

Электронный атлас «Биоразнообразии животного и растительного мира Сибири» *

Коропачинский И.Ю., Шокин Ю.И., Шумный В.К., Ермаков Н.Б., Колчанов Н.А., Федотов А.М.
E-mail: fedotov@ict.nsk.su

1 Введение

Не секрет, что информация является стратегическим ресурсом, таким же как и традиционные материальные и энергетические ресурсы. Информационные ресурсы, переведенные в электронную форму, приобретают новое качество, которое обеспечивает им более широкое распространение и эффективное использование. Современные информационные технологии позволили приступить к широкомасштабному переводу накопленной человечеством информации в электронную форму и созданию принципиально новых видов информационных ресурсов, каким являются электронные коллекции и библиотеки. Создание и организация доступа к электронным коллекциям стала одной из важнейших задач информационного обслуживания науки и образования.

Биологи в процессе своей научной деятельности собирают гигантский фактический материал, характеризующий различные стороны биоразнообразия животного и растительного мира. В настоящее время этот оригинальный материал находится в виде бумажных архивов и не представляет организованную информационную среду, которая является основой для современных научных исследований. Между научными подразделениями как в России, так и за рубежом отсутствуют надежные и быстродействующие каналы обмена информацией, что является существенным препятствием на пути развития интеграционных проектов и других форм научного взаимодействия. Многие важные результаты биологических исследований публикуются в региональных и местных научных изданиях, которые малодоступны для широкого круга исследователей.

Электронная публикация биологической коллекции данных представляет собой новую форму хранения и обмена информацией. Для нее характерны прежде всего динамичность (возможность обновления) и глобальный доступ (через компьютерные сети). На сегодняшний день электронные публикации не преобладают в общем объеме опубликованных информационных ресурсов, но их

* Работа выполнена в рамках проекта «Электронная библиотека Сибирского отделения РАН», а также в рамках проекта РФФИ 99-07-90222.

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

доля стремительно растет. При этом возрастает и качество таких публикаций.

Изначально электронная публикация появилась как электронный вариант бумажной публикации. В последние годы она приобрела самостоятельное значение. Более того, часть научных проектов, например, «Flora of North America», в процессе их выполнения была переориентирована с бумажной формы опубликования на электронную. В электронной форме планируется опубликовать и новую «Флору России». Во всяком случае, электронная форма опубликования признается уже сейчас как приоритетная по сравнению с типографской. Отсюда становится очевидным приобретение электронной формой опубликования самостоятельного и первоочередного значения. Интеллектуальные и технические ресурсы в ближайшие годы будут направлены на развитие новых возможностей и повышение качества электронной публикации данных.

В последние годы, преимущественно за рубежом, были предприняты попытки издания биологических баз данных на CD-ROM дисках по разным тематикам (например, «North American Birds»). Как правило, они имеют общеобразовательный либо прикладной характер, ориентированный на коммерческое использование (последние, обычно, предусматривают своевременное пополнение на дополнительных жестких носителях). Имеются и отечественные разработки в этой области («Мир жуков», «Деревья Германии», «Бобовые России»).

С другой стороны, все большее количество ученых-биологов используют электронные базы данных для оптимизации работы, систематизации предметной области, быстрого и качественного оформления результатов своей деятельности. Необходимо создание и внедрение системы, способной быстро публиковать научные данные в Интернет. В России разработку такой системы выполнила фирма FLORIN/DataX, возглавляемая ботаниками, сотрудниками ГБС РАН и МГУ (Серебряный М.М., Савов К.П., Майоров С. и др.). Распространению FLORIN (FLORistic INvestigations) в России препятствует высокая цена на программные продукты, входящие в ее состав. С помощью этой системы созданы электронные публикации по типовым образцам Гербария МГУ, мхи бывшего СССР, палинология рода *Sentaulgea* и др.

Важной проблемой выступает организация разнородной информации в удобном для конечного пользователя виде, что требует новых исследований и разработок интерфейсов для корректного отражения предметной области. Другой задачей обслуживания электронных коллекций и библиотек является стандартизация данных, раз-

работка технологических решений и юридических аспектов использования информации, включая вопросы интеллектуальной собственности. Отличительной чертой электронной библиотеки является возможность параллельного использования различных поисковых механизмов и средств доступа к гетерогенным банкам электронных данных. Так как в качестве ответа на запрос к электронной библиотеке пользователю может быть представлен далеко не единственный электронный документ и совсем не обязательно в текстовой форме, необходимы такие информационные системы, которые бы обеспечивали эффективный комплексный поиск и анализ информации в коллекциях гетерогенных объектов.

