXTH FRAMEWORK PROGRAMME**

FP6-2003-MOBILITY-CENTRES

EUROPEAN NETWORK OF MOBILITY CENTRES

Thematic Priority **Human Resources & Mobility**

Deliverable D 2.1 Portal Architecture Specification

Due date of deliverable: January 2005
Actual submission date: January 2005

Start date of project: 01.09.2004 Duration: 36 months

Organisation name of lead contractor for this deliverable: Sofia University

Revision 2

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

TABLE OF CONTENTS

1. Introduction

The objective of Work Package 2 of BulRMCNet project is to create a National Research Mobility web Portal that will support on a national level via Internet the provision of information and services to mobile researchers.

This document represents the first output of Work Package 2 focused on the Portal Architecture Specification. According to the specific tasks set in Annex I of BulRMCNet project, the document aims at providing a description of the National Research Mobility web portal – the data requirements and system specification. As the Portal is considered to mirror the European Researcher's Mobility Portal (RMP), the consideretions of its designers, namely the specifications constituted by Reggiani (a project deliverable ruled by EC Framework Contract No RTD-JRC / 00-04), are taken into account.

This document is structured in three parts providing analysis related to the European RMP requirements, presenting the designers view on portal architecture specification and the web site architecture to be followed. It provides a general idea on the National RMP – its services, databases, programming languages and technologies to be used, as well as organisation of the web accessible content.

2. European RMP Requirements Analysis

The National RMP is considered to follow the structure and graphical design of the Eurpean RMP and to deliver data for it (using RJOE, data-exchange software, delivered by Reggiani e-Division). As participant in the ERA MORE Network, Sofia University is going to commit to the requirements of the European Qality Charter, and to build the National RMP according to the charter. Subsequently, the main issues to ber taken into account include:

- Interoperability
- Ownership
- Quality assurance
- Accessibility
- Principle of non-discrimination
- Confidentiality

2.1 General

For building the Architecture of the National RMP were analysed the European Researcher's Mobility Portal requirements. It was considered that the National RMP should fulfil the following:

- the content should be available in English
- the page should have no frames structure
- clear and suitable set of meta-data (web-page description)

2.2 Data Exchange

The National RMP has a dual role – to be an input point for collecting data in the respective areas at national level and to to deliver data to the European RMP. For the data exchange with the European RMP will be used the Research Jobs Opportunity Editor (RJOE). Subsequently, the data will be delivered in a format compatible with RJOE. The data exchange scheme between the National RMP and the European RMP, as well as between the National RMP and the respective services at regional research mobility centers in Bulgaria is shown on figure 2.1.

