

Проблемы географической привязки цифровых объектов в электронных библиотеках *

© О.Л. Жижимов¹, Н.А. Мазов²

¹Институт вычислительных технологий СО РАН, г. Новосибирск

²Институт нефтегазовой геологии и геофизики им. академика А.А. Трофимука СО РАН, г. Новосибирск

zhizhim@sbras.ru, MazovNA@ipgg.nsc.ru

Аннотация

Рассматриваются вопросы, связанные с географическим аспектом цифровых объектов электронных библиотек. Обсуждаются возможности их географической привязки как для контента, так и для контекста. Рассматриваются распространенные схемы данных в части их поддержки географического аспекта информации. При этом основной акцент делается на координатный способ реализации географической привязки. Обсуждаются существующие программные комплексы и информационные ресурсы.

1 Введение

Развитие графических пользовательских интерфейсов, основанных на взаимодействии пользователя с географическими картами с развитым инструментарием по навигации, масштабированию и поиску информации с использованием ее географической привязки, позиционирует сервис работы с упомянутой информацией как неотъемлемую часть любой современной информационной системы. Примером могут служить не только популярнейшие справочные системы информирования мобильных пользователей с использованием спутниковой навигации, не только сервисы информационных монстров типа Google или Yandex, предоставляющие как возможность просмотра детализированной информации о земной поверхности, так и создание пользовательских мультимедиа архивов с координатной привязкой документов, но и всевозможные так называемые геоинформационные системы (ГИС), полностью основанные на работе с информацией, имеющей географическую привязку. Наверное, можно утверждать, что географическая привязка информации и способы работы с такой информацией – достаточно актуальная задача для разработчиков современных информационных систем. Можно

также утверждать, что эта задача актуальна и для так называемых электронных библиотек (ЭБ), поскольку в нашем понимании последние являются ничем иным, как хорошо регламентированными информационными системами, ориентированными на работу с электронными документами.

Содержание цифрового контента и его описаний определяет возможности использования геоинтерфейсов (основанные на масштабируемых изображениях земной поверхности или цифровых картах графические интерфейсы для работы с информацией, имеющей географическую привязку) при работе с электронными библиотеками. При этом возникают следующие вопросы:

1. Как должна выглядеть пресловутая географическая привязка для обеспечения минимальной функциональности геосервисов?
2. Что должна означать географическая привязка цифрового объекта, где в электронной библиотеке должна содержаться информация о географической привязке цифровых объектов?
3. Как географическая привязка цифрового контента электронных библиотек связана с текущими регламентирующими документами по созданию информационных ресурсов?
4. Какие программные продукты управления цифровым контентом электронных библиотек в России, претендующие на первые роли, сегодня могут обеспечить минимальный геосервис?

Ниже сделана попытка сформулировать ответы на поставленные вопросы.

2 Как должна выглядеть географическая привязка цифрового объекта

Географическая привязка цифрового объекта должна определять связь цифрового объекта с некоторой областью на земной поверхности. Сразу следует заметить, что эта область может представлять собой:

- точку, определяемую парой координат;
- замкнутый контур, определяемый последовательностью пар координат;
- некоторый нормализованный термин, ассоциируемый с этой областью.

Труды 12^й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010

Очевидно, что последнее представление географической привязки не является однозначным, поскольку:

- географические названия зависят от времени и языка;
- любая область (контур) может содержать в себе множество названий включаемых в себя областей.

Следует признать, что однозначная географическая привязка может быть реализована только в терминах географических координат.

Несомненно, реализация географического поиска возможна лишь в том случае, если в информационных массивах содержатся элементы данных, по которым этот поиск и будет производиться. В качестве таких элементов данных (метаданных) должны выступать элементы, содержащие информацию о географических координатах или о географических названиях. Такие элементы могут содержаться в метаданных цифрового объекта и заполняться независимо от структуры этого объекта. Метаданные создаются в процессе явной или неявной каталогизации и, как правило, соответствуют той или иной общепринятой схеме. Для разных типов объектов приняты различные схемы метаданных [4]. Ниже (табл. 1) приведены некоторые из них.