2 Проблемы биологических электронных публикаций

Публикация биологических данных в электронном виде принципиально не отличается от электронной публикации других данных, но имеет некоторые особенности: большой объем текстовой и графической информации, географическая привязка данных к местности, слабая формализация описательных данных, зависимость системы сбора информации от конкретного исследователя и ряд других.

Несмотря на кажущуюся очевидность преимуществ электронной публикации над обычной широкое применение электронных публикаций биологических данных имеет некоторые проблемы, которые зависят от ряда причин. Прежде всего, традиционно сложившаяся схема исследований, ориентированная на бумажные носители информации, а также профессиональные трудности в освоении вычислительной техники и недостаток технических средств. Существует пока непреодоленная боязнь потери данных на магнитных носителях, с одной стороны, и боязнь потери контроля над данным, с другой стороны. Негативную роль играет недостаточное знание современных информационных технологий и методов анализа информации.

Современные информационные технологии в биологии и экологии представляют широкие возможности для управления большими объемами данных и для их сложной математической обработки. Однако для интенсивного использования этих возможностей необходимо пройти рутинный этап адаптации и определенной организации накопленных (как правило, эмпирических) эколого-биологических данных к современным требованиям информатики. Под адаптацией данных подразумевается прежде всего их формализация на основе общепринятых стандартов, форматов и устоявшихся понятий. Под организацией данных понимается принятие определенной структуры, т. е. установление определенных иерархий и связей (вертикальных и горизонтальных) между информационными блоками. Все это в полной мере относится к таким эколого-биологическим наукам как геоботаника (фитоценология).

Биологическая информация имеет следующие отличительные черты:

- Данные — это большой объем текстовой и графической информации.
- Данные имеют географическую привязку.
- Со временем данные подлежат обновлению, причем каждое обновление должно документироваться.

- Данные имеют строго определенную структуру, которая плохо приспособлена для образовательных и исследовательских целей, так как исторически обусловлена и изменяется со временем.
- Данные имеют авторство и охраняются авторским правом.

Поэтому для более широкого применения электронной формы публикации биологических данных необходимо:

1. Осознание биологами новых возможностей представления данных, возникающих при их опубликовании в электронном виде: цветные иллюстрации, многоходовый доступ, развитая поисковая система, оформление перекрестных ссылок, географическая привязка данных с использованием геоинформационных систем, обеспечение связи с другими информационными ресурсами по данной тематике и др.
2. Осознание биологами новых возможностей доступа к данным и обмена данными.
3. Осознание авторами крупных проектов, таких как коллективные «Флоры» и «Фауны», принципиально новых возможностей создания коллективных монографий в режиме удаленного доступа.
4. Представление о том, что по завершении работы готовый продукт автоматически становится валидной публикацией, которую можно включать в список опубликованных работ и делать на нее ссылки.
5. Разработка новых, более эффективных программных средств организации и хранения разнородной биологической информации, удобного и быстрого доступа к ней (дружественный интерфейс), автоматизация поиска и анализа информации, создание экспертных систем, помогающих исправлять ошибки и генерировать новую информацию по имеющимся данным.
6. Необходима организация администрирования баз данных с оформлением прав доступа к информации по схеме пользователь (только чтение), автор (изменение и дополнение информации), администратор базы (изменение структуры данных), защита данных от несанкционированного доступа, документирование работы каждого автора.
7. Привлечение к совместной работе широкого круга специалистов из других городов и стран.

3 Проект “Электронный атлас Биоразнообразия растительного и животного мира Сибири”

В начале 1998 года в Сибирском отделении РАН была сформирована программа развития информационных ресурсов отделения под общим названием “Электронная библиотека Сибирского отделения РАН” [7]. В рамках данной программы разработан проект создания Электронного атласа “Биоразнообразие растительного и животного мира Сибири” [3].

Электронный атлас призван обеспечить систематизацию и организацию широкого доступа к разнообразной информации по биоразнообразию растительного мира Сибири на основе современных информационных технологий. Предполагается разработать общие подходы по

систематизации и сохранению разнородной ботанической и экологической информации, интегрировать эту информацию в Геоинформационную систему (ГИС), сделать доступными через сеть *Internet* массивы данных материалов и коллекций [5].

