

## Четыре измерения метаданных: проект RUSSIAinfo

### The Four Dimensions of Metadata: Case RUSSIAinfo

### Чотири виміри метаданих: проект RUSSIAinfo

*Таня Пурсиайнен*

*Университет Хельсинки, Хельсинки, Финляндия*

*Tanja Pursiainen*

*University of Helsinki, Helsinki, Finland*

*Таня Пурсиайнен*

*Університет Хельсінкі, Хельсінкі, Фінляндія*

Каждый, кто занимается разработкой онлайн-информационных услуг, рано или поздно придет к рассмотрению вопросов метаданных с 4-х точек зрения: взаимодействия, использования, разнообразия ресурсов и простоты каталогизации.

В докладе анализируются процессы построения метаданных и показано, как можно дополнять взятый за основу стандарт Дублинского ядра, и каким образом творческий подход к доработкам элементов ДЯ значительно улучшает использование, не мешая при этом взаимодействию. В докладе указывается на кажущуюся недостаточность ДЯ для описания многоязычных ресурсов, и предлагаются способы разрешения этой проблемы. Проект RUSSIAinfo ([www.russiainfo.org](http://www.russiainfo.org)) взят для иллюстрации перечисленных в докладе проблем.

Everyone designing on-line information services sooner or later comes to consider the issue of metadata from the four angles: interoperability, usability, resource diversity, and ease of cataloguing.

The paper analyses the process of metadata construction. It demonstrates, how the Dublin Core standard, used as the basis, can be augmented, and how creative use of DC element refinements considerably improves usability without hindering interoperability. The paper points out the seeming insufficiency of DC for multilingual resources' description, at the same time suggesting means to deal with the problem. The RUSSIAinfo project ([www.russiainfo.org](http://www.russiainfo.org)) is used to illustrate the points stated in the paper.

Кожен, хто займається розробкою онлайн-інформаційних послуг, рано чи пізно прийде до розгляду питань метаданих з 4-х точок зору: взаємодії, використання, різноманітності ресурсів і простоти каталогізації.

В докладі проаналізовано процеси побудови метаданих і показано, як можна доповнювати взятий за основу стандарт Дублінського ядра, і яким чином творчий підхід до доробок елементів ДЯ значно покращує використання без шкоди для взаємодії. В докладі вказано на уявну недостатність ДЯ для опису багатомовних ресурсів, і запропоновано способи вирішення цієї проблеми. Проект RUSSIAinfo ([www.russiainfo.org](http://www.russiainfo.org)) взято для ілюстрації перелічених у докладі проблем.

Everyone designing on-line information services discovers that the issue of metadata has to be approached at least from four different perspectives: interoperability, resource diversity, usability, and resource management. This paper analyses, from a managerial point of view, the process of metadata construction for the RUSSIAinfo ([www.russiainfo.org](http://www.russiainfo.org)) information service. ruSSIAinfo is a Finnish national information service providing quality resources on Russia. It is mainly targeted at scholars, government officials as well as media and business communities, and concentrates on providing in-depth, rather than time sensitive information.

#### **Nobody has to create metadata which is already created**

The Dublin Core standard has been largely used in recent years for description of web-based resources. Standardized metadata is recognized as prerequisite of interoperability of information services: same resource, once described, can be «borrowed» by other services, thus enhancing the opportunities of an end user to find the resource, and relieving information services' designers of unnecessary work. The standard metadata encoded in standard way, such as for instance RDF/XML schemas, can be also used in services with other metadata formats through metadata conversion programs, which there are plenty even now, and more being created all the time.

In the RUSSIAinfo project the Dublin Core was selected a metadata format first of all for the purpose of interoperability, so that the service can become part of larger services and exchange metadata with similar services. For example, RUSSIAinfo's resources can be used as such in the Virtual Library of Finland, where RUSSIAinfo fits perfectly as a subject gateway «Russian studies». The Virtual Library of Finland's resources can be found at [www.linkkitalo.fi](http://www.linkkitalo.fi).