Таблица 1

Тип физического или цифрового объекта	Схема метаданных
Библиографическая информация	MARC21 [5-6], RUSMARC [7]
Архивы, коллекции документов	OAE, Digital Collection, DC [8]
Описания предметов культурного наследия	CIMI, CIDOC CRM
Информационные ресурсы WWW	DC, GILS
Цифровые карты, космические снимки, данные дистанционного зондирования	CSDGM [12], ISO-19139, ISO-19115, GEO[11]

И если в схемах метаданных для цифровых карт, космических снимков и т. п. (CSDGM, ISO-19139, ISO-19115, GEO и др.) элементы, описывающие географическую привязку цифрового объекта, с необходимостью присутствуют, то в других схемах наличие подобных элементов не очевидно, а заполнение их – не обязательно.

Ниже рассмотрены некоторые из вышеперечисленных схем данных, которые могут содержать элементы с географической привязкой. При этом не рассматриваются элементы, содержащие географическую привязку методом географического назва-

ния области ¹. Дело в том, что для корректной работы с такими элементами (например, использование их для поиска) необходима их дополнительная обработка – перевод из терминов географических названий в термины географических координат. Эта процедура не всегда корректна.

2.1 USmarc, marc21

USMARC, MARC21 – американский стандарт на библиографические описания различных объектов, ориентированный на структуру записи ISO-2709. В табл. 2 представлен фрагмент описаний полей с примерами представления географических координат в формате MARC21.

Наличие огромных информационных массивов, описанных в соответствии с правилами MARC21, с незаполненными полями **034** и **255** обедняет возможности географического поиска информационных ресурсов. Поэтому сегодня выполняется ряд проектов по восстановлению в записях значения поля **034** в записях MARC21 на основе информации, содержащей географические названия, например, из поля **043** (географический код). Подобный проект сегодня выполняется совместно Библиотекой конгресса США и OCLC (см. <http://www.loc.gov/cds/notices/2010-04-19.pdf> от 19 апреля 2010 года).

2.2 RUSmarc

RUSmarc – российский вариант схемы описания библиографических данных UNIMARC, ориентированный на структуру ISO-2709, учитывающий национальные правила каталогизации. Географические координаты могут присутствовать в полях **123** и **206**. В табл. 3 представлены описание и примеры заполнения соответствующих полей.

Следует заметить, что поле **123** содержит те же данные о масштабе и координатах, которые записываются в поле **206**, но в кодированной форме. Следует также обратить внимание, что представление координат в поле **123** в RUSMARC совместимо с представлением MARC21 (поле **034**). Однако обратной совместимости нет! Возможности представления координат в MARC21 существенно шире и не ограничены жестким форматом. Более того, точность представления координат в RUSMARC до одной угловой секунды сегодня является явно недостаточной.

2.3 GILS

Более совершенной в части географической информации является схема GILS – Government Internet Locator Service [8].

```
<spatialDomain>
<boundingCoordinates>
<westBoundingCoordinate>112.5
</westBoundingCoordinate>
```

¹ См., например, ISO-3166, ГОСТ 7.67 – коды названий стран; TGN (Getty тезаурус географических названий) – <http://www.getty.edu/research/tools/vocabulary/tgn>