Основные источники информации — данные собранные исследователями в течении нескольких десятилетий в биологических институтах отделения, научные журналы, монографии, учебники, материалы, хранящиеся в зоологических музеях и гербариях, а также результаты экспериментальных исследований и полевые журналы экспедиций. В рамках библиотеки создаются базы данных со средствами поиска по важнейшим группам растений и животных, растительных сообществ и биоценозов. В частности, баз данных «Редкие виды растений Сибири», «Охраняемые природные территории», «Определители таксономической принадлежности», «Лекарственные растения», «Природно очаговые болезни», «Тенофонд сельско-хозяйственных растений и животных» и т.д.

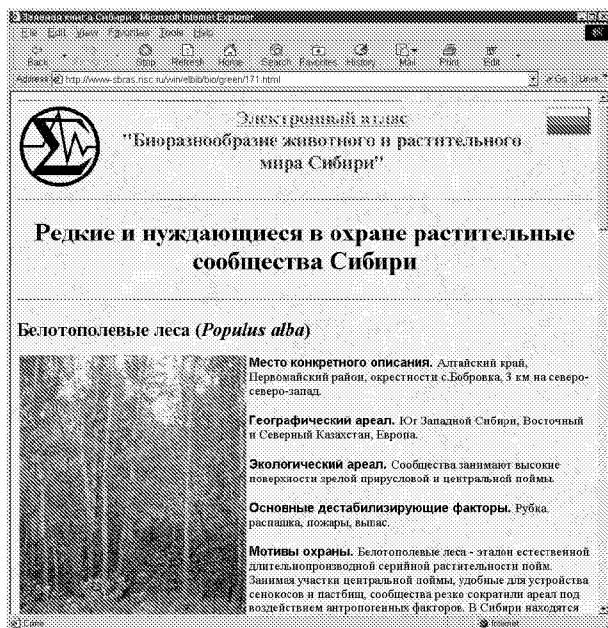


Рис. 1: База данных «Зеленая книга Сибири»

В качестве примера, уже созданных баз данных, можно указать Информационные системы «Редкие и нуждающиеся в охране растительные сообщества Сибири» [1] (База данных «Зеленая книга Сибири» рис. 1), «Флора Новосибирской области» (см. рис. 2) и «Интернет-каталог растений Сибири» (см. рис. 3).

В настоящий момент на этих базах данных отработываются основные технологические подходы для создания электронного атласа и создание автоматически актуализируемых баз данных биологических видов и на основе системы классификаторов, а также поддержка распределенной работы и кооперации исследователей, находящихся в разных регионах.

Составляющими электронного атласа являются тематические связанные базы данных (БД), содержащие информацию уникальных коллекций и фондовых материалов, литературные данные по биоразнообразию растительного мира Сибири. Основные разделы электронной библиотеки содержат эталонные названия таксонов, информацию о видовом разнообразии и разнообразии

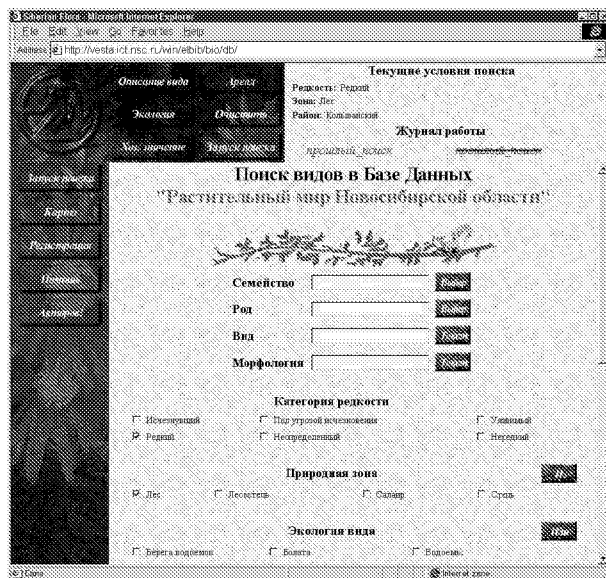


Рис. 2: База данных «Флора Новосибирской области»

сообществ, данные о видах, имеющих ресурсную ценность (ресурсное разнообразие), информацию по природоохранной тематике и интродукции видов.