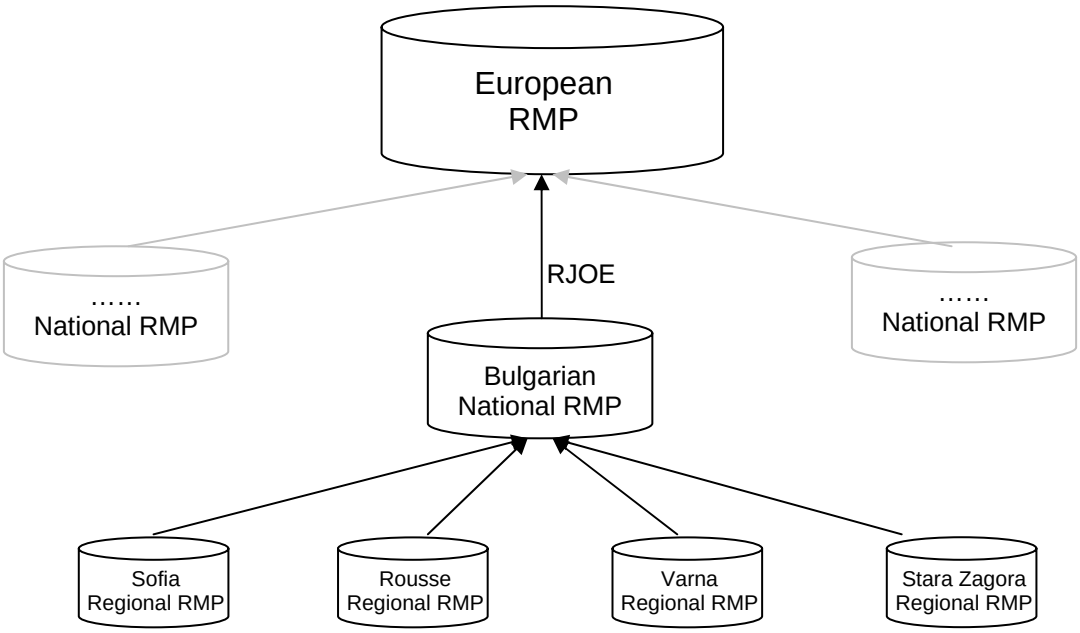

Figure 2.1 Data exchange scheme

2.2.1. Interoperability

It is standartisation on core sets of inter-operational protocols – the most important part of data exchanging and it is based on:

- Standard technologies;
- Standard-based formats.

2.2.2. Content Quality

The data published on the National RMP should be accurate and up-to-date. For this purpose Sofia University as the appointed bridgehead organisation in Bulgaria will contact trusted sourses in the state administartion and agencies, will use data and information from well recignised sources in Bulgaria. It is essential to note also that a big part of the content will be provided by research organisations and researcheres themselves, for which content the University will have no responsibility.

2.3 Information categories

The portal collects different types of information that can be gouped intwo main:

- **Unstructured data** - all the data without any particular format.
- **Structured data** - those data with an agreed format as curriculum vitae, or job vacancies.

3. Portal Architecture Specification

3.1 Actors of the System

This section summarises the main requirements related to the information handled and the different actors of the system. Generically, *Users* will refer to all people who will use the National Portal Services.

For the purposes of this docuemnt will be distinguished the following groups of users:

- Contributors - Refer to users that provide information to the National RMP. The contributors can be:
  - Research institutions, Universities;
  - Companies with research orientation.
- Researchers - Refer to qualified users, who are searching for jobs opportunities and/or are publishing their CVs on the National RMP.
- Visitors- Refer to Occasional users. They can be registered and not registered.

3.2 System Services

The main services that will be provided at the National RMP include:

- Brokering Service
- Communication Service
  - External - Data Exchange (RJOE)
- Internal
  - Forum
  - Mailing List
  - Chat
- Notification Service
- Publishing Service
  - CV Publishing
  - Vacancy and Job opportunity Publishing
- Logging and Activity Tracking Service

3.3 Database Architechture

Database Access Models

There have been two distinct approaches identified:

- central store
- distributed access

Central Data Store

In this model all data will be held in central database within the National RMP. Initially, the information will be captured from providers/contributors and then forwarded to the central database. The information is then in one definitive location and can be accessed from all services within the National RMP.

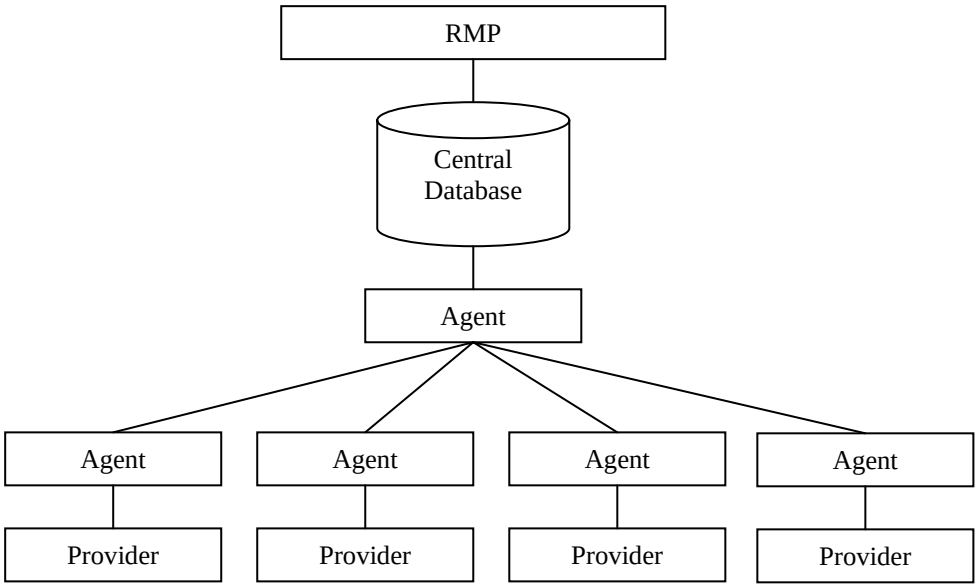

Figure 3.1 Central Data Store

Distributed Data Store

In this model each partner captures and stores information in its own data warehouse. In order to get information will be used agents which will pack and retrieve data in correct format from the different providers/contributors.