Поле	Подполе	Описание	Примеры
034	\$d – западная долгота \$e – восточная долгота \$f – северная широта \$g – южная широта	Coded Cartographic Mathematical Data. Подполя \$d, \$e, \$f, и \$g всегда присутствуют вместе. Координаты могут быть записаны в форме <i>hddmmss</i> (полушарие-градусы-минуты-секунды), однако другие формы записи также допустимы. Выравненные вправо подэлементы заполняются лидирующими нулями. Односимвольный код полушария: W(-) = запад, E(+) = восток, N(+) = север, S(-) = юг. Возможны обозначения +/-. Лидирующий символ + может быть опущен.	<i>hddmmss</i> : 1#SaaSb22000000\$dW1800000\$eE1800000\$fN0840000\$gS0700000 <i>hddd.ddddd</i> : 1#SaaSdE079.533265\$eE086.216635\$fS012.583377\$gS020.419532 <i>+ -ddd.ddddd</i> : 1#SaaSd+079.533265\$e+086.216635\$f-012.583377\$g-020.419532 Без лидирующего + 1#SaaSd079.533265\$e086.216635\$f-012.583377\$g-020.419532 <i>hddmm.mmmm</i> : 1#SaaSdE07932.5332\$eE08607.4478\$fS01235.5421\$gS02028.9704 <i>hddmmss.sss</i> : 1#SaaSdE0793235.575\$eE0860727.350\$fS0123536.895\$gS0202858.125
255	\$c – сведения о координатах.	Cartographic Mathematical Data	##\$aScale 1:41,849,600\$c(W 180°--E 180°/N 90°--S 90°).

```
<eastBoundingCoordinate>129
</eastBoundingCoordinate>
<northBoundingCoordinate>=-13.5
</northBoundingCoordinate>
<southBoundingCoordinate>-35.5
</southBoundingCoordinate>
</boundingCoordinates>
<place>
<placeKeywordThesaurus>
</placeKeywordThesaurus>
<placeKeyword></placeKeyword>
</place>
</spatialDomain>
```

2.4 Dublin Core

DC (Dublin Core) – наиболее известная схема данных для каталогизации информационных ресурсов общего назначения. Информация о географической привязке объекта может содержаться в элементе COVERAGE. Допускается применение элементов типа «точка» и типа «прямоугольник». Ниже приведен пример XML-представление этого элемента схемы DC.

```
DCMI Point 2
<coverage type="spatial"
scheme="DCMI Point">
<point>
<east> 115.85717 </east>
<north> -31.95301 </north>
</point>
</coverage>
```

² DCMI Point – схема для указания географических координат, <http://dublincore.org/documents/dcmi-point>

DCMI Box ³

```
<coverage type="spatial" scheme="DCMI Box">
<box name=" Western Australia">
<northLimit> =-13.5 </northLimit>
<eastLimit> 129 </eastlimit>
<southlimit> -35.5 </southlimit>
<westlimit> 112.5 </westlimit>
</box>
</coverage>
```

2.5 CIMI

Схема CIMI – Computer Interchange of Museum Information. Информация о пространственных координатах описываемого объекта в схеме CIMI может присутствовать в следующих элементах:

```
objectInfo
digitalObject
actualDo
spatialReferencingSystem
xCoordinateInSpatialReferencingSystem
yCoordinateInSpatialReferencingSystem
```

Кроме этого схема CIMI включает на верхнем уровне все элементы Dublin Core, что позволяет использовать элемент COVERAGE для географической привязки контента.

2.6 GEO

GEO – профиль Z39.50 [1, 10] соответствует стандарту представления пространственных мета-

³ DCMI Box – схема для идентификации пространственного региона, используя географические координаты, <http://dublincore.org/documents/dcmi-box>

Таблица 3

Поле	Подполе	Описание	Примеры
123	\$d – западная долгота \$e – восточная долгота \$f – северная широта \$g – южная широта	<p>Координаты для планетарных и земных картографических материалов. Обязательное для картографических материалов. Повторяется, если материал содержит данные в различных масштабах и с разными координатами. Каждое подполе координат (\$d-\$g) имеет фиксированную длину в 8 символов и не повторяется. Каждое подполе содержит следующие данные:</p> <ul style="list-style-type: none"> – Позиция символа 0. Стороны света. Односимвольный код: w = запад, e = восток, n = север, s = юг – Позиции символов 1-3. Градусы. 3 цифровых символа, выравниваемые вправо заполнителями - нулями. – Позиции символов 4-5. Минуты. 2 цифровых символа, выравниваемые вправо заполнителями - нулями – Позиции символов 6-7. Секунды. 2 цифровых символа, выравниваемые вправо заполнителями - нулями 	<p>123 2#\$aa\$b150000\$b25000\$d e0150000\$ee0173045\$fn0 013012\$gs0023035</p> <p>Часть карты Заира с линейными масштабами 1:150000 и 1: 25000, долгота - от 15°В до 17°30'45" В, широта - от 1°30'12" С до 2°30'35" Ю.</p>
206	\$d – сведения о координатах. Факультативное, не повторяется	<p>Обязательное для картографических материалов.</p> <p>Индикатор 1: Определяет, структурированы ли данные в поле</p> <p># – Данные представлены в неструктурированном виде</p> <p>0 – Данные представлены в структурированном виде</p> <p>Индикатор 2: # (не определен)</p>	<p>Неструктурированное представление</p> <p>206 ##\$aScale 1: 6 336 000 (W 170° -W 50°/N 80° -N 40°)</p> <p>Структурированное представление</p> <p>206 0#\$bScale 1: 6 336 000\$d(W 170° -W 50°/N 80° -N 40°)</p>