Основой атласа являются уникальные коллекции и массивы данных по биоразнообразию растительного мира Сибири. Так, в гербариях ЦСБС хранится около 310 тыс. образцов высших сосудистых растений, собранных на территории Сибири, в числе которых и уникальные типовые образцы. Имеются коллекции мхов, грибов, лишайников, низших растений. В фитоценологе хранятся материалы более чем 10 тыс. описаний разнообразных растительных сообществ. Собрание древесных растений ЦСБС насчитывает около 800 видов, гибридов и форм, что в 2.5 раза превышает количество видов дендрофлоры Сибири. Уникальна информация по интродукционным испытаниям более 2000 видов, гибридов и форм, из которых 136 рекомендовано к использованию в Южной Сибири. Ценные данные содержат полевые дневники исследователей.

4 Используемые технологии

В основу создания Электронной библиотеки СО РАН положен принцип *Internet/Intranet* технологии [6]. Использование этих технологий при создании информационных ресурсов и построении информационных систем различного назначения в ближайшее время станет доминирующим в мировом информационном пространстве. Что самое главное *Internet* технология позволяет оперативно управлять и актуализировать информацию, хранящуюся в базах данных через просмотрщик WWW страниц и создавать динамические системы формирования коллекций.

В настоящее время представление информации в виде статических HTML-страниц для полномасштабных информационных систем, каковыми являются электронные библиотеки и коллекции, теряет актуальность, поскольку любое небольшое ее изменение или обновление зачастую влечет за собой весьма энергоемкую процедуру управления множества файлов, а создание системы поис-

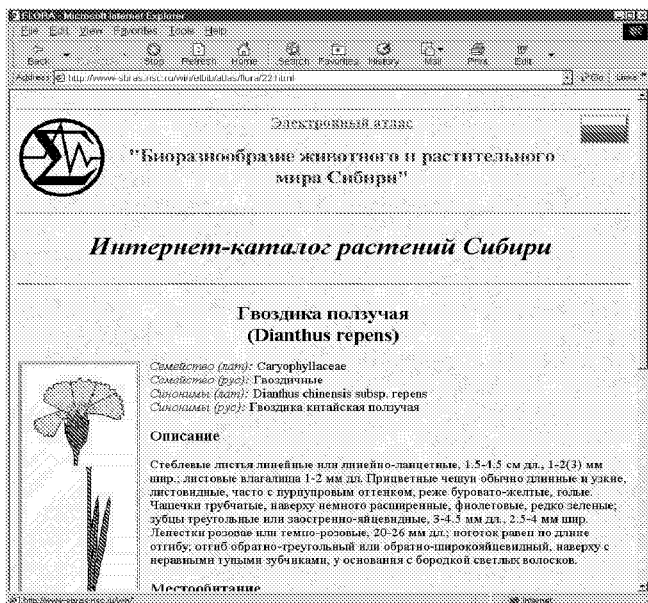


Рис. 3: База данных «Интернет-каталог растений Сибири»

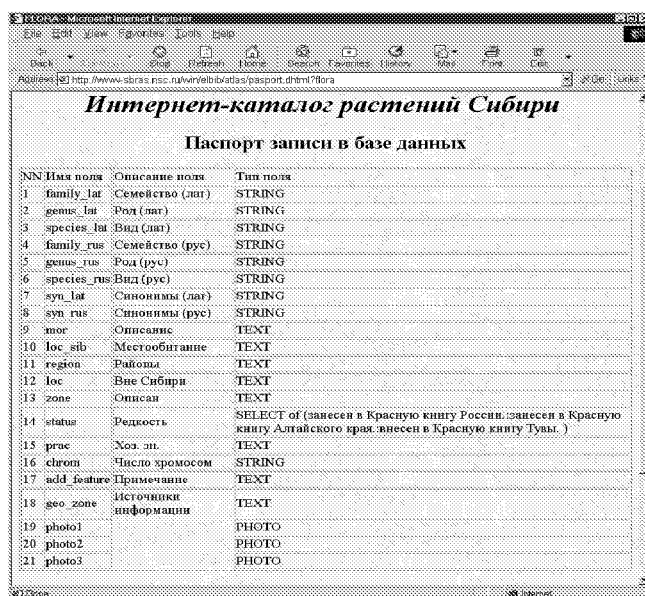


Рис. 4: Паспорт документа в «Интернет-каталоге растений Сибири»

ка упирается в организации индексных файлов, размер которых может быть сравним с размером основной информации.