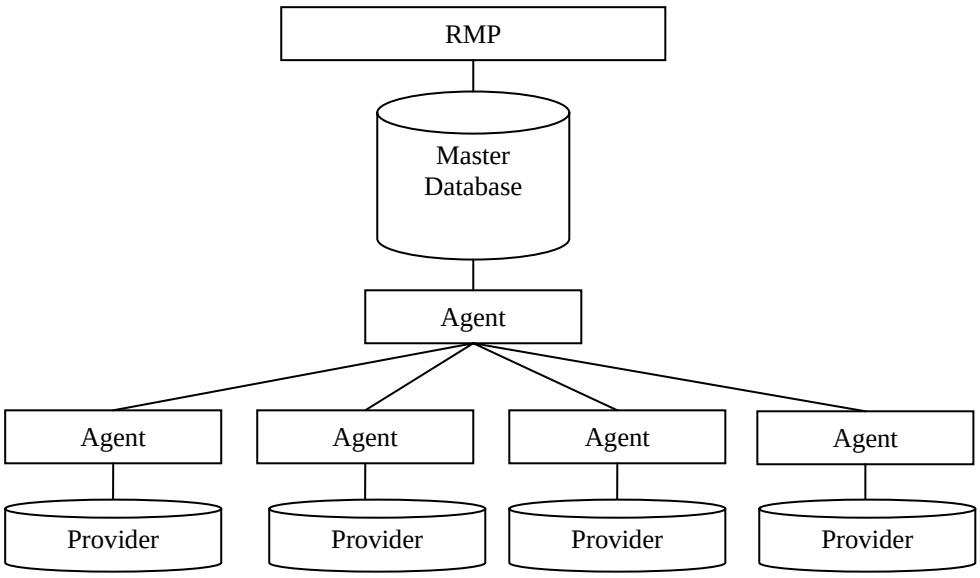

Figure 3.2 Distributed Data Store

	Advantages	Disadvantages
Central Data Store	Only one access point	Single point of failure
	Overcomes the problem of matching all available vacancies and job opportunities published from different provides or contributors	Increase Cost, in terms of setting up, running and managing such a database.
	Lower consumption of net bandwidth.	
Distributed Data Store	Data remains in the provider/contributor that created it and therefore will always be current	Higher consumption of net bandwidth.
	Failure of single database is not crucial for overall system work	One failure leads to data base inconsistancy.

A centralised repository will be created to store the relevant data. The storage will not imply duplication of data on the RMP.

All RMP's own data (job offers, profiles, search profiles, etc.), will be stored on the central RMP's repository.

3.4 Ontology

Ontology and namespaces might be different domains requiring different data models to be stored. For data description will be used the Orthelius classification for research fields recomended by the European Commission within ERA MORE technical meetings.

The present approach is to define ontology concepts using XML document type definitions (DTDs). DTDs within the W3 consortium aim to specify a schema definition language that describes the structure of XML documents in a machine-readable form.

An XML schema in the ontology of the National RMP describes the concepts, the national language bindings, the different forms for brokering vacancies and job opportunities. XML schemas will therefore be the building blocks of the ontology. Since XML and current

schema languages are quite flexible in their expressiveness it could be considered that this approach will allow to cover most of the needed service domains.

There are three types of data which have to be described

- Researchers' curriculum vitae.
- Job Vacancies
- Practical Information

3.5 Portal Security Architecture

The security is essential part of the portal architecture because of the portal nature as a storage of different data.

Logging is an essential part of documenting system activity enabling to reconstruct past systems states after unusual events. A good logging policy and implementation helps to avoid future malfunctions by analysing past failures. Usually there is a natural trade-off between the kind and amount of data logged and the system performance penalty and log backup procedures.

It should be considered that important logging data should be moved from operational systems to secondary storage for further archival. For this purpose archival mechanisms based on tape-drives and CD-RW drives should be considered as potential options.

Constant system monitoring is another vital activity needed for proper operation. Besides control of system resources such as disk and data space, network problems concerning IP, ICMP, and DNS should be controlled in an automated manner. Several tools are available that constantly check systems for unusual conditions and offer mean to notify system administrators of potential problems.