данных FGDC Content Standards (CSDGM) [12]. В табл. 4 приведено соотношение полей, содержащих граничные географические координаты для профиля Z39.50 GEO, стандарта FGDC Content Standards и MARC21 [14, 15].

Форма представления координат для профиля GEO соответствует форме MARC21. Также следует заметить, что стандарты ISO-19115, ISO-19139 содержат аналогичное описание географической привязки контента информационного объекта.

Таким образом, географический аспект может присутствовать в метаданных, он выражается в виде географических названий или географических координат. При этом чаще всего географические координаты указывают или на точку, или на ограничи-

вающий область четырехугольник. Представление в метаданных более сложных областей, например, области, заданной произвольной замкнутой кривой на поверхности, возможно лишь в специализированных схемах данных (FGDC, ISO-19115 и др.), которые изначально создавались для географической информации. Тем не менее, информации в виде точки и граничного четырехугольника вполне достаточно для организации поиска с использованием пользовательских графических интерфейсов, основанных на географической карте. Поэтому следует сделать вывод, что практически все используемые схемы метаданных (библиотеки, архивы, музеи и др.) допускают интеграцию с географическими поисковыми системами.

Элемент схемы	Содержание	FGDC	MARC21
idinfo/spdom/bounding/westbc	West Bounding Coordinate	1.5.1.1	034 d
idinfo/spdom/bounding/eastbc	East Bounding Coordinate	1.5.1.2	034 e
idinfo/spdom/bounding/northbc	North Bounding Coordinate	1.5.1.3	034 f
idinfo/spdom/bounding/southbc	South Bounding Coordinate	1.5.1.4	034 g

3 Что должна означать географическая привязка цифрового объекта

Географическая привязка цифрового объекта должна означать, что все элементы его описания, имеющие прямое или косвенное отношение к географии, должны иметь возможность содержать географическую привязку. Это может относиться к описанию как информационного контента объекта, так и к описанию контекста существования объекта.

В качестве примера последнего можно привести следующее. Несомненно, явную географическую направленность имеет элемент «Место публикации» (MARC21 260\$a) в традиционном библиографическом описании. Этот элемент не описывает информационный контент, но он описывает контекст (что, где, когда и пр.). К сожалению, в существующих библиографических схемах данных нет возможности указать явную географическую привязку в виде географических координат как для этого элемента, так и для других элементов контекста.

Для описаний различных типов объектов элементами, содержащими описание контекста, могут быть элементы, указывающие на категорию местоположения в географическом смысле, например:

- место создания;
- место публикации;
- место проведения;
- место находки;
- место хранения;
- место обитания;
- место выставки;
- место реставрации;
- место съемки;
- и др.

Заметим, что ни для одной из перечисленных категорий контекста ни в одной из упоминавшейся выше схемах данных не предусмотрено использование географических координат. В лучшем случае допускается использование нормированных географических названий, а в идеальном случае эти нормированные названия содержат указания на соответствующий географический тезаурус.

Следствием этой неполноты схем метаданных является тот факт, что сегодня, оставаясь в рамках действующих стандартов на метаданные и используя существующие информационные массивы, не-

возможно сформулировать разумный поисковый запрос (и, тем более, получить разумный ответ) о представлении информации, например, о всех артефактах, найденных на территории, ограниченной координатами (x_1, y_1, x_2, y_2) , или о всех литературных шедеврах, созданных на той же самой территории.