В связи этим, в основу создания электронных коллекций в наших работах положена концепция динамических документов, в которой каждый тип документов, содержащих информацию о конкретных фактах, представляется в виде набора объектов со своими характеристиками и атрибутами (наподобие тому, как это принято в объектных языках программирования).

Для занесения в электронную коллекцию фактографической информации, последняя подлежит паспортизации — формальному описанию структуры объектов, составляющих документы (см. рис. 4).

Разработанная технология создания динамической системы публикаций позволяет оперативно управлять и актуализировать информацию, хранящуюся в разнородных и распределенных по сети базах данных, организовать гибкий поиск, что самое главное использовать достаточно удобный интерфейс для ее наполнения.

5 Подходы к формированию информационных систем

Основные информационные задачи изучения и сохранения биологического разнообразия планеты связаны с составлением соответствующих каталогов растительного и животного мира и с описанием пространственного распределения организмов в биогеоценозах. Рассмотрим на примере создания информационной системы о растительных сообществах Сибири организационные принципы формирования информации.

Важнейшим элементом организации данных о растительном покрове является оценка уровней геоботанической информации. Выявляется три основных информационных уровня:

1. Уровень информации о конкретном растительном сообществе. Основной элемент — полное описание однородной растительности на определенном

участке земной поверхности (документ растительности).

2. Уровень информации о типологических категориях растительности, полученных в результате типизации (классификация) элементов первого уровня (описаний растительности). Основным элементом — стандартная характеристика синтаксона.
3. Уровень информации о составе и структуре растительности определенных территорий. Основным элементом — охарактеризованный в категориях предыдущего уровня контур растительности на карте определенного масштаба. Наиболее обычная форма представления информации этого уровня — карты растительности территорий.

На каждом из информационном уровнях решается ряд как научных задач, так и прикладных. Среди последних особо выделяются: оценка глобальных изменений биоты (мониторинг), оценка ресурсного потенциала растительности и сохранение биоразнообразия.

Структура электронного атласа разработана таким образом, чтобы максимально охватить все три информационных уровня и отразить особенность потоков информации между ними.

Наука о растительности тесно связана с географией, потому что ее основной объект — растительное сообщество — реально существующая часть поверхности Земли. Поэтому разработка электронного атласа на всех информационных уровнях поддерживается современными ГИС технологиями.

Рассмотрим структуру информационных уровней:

1-й уровень информации — база данных геоботанических описаний.

Основным элементом этого уровня выступает геоботаническое описание конкретного участка растительности. Функции введения, хранения и экспорта описаний выполняет стандартный Европейский пакет фитосоциологических данных TURBOVEG. Этот пакет как базу

данных описаний так и систему для анализа фтосоциологической информации. Развитие этого пакета поддержано рядом проектов, она широко распространена в Европе и России и была принята Рабочей группой по изучению растительности Европы.

TURBOVEG содержит большинство данных в таблицах dBase формата (DBF), которые могут быть напрямую использованы базами данных более высоких информационных уровней.

2-й уровень информации — база данных синтаксонов.

В настоящее время наиболее разработанная базой данных синтаксонов растительности находится в Ланкастерском университете (Великобритания) и реализована на стандартном пакете ACCESS-2. Эта база данных содержит разностороннюю информацию о синтаксонах растительности Европы (включая Россию) и разработана в рамках деятельности Рабочей группы по изучению растительности Европы.

Элементом данного уровня информации выступает типологическая единица растительности (тип экосистемы). Отсутствие в настоящее время единых подходов к классификации растительности создает определенные трудности при идентификации синтаксономических единиц и составления стандартного списка синтаксонов общего пользования. Поэтому для каждого вводимого синтаксона ставится идентификатор (код) в той системе классификации, в которой он был описан. В настоящей базе данных реализована идентификация единиц растительности в двух системах: эколого-фитоценотической (ассоциация, группа ассоциаций, класс ассоциаций, формация, группа формаций, класс формаций, тип растительности), система классификации Браун-Бланке (вариант, субассоциация, ассоциация, подсоюз, союз, подпорядок, порядок, класс). Кроме этого в базе данных предусмотрено введение единиц растительности без четкого ранга, которые обычно используются в картографии растительности. Для каждой единицы предусмотрены поля для введения синонимов. Для единиц системы Браун-Бланке предусмотрена связь с номенклатурным блоком.