Since in a real deployment scenario a user's sensitive information will be hosted by the National RMP, protection of information may be of particular concern. Internet users are very sensitive to downtimes of particular services. Special care must be taken to keep possible downtimes of the National RMP as small as possible. Since for the moment Linux based operating system is selected for hosting the National RMP a packet-filtering firewall could significantly simplify the security-related administration of a the Portal.

3.6 Portal Platform Architecture

The National RMP Architecture provides the framework to host the initially delivered and all subsequently delivered content in a standard, easily maintainable manner. It utilizes and provides support for industry standards for communications and interoperability.

3.6.1. Operating System

A review and comparative analysis of several operating systems has been made, and comparisson of their advantages and disadvantages. Subsequently, a decision was taken based on the following criteria:

- **Stability and reliability:** BSD systems are extremely robust. Linux is also very stable system. Both can stay up for years. Some of the major issues have been fixed in Windows 2000 Server but still the main problems like memory leaks and filesystem fragmentation remain.
- **Security:** some investigations show that the ration of the hacked sites based on Windows, Linux and Unix like operation systems is 5:2:1.
- **Performance:** BSD systems outperform other systems when running on equivalent hardware. Linux performs well for most applications; however the performance is not optimal under heavy network load. Windows is adequate for routine desktop apps, but it is unable to handle heavy network loads.
- **Total cost of ownership:** Most BSD and Linux systems are free. They are much more secure than Windows OS. Windows can be made more secure than its defaults but that brings additional costs. The more Windows systems are asked to do, the less stable they will be compared to UNIX counterparts, performing a similar diversity of tasks. Troubleshooting unstable Windows systems may be the most labour intensive task.

Based on the above analysis it could be concluded that the most appropriate Operating System for the purposes of the National RMP is BSD (OpenBSD or FreeBSD distribution). Linux can also be used as a second choice.

3.6.2. Database Server

Considering the results from comparative analysis depicted in the table bellow a decision was taken to use MySQL RDMS server. It meets the stated requirements as a reasonable choice for the database server of the National RMP. Structured data in a MySQL database can be accessed by Internet-based applications through a JDBC interface and standard SQL. Stored procedures may be written either in PL SQL.

	PgSQL	**MySQL**	**Commerical**
Data Integrity			
ACID compliance	X	X	X
Row-level locking	X	X	X
Partial rollbacks	X	X	X
Advanced Features			

Stored procedures	X	In 5.0	X
Views	X	In 5.0	X
Triggers	X	In 5.1	X
Sequences	X	In 5.1	X
Cursors	X	In 5.0	X
Indexes			
Single column	X	X	X
Multi-column	X	X	X
Primary key	X	X	X
Full text	X	X	X
Replication			
Single-master	X*	X	X
Multi-master	X**		X
Interface Methods			
ODBC/JDBC	X	X	X
C/C++, Java	X	X	X

* Open Source, but comes from another vendor

** Solutions exist, but they are commercial.

3.6.3. Application Server

Jakarta Tomcat is an open-source application server that is produced by the Apache Software Foundation. Tomcat is the reference implementation for the Java Servlet and Java Server Pages technologies.

The advantages of Jakarta Tomcat Application Server are:

- Keeps the system secure and stable.
- Makes it easy for developers to add applications.
- Provides access to databases over JDBC and ODBC interfaces to RDBMS such as Oracle, MySQL, PostgreSQL, DB2, etc.
- Provides redundancy so that applications could continue running despite the outage of a single server.
- Open Source and cost free

3.7 Programming Language and Technologies

The system will be designed according to the traditional 3-tiered Web architecture model and all the client-server communication is to be handled through a web server.

The software implementation will use Java freeware technologies thus keeping the system open, portable, and easy to integrate at a reasonable cost.

The business layer will be implemented by JSP and servlets following JSP model 2 so it can run not only under any Java application server but as well as under JSP engines such as Tomcat.

The presentation layer will use HTML/CSS and JavaScript but can be implemented for any markup language by adding new XSL templates.

Here are the core programming languages and technologies we will use:

Java

The Java Programming Language is a general-purpose, concurrent, strongly typed, class-based object-oriented, platform independent language. It is normally compiled to the bytecode instruction set and binary format defined in the Java Virtual Machine Specification.