Приведенный пример показывает, что существующие схемы метаданных требуют существенной переработки для представления информации о контексте в части его географического аспекта. Отсутствие возможности строгой фиксации этого аспекта существенно обедняет поисковые возможности географических информационных систем.

Справедливости ради следует отметить, что географическая координатная привязка контекста (класс E47) присутствует в определениях онтологии в сфере документов по культурному наследию (CIDOC CRM [16]). Это не удивительно, т. к. в описаниях объектов культурного наследия описание контекста имеет не меньшее значение, чем описание контента. К сожалению, в CIDOC CRM геометрия всех объектов сводится к точке.

4 О регламентирующих документах

Регламентирующими документами для создания электронных ресурсов и электронных библиотек с большой степенью натяжки можно сегодня признать лишь правила каталогизации для традиционных библиотек в части электронных ресурсов. Регламентация кодирования географической привязки выглядит даже в детализированных описаниях RUSMARC очень и очень бледно. Явный акцент при этом делается на фиксацию привязки только информационного контента посредством нормализованных географических тезаурусов. Числовым выражениям отводится второстепенная роль, а представление координат не соответствует мировой практике. В настоящее время привязка контекста регламентируется только в документах по культурному наследию, но опять-таки только на уровне географических названий.

5 О программных и информационных продуктах

Анализ распространенных программных продуктов для управления цифровым контентом электронных библиотек в России (и не только) показал,