Связь с 1-м уровнем геоботанической информации (с базой данных описаний) поддерживается через коды синтаксонов, которые проставляются в специальное поле каждого описания. Однако часть синтаксонов (особенно введенная из литературных источников или легенд карт) имеет только общие характеристики видового состава, типа местообитания и не представлена конкретными описаниями в базе данных первого уровня. Эта информация является менее ценной, однако она может быть полезной для представления слабо исследованных типов растительности.

Связь синтаксонов с ГИС осуществляется несколькими путями:

1. точечное представление локалитета описанного синтаксона на карте (аналогично предыдущему уровню информации),
2. контурное представление синтаксона.

Применяется в том случае, когда установлен ареал данного типа сообществ. Более грубая форма контурного представления может соответствовать контуру административного (или географического) района, где этот синтаксон был зафиксирован (описан). Синтаксон может быть также представлен множеством точек, т. е. локалитетами его описаний, взятых из базы данных предыдущего уровня.

Создание единого списка синтаксонов растительности Сибири в настоящее время наиболее возможно в системе Браун-Бланке, основанной на вполне определенных номенклатурных правилах типификации единиц и построения названий. Единственное ограничение — это недостаточный охват данной системой всех типов растительности и территорий Сибири. Основой современного списка синтаксонов служат уже опубликованные материалы. Для гемибореальных лесов существует список синтаксонов, который разрабатывался для всего ареала лесов этого географического типа.

Структура информации синтаксономического блока во многом аналогична предыдущему блоку и представлена двумя базисными таблицами.

Данный блок синтаксономической информации позволяет:

- оперативно составлять списки типов сообществ необходимые для инвентаризации биоразнообразия определенных территорий,
- получать карты распространения (точечные и контурные) типов сообществ различного ранга на любые территории,
- получать списки видов (с весовыми характеристиками каждого вида) определенных экосистем,
- оценивать ресурсную значимость определенных типов экосистем,
- получать списки и ареалы редких и находящихся под угрозой исчезновения сообществ на любой территории,
- получать разностороннюю информацию об экологических условиях, фитосреде и распространении редких, исчезающих, декоративных, ценных ресурсных видов (по поиску их в составе определенных типов экосистем).

Учитывая, что классификация методом Браун-Бланке интенсивно развивается на территории Сибири, предусмотрена возможность пополнения списка синтаксонов через специальный синтаксономический модуль с использованием технологии удаленного доступа Internet.

3-й уровень информации — база данных электронных карт растительности и информации дистанционного зондирования растительности территорий.

Это высший уровень обобщенных знаний о растительном покрове, так как любая карта — это модель растительности основанная на информации предыдущих двух уровней. Информация этого уровня является основой для принятия решений по рациональному использованию ресурсного потенциала и сохранению биоразнообразия регионов. Она также является основой решения таких глобальных научных проблем как Глобальные изменения биосферы и оценка биоразнообразия больших территорий.

В настоящее время в фондах многих научных, производственных и учебных заведений имеется значительное количество разнообразных карт растительности, выполненных в различных масштабах на различные территории. Однако для того, чтобы они стали элементом базы данных электронных карт необходимо осуществить несколько этапов их дальнейших преобразований:

- представление карты в электронном виде в формате базы данных (векторизация и т. д.);

- приведение к определенной системе координат;
- приведение к определенной проекции;
- установление связей легенды с элементами баз данных двух предыдущих уровней.

Аналогичные условия необходимы также и для введения в базу данных элементов дистанционного зондирования аэро-космической информации.