The Virtual Library of Finland, in its turn, is part of Renardus, the European service that provides a trusted source of selected, high quality Internet resources for those teaching, learning and researching in higher education in Europe.

The smaller constituency, which would considerably augment RUSSIAinfo, is Karelia Information Service, at present under construction in the Joensuu Polytechnic. As the Karelia Information Service project started later than RUSSIAinfo, it has an advantage of using the schema developed in the RUSSIAinfo.

Also in Russia there are projects that describe Internet resources in the Dublin Core, for example collections of on-line magazines. Needless to say, with the help of simple conversion program these resources too can be made available through RUSSIAinfo.

Out of complete set of Dublin Core metadata elements RUSSIAinfo has accepted 13: Title, Subject (repeated as many times as particular resource requires), Creator, Description, Rights, Publisher, Relation, Date, Coverage (both temporal and spatial), Identifier, Type, Format and Language. The use of these metadata elements in the RUSSIAinfo is described in more details further in this paper.

### **Resource diversity dictates new requirements**

Element refinements of the Dublin Core elements were chosen on two considerations: types of resources we are describing and usability questions. RUSSIAinfo provides access to resources of types such as single full text documents (in PDF and Word format), collections of full text documents, single web pages, collections of web pages (web sites, portals and on-line databases).

The degree of granularity of resource description varies from single documents to big document collections according to the relevance of particular resources to our target audience. For example, a RUSSIAinfo user can sooner be interested in «Dynamics of foreign investments in Kaliningrad region in 2005» than in the «General economic trends of North-West Russia».

Some of the RUSSIAinfo's resources, documents in both PDF and HTML formats, are published simultaneously in several languages. In order to convey this information in Dublin Core language, one can either catalogue all language versions separately, or use element refinements, such as for Title: alternative, or for Relation: Has Version.

Let's consider the example of Economic Monitoring Reports on North-West Russia that are published in three languages: English, Finnish and Russian.

1. In English: Leningrad Region in the first half of 2003: Biannual monitoring review.

ID: [http://www.economicmonitoring.com/data/reports/7\\_2\\_en.pdf](http://www.economicmonitoring.com/data/reports/7_2_en.pdf)

2. In Finnish: Leningradin alue vuoden 2003 alkupuoliskolla: Talouden puolivuotiskatsaus

ID: [http://www.hkkk.fi/netcomm/ImgLib/2/72/Lenobl\\_Nov03\\_FIN.pdf](http://www.hkkk.fi/netcomm/ImgLib/2/72/Lenobl_Nov03_FIN.pdf)

3. In Russian: ЛЕНИНГРАДСКАЯ ОБЛАСТЬ в первом полугодии 2003 года: Полугодовой аналитический обзор

ID: [http://www.economicmonitoring.com/data/reports/7\\_2\\_ru.pdf](http://www.economicmonitoring.com/data/reports/7_2_ru.pdf)

Each language version has its own ID, which means that if we'd employed Title:alternative, we'd be in trouble with referencing each alternative title with the correct ID. Instead it was decided to use element refinement Has Version for Relation field, describe only English language report, and in the field Relation: Has Version provide URL:s of Finnish and Russian versions, marking them for languages. The Dublin Core Documentation does not directly recommend to use this element refinement for language versions, but neither does it forbid such use. So for each set of three documents we only have one record. To user it looks like this:

[Leningrad Region in the first half of 2003: Biannual monitoring review.](#)

Regular, comprehensive and systematic reporting on the socio-economic development in the region.

[Finnish language version](#)

[Russian language version](#)

Some relevant web sites also have more than one equal language version. Those are treated in the same manner.

Another case RUSSIAinfo considered as «special» is the following: same resource is published in several formats by several media. For example, a researcher publishes an article in a magazine which exists on paper only, and by agreement with that magazine the article is published on his/hers research institute's web site.

The solution for this case was to use another element refinement for Relation: Is Format Of. That is, the record is created for the web version of the article, and in the field Relation: Is Format Of the name of the original publishing venue is provided. Thus the end user can easily compile source data, if the article is, for instance, quoted in his/her work.