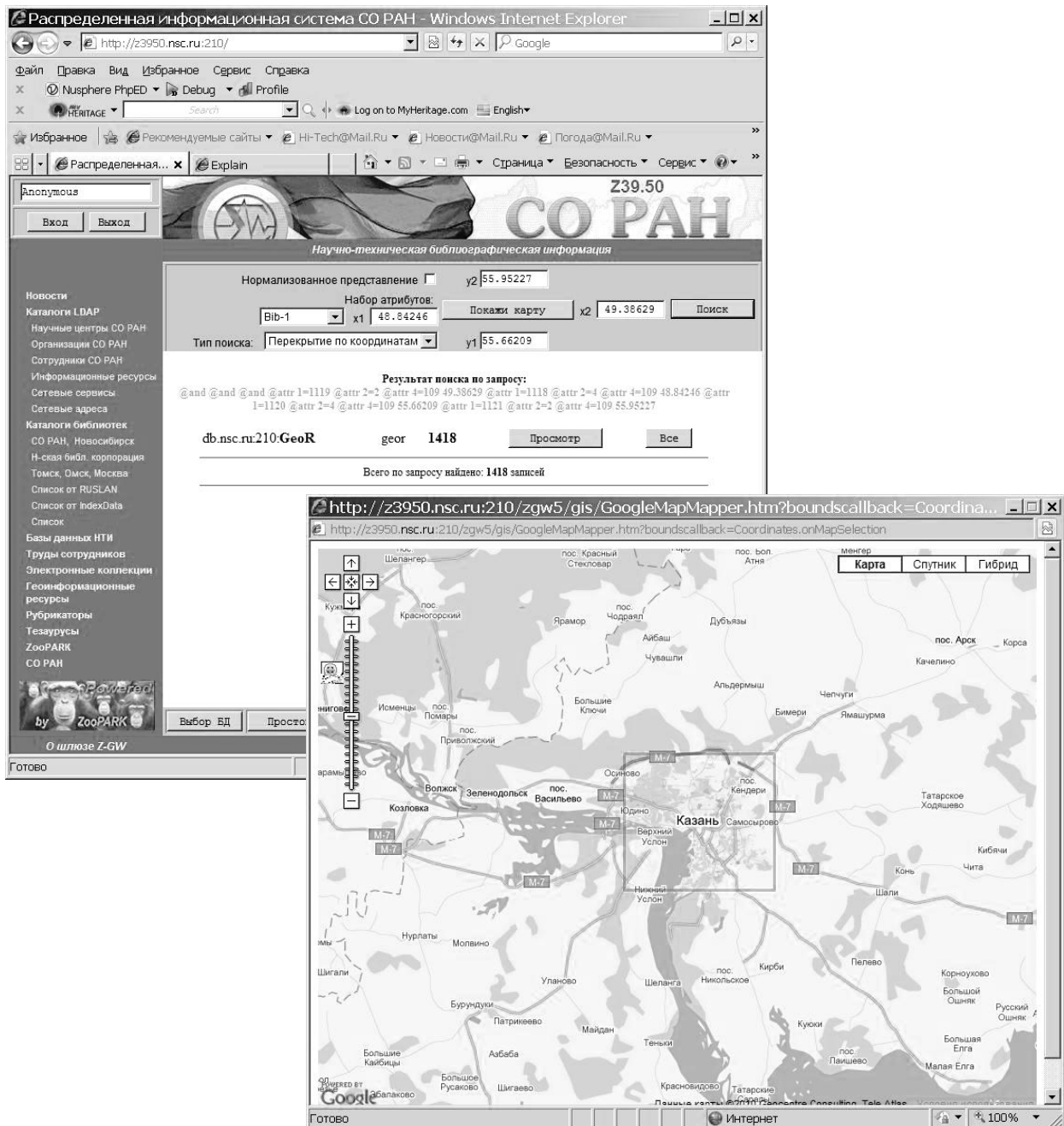


Рис. 1. Интерфейсы поиска информации шлюза Z-GW комплекса ZooPARK с использованием карт Google

что практически ни один из этих продуктов не предоставляет сервисов для управления географической привязкой.

Из известных авторам программных комплексов единственным свободно распространяемым программным продуктом, который позволяет работать с географическим аспектом информационного контента, является программный комплекс GeoNetwork [17]. Несмотря на то, что основное его предназначение – обеспечение управления специализированным цифровым контентом на основе метаданных FGDC [12, 13], ISO-19115 и ISO-19139, он допускает час-

точное использование дополнительных XML-схем данных и управление различными типами цифровых объектов (снимки, текстовые документы и т. д.).

Из известных авторам библиографических ресурсов, содержащих географическую привязку информационного контента, является база данных Американского геологического института GeoRef [18], содержащая на сегодняшний момент около 3 млн. записей.