К сожалению, в настоящее время для многих существующих карт эти условия являются невыполнимыми или трудно выполнимыми (т.е. с большими погрешностями или допущениями). Многие опубликованные в открытой печати карты растительности (масштабов 1:1.500.000 и др.) выполнены без указания проекции, системы координат, координатной сети, а иногда и с сознательно внесенными искажениями в картографическую основу (по требованию госгеонадзора). Большинство карт не имеют четкой связи с информацией первых двух уровней, в результате чего снижается точность понимания единиц авторской легенды и резко снижается возможность разностороннего использования карты. В лучшем случае существует связь геоботанических карт с информацией второго уровня, когда опубликованная карта сопровождается монографией (пояснительным текстом), в которой приводятся разной степени детальности характеристики единиц растительности, отраженные в легенде. Связь опубликованных геоботанических карт с первым уровнем информации как правило отсутствует, т.к. характеристики единиц легенды редко документируются конкретными описаниями. В настоящее время из электронных карт Сибирской растительности известна только Карта Растительности Новосибирской области (1:500.000), выполненная в формате Arc/Info.

В настоящее время существует быстро появляются новые возможности получения разнообразной информации о растительном покрове, имеющей четкие географические привязки методами дешифрирования комической информации, которая благодаря современной вычислительной технике легко трансформируется во многие географические стандарты.

Характер информации этого третьего уровня определяется прежде всего понятием размерности растительного покрова: топологической, региональной, субпланетарной и планетарной.

6 Заключение

В создаваемый атлас, как один из важных объектов био-разнообразия, включены также базы данных, связанные с описанием генетико-селекционных объектов (см. рис. 5): сельскохозяйственных (сорта и породы) и модельных (хириноиды и др.).

В заключении отметим, что в конце апреля 1999 года в Новосибирском академгородке состоялась Четвертое рабочее совещание по электронным публикациям, которое проводилось с использованием *динамической системы публикаций* и на котором были представлены создаваемый атлас и описание используемых технологий (см. более подробную информацию в материалах совещания и в тезисах докладов [4]).

Библиография

- [1] База данных «Зеленая книга Сибири». [<http://www-sbras.nsc.ru/win/elbib/bio/green/>]

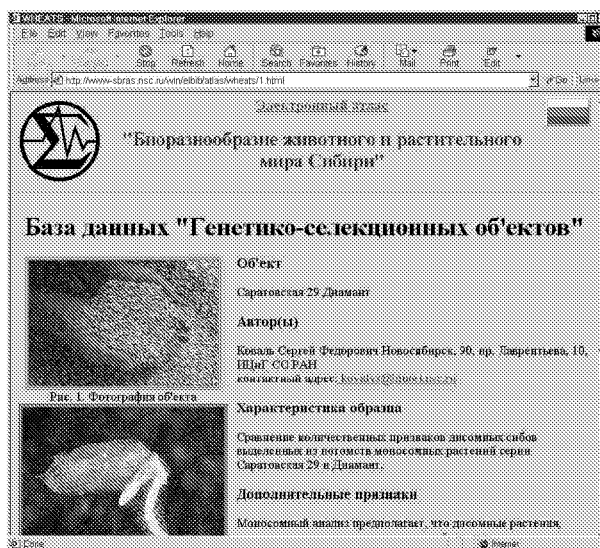


Рис. 5: База данных «Селекционно-генетических объектов»

- [2] Информационный сервер Сибирского отделения РАН. [<http://www-sbras.nsc.ru/>].
- [3] Коропачинский И.Ю., Шокин Ю.И., Шумный В.К., Ермаков Н.Б., Колчанов Н.А., Федотов А.М. *Электронный атлас "Биоразнообразие животного и растительного мира Сибири"*. [<http://www-sbras.nsc.ru/win/elbib/bio/>]
- [4] *Материалы рабочего совещания по электронным публикациям EL-PUB'99* [<http://www-sbras.nsc.ru/win/mathpub/ws-ep99.html>]
- [5] Федотов А.М., Артемов И.А., Ермаков Н.Б., Красников А.А., Потемкин О.Н., Рябко Б.Я., Федотов А.А., Хорев А.Г. *Электронный атлас "Биоразнообразие растительного мира Сибири" // Вычислительные технологии*, Том 3, N 5, 1998. [http://www-sbras.nsc.ru/win/elbib/bio/atlas_flora.html]
- [6] Шокин Ю.И., Федотов А.М. *Распределенные информационные системы // Вычислительные технологии*, Том 3, N 5, 1998. [http://www-sbras.nsc.ru/win/gis/publ/inf_sys.html]
- [7] Шокин Ю.И., Федотов А.М. *Электронная библиотека Сибирского отделения РАН (проект)*. [<http://www-sbras.nsc.ru/win/elbib/>]