XML

XML stands for Extensible Mark-up Language. XML is a framework language (or meta-language) that is a subset of SGML. It conserves essentially the basic elements and separates structure, content, and presentation. The relationship between HTML, SGML and XML is the following: All three are mark-up languages. XML is a proper subset of SGML. Both are meta-languages. HTML is an application based on SGML (and by extension of XML).

XML offers a variety of advantages and can be used as the base on which to build other utilities and applications, in particular:

- Metadata management
- Database integration
- Localization and Unicode
- Document management using the DOM model and DTDs / XSD

XML will be used as tenabling format for the interaction of the system with other applications and software tools. It is associated with XML schemas or DTD that defines the rules, the grammar to be respected and enforces a document that refers to it to conform to the rules it defines.

XSL

XSL is a language for expressing stylesheets. It consists of two parts:

- a language for transforming XML documents, and
- An XML vocabulary for specifying formatting semantics.

An XSL stylesheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary.

XSLT

An XSL stylesheet processor accepts a document or data in XML and an XSL stylesheet and produces the presentation of that XML source content that was intended by the designer of that stylesheet. There are two aspects of this presentation process: first, constructing a result tree from the XML source tree and second, interpreting the result tree to produce formatted results suitable for presentation on a display, on paper, in speech, or onto other media. The first aspect is called tree transformation and the second is called formatting. The process of formatting is performed by the formatter. This formatter may simply be a rendering engine inside a browser. Tree transformation allows the structure of the result tree to be significantly different from the structure of the source tree. For example, one could add a table of contents as a filtered selection of an original source document, or one could rearrange source data into a sorted tabular presentation. In constructing the result tree, the tree transformation process also adds the information necessary to format that result tree. Formatting is enabled by including formatting semantics in the result tree. Formatting semantics are expressed in terms of a catalog of classes of formatting objects. The nodes of the result tree are formatting objects. The classes of formatting objects denote typographic abstractions such as page, paragraph, table, and so forth. Finer control over the presentation of these abstractions is provided by a set of formatting properties, such as those controlling indents, word and letterspacing, and widow, orphan, and hyphenation control. In XSL, the classes of formatting objects and formatting properties provide the vocabulary for expressing presentation intent. The XSL processing model is intended to be conceptual only. An implementation is not mandated to provide these as separate processes. Furthermore, implementations are free to process the source document in any way that produces the same result as if it were processed using the conceptual XSL processing model.

Java Server Pages

Java Server Pages (JSP) is Sun's solution for developing dynamic web sites. JSP provides excellent server side scripting support for creating database driven web applications. JSP enables the developers to directly insert java code into jsp file, this makes the

development process very simple and its maintenance also becomes very easy. JSP is efficient, it loads into the web servers memory on receiving the request very first time and the subsequent calls are served within a very short period of time.

In today's environment most web site servers dynamic pages are based on user request. Database is very convenient way to store the data of users and other things. JDBC provides excellent database connectivity in heterogeneous database environment. Using JSP and JDBC it is very easy to develop database driven web application.

Java is known for its characteristic of "write once, run anywhere." JSP is an independent platform.

JavaScript

Client-side scripting languages are well suited for the implementation of simple interfaces based upon HTML. Scripting languages will be used in order to add interactivity to HTML pages and implement the validation of field information in the DMS or Web portal forms. Such strategy lowers the server workload and avoids classic Internet delays when users submit searches with wrong parameters or criteria. With scripting languages the browser itself is aware of some possible inconsistencies.

For achieving cross-browser compatibility and addressing at least MS Internet Explorer and Netscape Navigator browsers, we will use the common subset of JavaScript and JScript languages.

HTML

This W3C standard is universally supported and allows for dissemination of information to a broad audience. HTML itself is simple, but several extensions are also possible. Improved interactivity, 3D graphics, database interaction, state-of-the-art GUIs can easily be integrated with HTML pages.

The Dissemination System interface will generally use the HTML 4.01 standard. This version is fully supported by fourth-generation Web browsers. HTML 4.01 helps developing a better presentation of content and offers enhanced interactivity.

The HTML pages produced by the Publishing Scripts will be well-formed, i.e. they will conform to the rules of XML.

4. Web Site Architecture

In the project BulRMCNet is considered the National Research Mobility Portal to mirror the European Researcher's Mobility Portal.