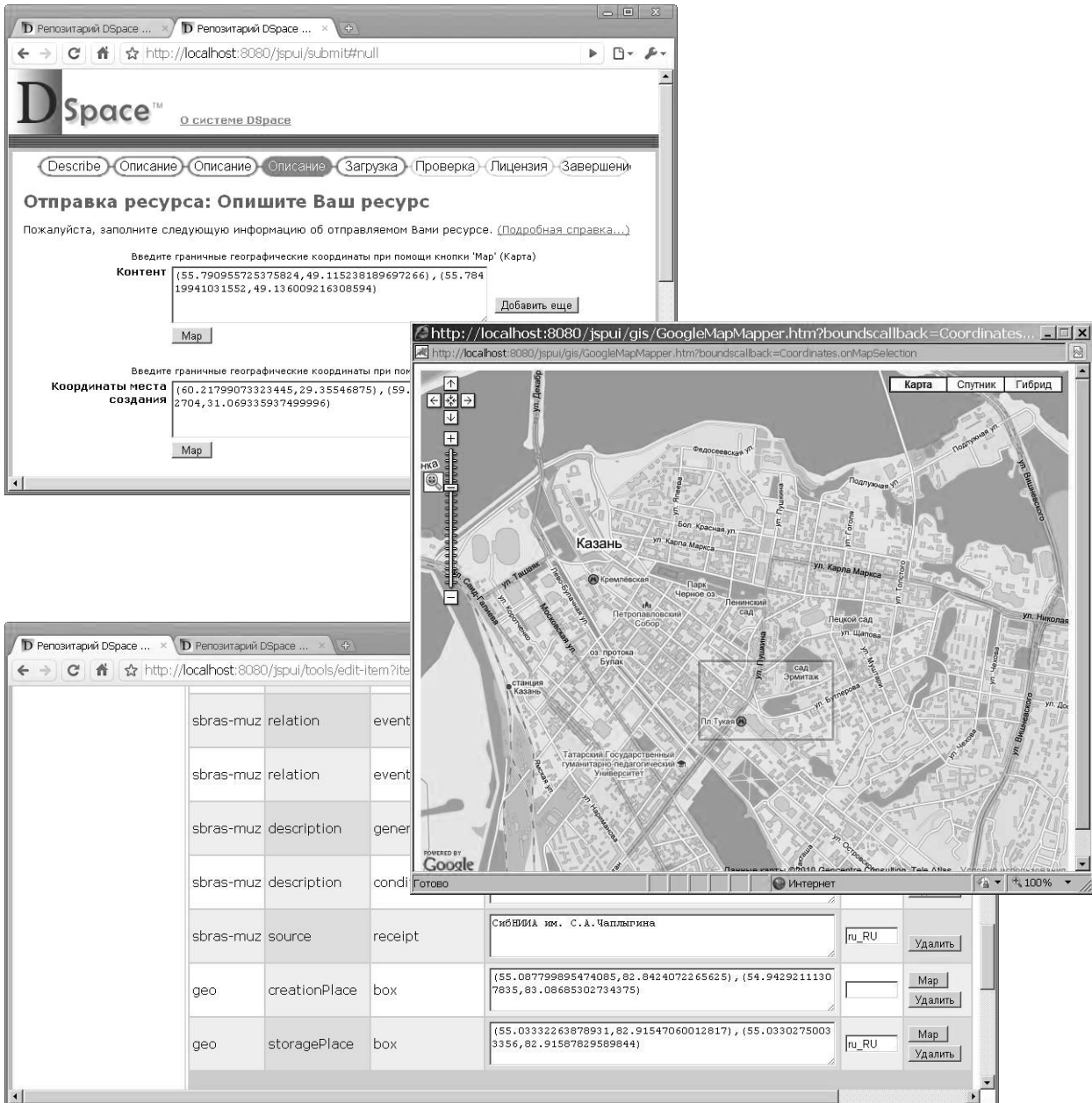


Рис. 2. Интерфейсы ввода и редактирования метаданных в модернизированной DSpace с использованием карт Google

6 О собственных разработках

Следует заметить, что проблемы географической привязки цифровых объектов интересуют авторов давно. В частности, еще в 2000 году на RCDL был озвучен доклад [19] о специфической информационной системе, включающей поиск информации по географическим координатам.

В настоящий момент работы ведутся по следующим направлениям:

- разработка расширений существующих схем данных для работы с географической информацией;
- разработка административных интерфейсов управления цифровыми объектами, содержащими координатную привязку;

- разработка пользовательских интерфейсов доступа к информации, в том числе для поиска информации;

- адаптация и модернизация существующего свободно распространяемого программного обеспечения для возможности управления цифровыми объектами к географическим потребностям;

- модернизация интерфейсов программного комплекса ZooPARK собственной разработки.

В качестве примера интерфейсов доступа к географической информации на рис. 1 приводятся интерфейсы шлюза ZooPARK (доступ к данным GeoRef).

Другой пример демонстрирует интерфейсы модернизированной системы DSpace [20]. Информационная система DSpace обладает широкими воз-

возможностями по управлению цифровым контентом, но не содержит интерфейсов для работы с географическими координатами. Учитывая, что DSpace широко используется для создания электронных библиотек, мы не могли пройти мимо соблазна модифицировать эту систему для придания ей дополнительной функциональности. На рис. 2 показаны пользовательские интерфейсы для ввода и редактирования географической информации в модернизированной системе DSpace. При этом достигнутая функциональность системы позволяет реализовать географическую привязку как для контента, так и для контекста. Поиск информации по различным критериям осуществляется через интерфейсы ZooPARK, который напрямую связан с данными DSpace.