For cases, when RUSSIAinfo resources are a part of a bigger constituency (a web site of a research organization, a series of on-line publications, one publication produced within the framework of a research program), another refinement of Dublin Core Relation element is used: Is Part Of

### **The user is always right**

After issues of interoperability and resource diversity were dealt with, the next step in the process of metadata design in the RUSSIAinfo project was matching the fields with user requirements. The feedback that was received

from the potential users of the service revealed the need for «meta» information not accounted for in the Dublin Core format. Thus two new fields were added to the RUSSIAinfo's metadata: «Country of origin» (country, where the document was published) and «auxiliary subject», an element which semantic field intersects with both that of «Subject» and «Type». These auxiliary classes for RUSSIAinfo are: news, statistics, reference books, legislation, public institutions, research institutes and organizations.

Obviously, this data will be «lost» for other services in conversion, but not necessarily in all other services. For example, the above mentioned Renardus service also has the field «Country» (the country where the resource originates from).

Research shows that total majority of users tend to rely on the most basic features of information services and expect consistency in them. This is why special attention was paid to the order of appearance of metadata fields, as well as for which fields should hyperlinks be provided.

One more issue that was closely looked into in connection with usability design was mapping of Dublin Core terms into human languages, in the RUSSIAinfo case English and Finnish.

As a result, the short record the user views looks like this:

[Nuclear Safety](#)

Nuclear safety — a prioritized area for the international collaboration in the Barents Region. Brief review of the problem. Links for further reading.

web page. English.2004

[view full record](#)

The link to full record encourages the user to follow it. This can be considered a part of users' information education: once introduced to a complete metadata of the resource, the user will hardly go without it in the future. It might even prompt the move to use advanced search features.

The full record in the RUSSIAinfo looks like this:

Name	<a href="#">Nuclear Safety</a>
Creator	Barents information service project
Subject	Politics, Environment
Subject (key words)	politics, security policy, radioactivity, ecology, environment
Description	Nuclear safety — a prioritized area for the international collaboration in the Barents Region. Brief review of the problem. Links for further reading.
Publisher	Barents information service project
Rights	Barents information service project
Date of publication	2004
Country of origin	Finland
Covers region	North-West Russia
Covers time	1990 — 2000
Language	English
Type	Text, article
Format	web page
Located at	<a href="http://www.barentsinfo.org/?deptid=15865">http://www.barentsinfo.org/?deptid=15865</a>
Published at	<a href="http://www.barentsinfo.org">www.barentsinfo.org</a>

### **Metadata has to be easily managed**

The metadata creation is a time consuming, expensive work. This means, that the editor for metadata creation has to be visual and easily updatable. Like most metadata editors, the RUSSIAinfo's editor gives a chance to automatically extract metadata from a resource being described (which isn't much help in this case), as well as to select key words from thesauri. One important feature of the RUSSIAinfo's editor is the template mode: every record can be used as template when there are similarities in content of resources to be described.

Many ready made tools for extracting «generic» metadata were tested in the RUSSIAinfo project, but none was taking into use. These tools do not work with PDF documents, which are the main bulk of our resources (unless, of course, a PDF document was metadescribed when created, but these are scarce). Neither do these tools work with the Russian language resources, which are, naturally, also plentiful in the RUSSIAinfo information service.

In the development plans of the RUSSIAinfo's Dublin Core editor there is, first of all, the storage of new resources, automatically detected by RUSSIAinfo on the providers' web sites, ready for metataging.

## References

1. Dublin Core Metadata Initiative documents, <http://www.dublincore.org/>
2. ed. Jacobs, N. and Huxley, L. (2004) *Online Information Services in the Social Sciences*.
3. Borgman, C. (2003) Designing Digital Libraries for Usability, in *Digital Library Use: Social Practice in Design and Evaluation*, ed. Peterson Bishop, A. at all.
4. Howarth, L. C. (2004). «Metadata Schemas for Subject Gateways», *International Cataloguing and Bibliographic Control*, 33(1), 8-12.
5. Day, M. (2001) «Metadata in a nutshell», *Information Europe*, 6 (2): 11.