The National RMP is published on http://www.eracareers-bg.net/. The simplest solution relies obviously on an accurate selection of the web site requirements:

- available in English, as static pages;
- without any access restriction;
- without frames;
- clearly with a suitable set of meta-data;
- each being page-specific when applicable.

The RMP Structure is given on Figure 4.1. It includes the following categories:

- Organisations – foreseeing the possibility for registering an organisation, posting research, publishing job vacancies, and finding the ideal candidate
- Reseachers –protected area for Registered Users with opportunity to publish the CV and search for research opportunities.
- Mobility Centers – access points to the portals of the Regional Resesarch Mobility Centers (RMC) in Bulgaria, as well as to the bridgehead organisations all around Europe.
- Research & Universities – provides general information on the Bulgarian research landscape and opportunity to explore Research Institutions and Universities in Bulgaria
- Relocation & Living in Bulgaria – provides practical information on: Visa and Residency in Bulgaria, Work & Living in Bulgaria, Health & Social Security in Bulgaria, Taxes and Financial Issues, and Educational Opportunities.
- Tourist Destinations – opportunity to explore Bulgaria: History, Culture, Folklore, Cuisine, Treasures, Nature, etc.
- Other Career Resourses – access point to other on-line job offers – not related tothe researchers profession.
- Discussion Forum – contact with other people
-

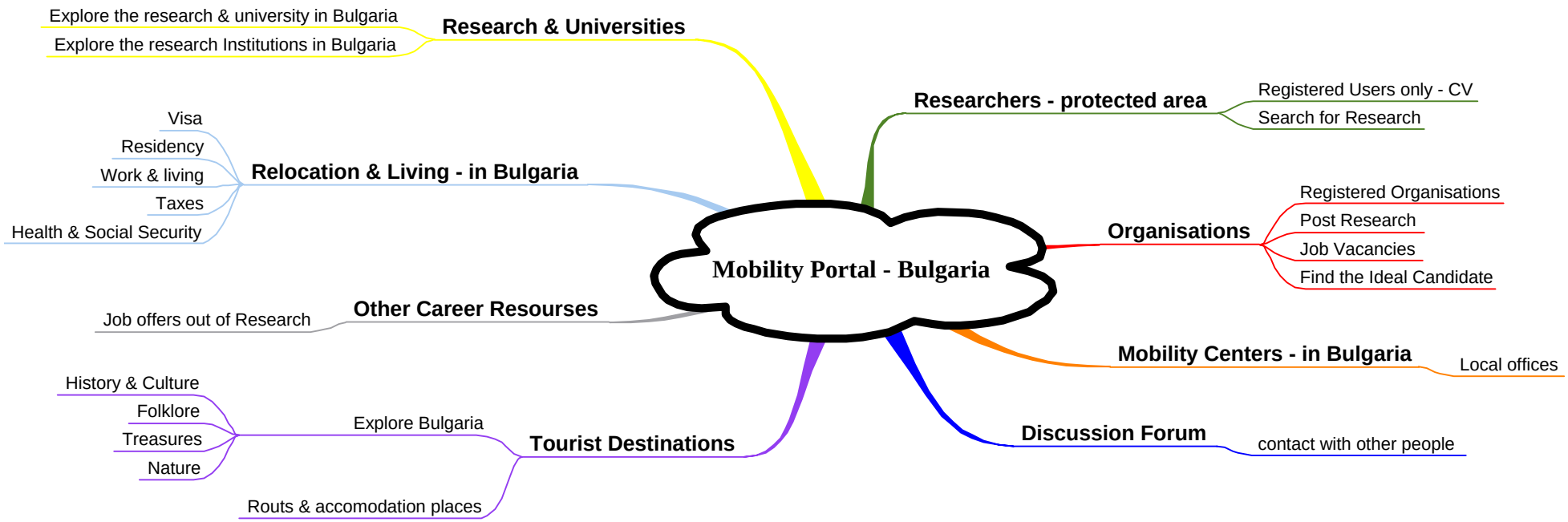

Figure 4.1 – National RMP Structure

4.1 Page Structure

Figure 4.2 is the graphical lay out proposed for the portal implementation. It is based on the use of the template of the European RMP.