7 Заключение

В заключение следует отметить, что игнорирование географической привязки цифровых объектов в электронных библиотеках и информационных системах является фактором, значительно их обедняющим. Этот фактор может быть исключен при правильном выборе схем данных, модернизации существующего программного обеспечения и, безусловно, разработке нового программного обеспечения, учитывающего географические потребности пользователей.

Литература

- [1] Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. – Новосибирск: ИВТ СО РАН, 2004. – 361 с.
- [2] Жижимов О.Л., Мазов Н.А. Об использовании географических координат при поиске библиографической информации // Науч. и техн. библиотеки. – 2009. – № 1. – С. 54-60.
- [3] Жижимов О.Л., Мазов Н.А. Об интеграции библиотечно-информационных и геоинформационных технологий // Вычислительные технологии. Совместный выпуск. Вестник КазНУ им. Аль-Фараби. Сер. математика, механика, информатика. – 2008. – Т. 13, Ч. II. – С. 97-101.
- [4] Мазов Н.А. Жижимов О.Л. Метаданные и их роль в распределенных информационных системах на основе использования протокола Z39.50 // Библиосфера. – 2006. – № 2. – С. 51-60.
- [5] MARC Standards. – <http://www.loc.gov/marc/>.
- [6] Форматы MARC21. – <http://marc21.rsl.ru>.
- [7] RUSMARC. Российский коммуникативный формат представления библиографических записей в машиночитаемой форме. – <http://www.bookresearch.ru/rusmarc.htm>.
- [8] DCMI – Dublin Core Metadata Initiative. – <http://www.dublincore.org/>.
- [9] Application Profile for the Government Information Locator Service (GILS), Version 2, Nov. 24, 1997. – http://www.gils.net/prof_v2.html.
- [10] ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification / Z39.50 Maintenance Agency Official Text for Z39.50-1995. – July 1995.
- [11] Douglas D. Nebert. Z39.50 Application Profile for Geospatial Metadata or "GEO"/ Version 2.2 / U.S. Federal Geographic Data Committee. – <http://www.blueangeltech.com/Standards/GeoProfile/geo22.htm>.
- [12] Content Standard for Digital Geospatial Metadata. – <http://www.fgdc.gov/metadata/constan.html>.
- [13] FGDC – Federal Geographic Data Committee. – <http://www.fgdc.gov>.
- [14] Crosswalk: USMARC to FGDC Content Standards for Digital Geospatial Metadata. – <http://www.alexandria.ucsb.edu/public-documents/metadata/marc2fgdc.html>.
- [15] Crosswalk: FGDC Content Standards for Digital Geospatial Metadata to USMARC. – <http://www.alexandria.ucsb.edu/public-documents/metadata/fgdc2marc.html>.
- [16] Definition of the CIDOC Conceptual Reference Model // Produced by the COM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group, Version 4.2.1, October 2006. – http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.1.doc.
- [17] GeoNetwork Opensource Community website. – <http://geonetwork-opensource.org/>.
- [18] The GeoRef database // American Geological Institute. – <http://www.agiweb.org/georef>.
- [19] Жижимов О.Л., Коджесян В.С., Мазов Н.А. Пример распределенной информационной системы на основе метаданных и международных стандартов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: 2-я Всерос. науч. конф., Протвино, 26 – 28 сентября 2000 г.: Сб. докл. – Протвино: ГНЦ ИФВЗ, 2000. – С. 102-106. – <http://rcdl.ru/doc/2000/014.pdf>.
- [20] Система DSpace, домашняя страница. – <http://www.dspace.org>.

Problems of a geographical binding of digital objects in electronic libraries

O.L. Zhizhimov, N.A. Mazov

The questions connected with geographical aspect of digital objects of electronic libraries are considered. Possibilities of their geographical binding both for a content, and for a context are discussed. Widespread schemes of data regarding their support of geographical aspect of the information are considered. Thus the basic accent becomes for a co-ordinate way of realisation of a geographical binding. Existing software complexes and information resources are discussed.

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 10-07-00302 и 09-07-00277) и Сибирского отделения РАН (интеграционный проект 121)