- The **European RMP header is** on the top of the page – it includes the last version of the standard IPG header including the following elements: banner, path and service tools.
- The **main menu on the left**

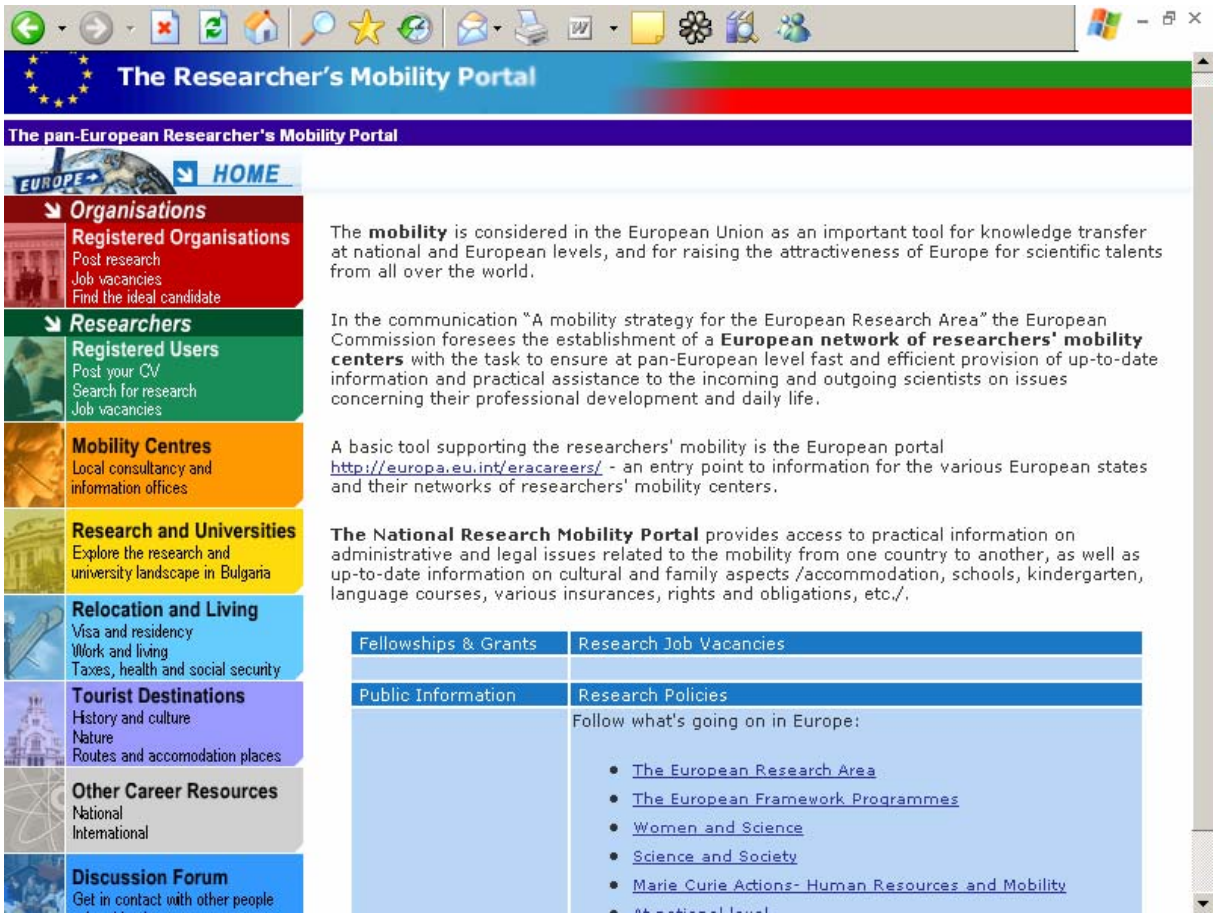

Figure 4.2 – Page Structure

4.2 Main Menu

The main areas of the National RMP related to the main areas of European RMP are:

- Organisations,
- Researchers,
- Regional Mobility Centers,
- Other Career Resources.

Additional areas of the National Research Mobility Portal are:

- Research & Universities,
- Relocation & Living,
- Tourist Destinations,
- Discussion Forum.

The page Mobility Centers contains a map of Bulgaria showing the location of the Regional Research Mobility Centers – members of the Bulgarian Network of RMC (figure 4.3).

The web pages providing Practical Information for Bulgaria (Research and Universities, Relocation and Living, and Tourist Destinations) are organised in similar manner – including interactive map of Bulgaria with regional contact points. In addition, they will contain centralised information on administrative procedures and legislation in the respective areas and links to the authorities and service providers.

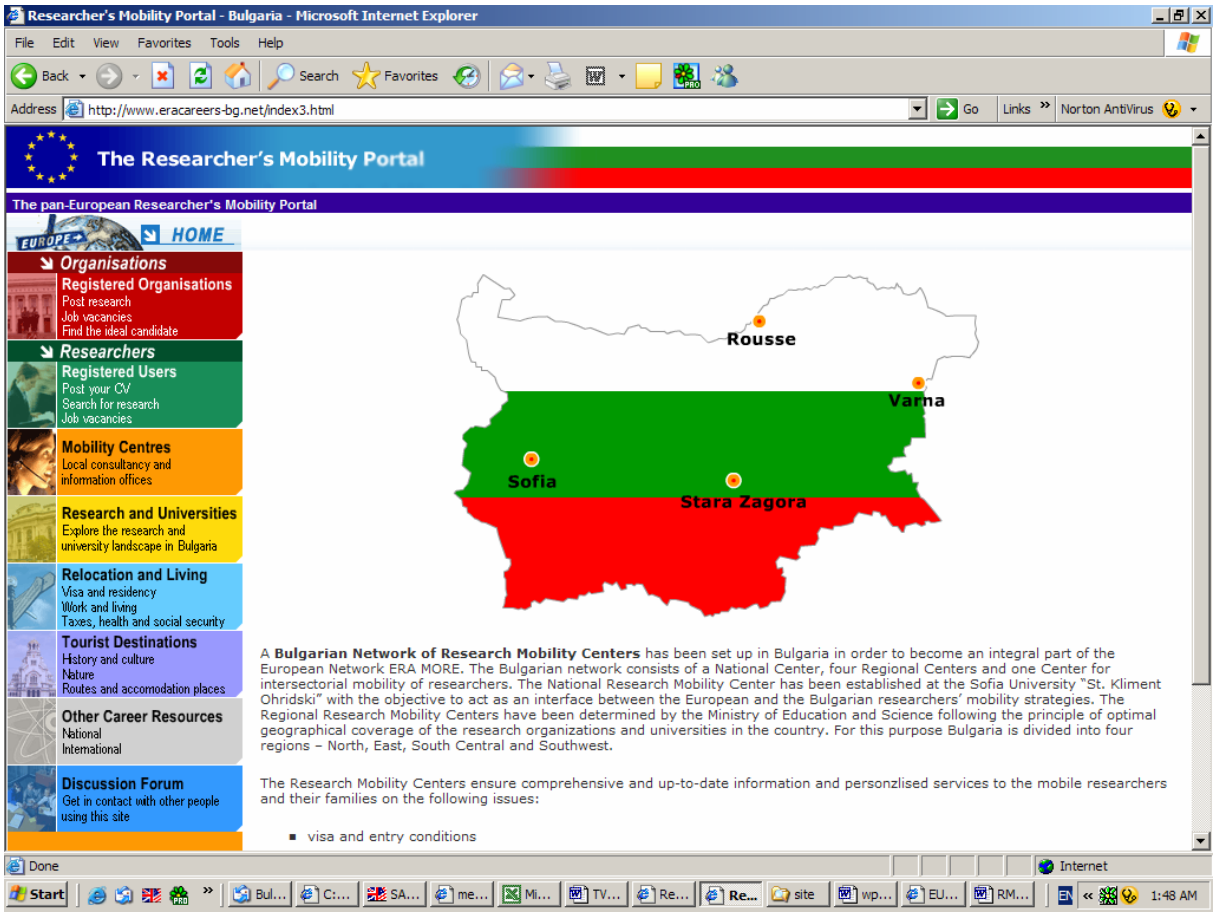

Figure 4.3 – Bulgarian Mobility Centre

5. Next steps

1. Working prototype of the National RMP will be created with Information database and services concerning Researchers Mobility in English.
2. For the data exchange with the European portal will be used the tool RJOE provided by the European Commission.

6. Conclusion

The web portal will have a key role for the effective cooperation and coordination of the RMC in Bulgaria as well as to all organizations, participating in such mobility activities. It will also be effectively linked to the European Researcher's Mobility Portal. Therefore, its design will mirror that of the European one. Subsequently, the portal will provide a set of practical information to mobile researchers and news of interest to the scientific community, as well as opportunity to search for research grants and/or job vacancies.