

## ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЕ СИСТЕМЫ ДЛЯ ЗАДАЧ БИОРАЗНООБРАЗИЯ

Ю.И. МОЛОРОДОВ, А.М. ФЕДОТОВ

## INFORMATION-ANALYTICAL SYSTEMS FOR BIODIVERSITY TASKS

Yu.I. MOLORODOV, A.M. FEDOTOV

*Институт вычислительных технологий СО РАН, 630090 Новосибирск, пр. акад. Лаврентьева, 6*

*Institute of Computational Technologies, SB RAS, 630090 Novosibirsk, Acad. Lavrentjev av., 6*

*Fax: +7 (383) 330-63-42; e-mail: fedotov@nsu.ru, yumo@ict.nsc.ru*

В работе обсуждаются идеи формирования электронных ботанических коллекций на основе материалов полевых работ. Прототип информационно-поисковой системы позволяет хранить связанную и формализованную информацию о видах растений. Описаны некоторые реализованные коллекции по биоразнообразию, представленные в сети Internet.

**Ключевые слова:** информационные ресурсы, электронные библиотеки, ботаника.

This paper discusses the idea of developing electronic botanical collections on the basis of field study. A prototype of information retrieval system allows to keep interrelated and formalized information on plant species. Some collections concerning biodiversity are available in the Internet.

**Key words:** information resources, digital libraries, botany.

### ВВЕДЕНИЕ

Важнейшей частью информационной поддержки исследований является создание и развитие информационной среды научного сообщества и собственных информационных ресурсов. Информационные ресурсы, переведенные в электронную форму в виде электронных коллекций, атласов и информационных систем, приобретают новое качество, обеспечивая 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, какими являются электронные коллекции и библиотеки.

#### Электронные коллекции и библиотеки

Электронные коллекции и библиотеки представляют собой новую форму хранения, представления и обмена информацией для проведения научных исследований. Для такой формы хранения информации характерны динамичность (возможность обновления) и глобальный доступ (через компьютерные сети). Электронные коллекции и библиотеки объединяют в себе свойства информа-

ционно-поисковой системы, системы обработки, анализа и интерпретации данных и обычной «бумажной» публикации результатов исследований. Развитию концепций, моделей и реализаций электронных библиотек (ЭБ) уделяется большое внимание во всем мире.

На сегодняшний день наиболее эффективным способом решения проблем организации доступа к распределенным информационным ресурсам является организация информации о них в информационных системах, обличенных в форму *электронных библиотек*<sup>1</sup>. Электронные библиотеки — это распределенные каталогизированные информационные системы, позволяющие хранить, обрабатывать, распространять, анализировать, а также организовывать поиск в разнообразных коллекциях электронных ресурсов<sup>2</sup>. В этом отношении,

<sup>1</sup> По-простому, понятие электронной библиотеки заключается в том, что любой ресурс должен быть стандартным образом каталогизирован, снабжен *метаданными* и уникальным идентификатором.

<sup>2</sup> Очевидно, что элементом «хранения» в каталоге электронной библиотеке может быть описание любого ресурса, представленного в глобальной сети передачи данных, будь то документ, база данных, вычислительная программа или компьютерный ресурс.

электронная библиотека определяется, как интегрированная единая среда распределенных ресурсов. Важной проблемой при создании ЭБ (коллекций) 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 библиотек является стандартизация данных, разработка технологических решений и юридических аспектов использования информации, включая вопросы интеллектуальной собственности.

Главным отличием ЭБ от прочих видов ИС является ее способность накапливать и использовать (например, распространять) информацию в виде документов. Принципиально новым подходом является создание таких информационных систем, которые обеспечат эффективный комплексный поиск и анализ информации в коллекциях разнородных данных.

### Схема работы в среде Интернет

Наиболее популярным способом информационного взаимодействия в сети Internet является работа с информационными ресурсами с помощью «тонких клиентов» — Web-браузеров (Internet Explorer, Mozilla Firefox, Opera и др.). Минимальный набор, необходимый пользователю ресурсов, — это наличие подключения к Internet и Web-браузер. При этом тип компьютера (персональный компьютер (ПК), ноутбук, карманный ПК, смартфон и т.д.) и тип операционной системы (Windows, Linux, Unix, MacOS, Solaris) не оказывают влияния на возможности информационного взаимодействия в сети Internet.

Для просмотра информационных ресурсов, предоставляемых Internet-сайтом, пользователь вводит в адресной строке Web-браузера адрес сервера и путь к сайту (например, <http://gis-app.ict.nsc.ru/www>), либо переходит по гиперссылке. В результате этих действий Web-браузер формирует http-запрос, который адресуется указанному серверу. В качестве сервера может служить компьютер с произвольной операционной системой. На нем запущена программа «http-сервер», которая обрабатывает полученный запрос клиента, формирует результат (обычно html-страница), и передает его клиенту, от которого был получен запрос, после чего клиентский Web-браузер показывает пользователю принятую html-страницу (рис. 1). В зависимости от клиентского запроса и от настроек http-сервера клиент может получить заранее подготовленную (статическую) html-страницу, либо html-страницу, сформированную динамически

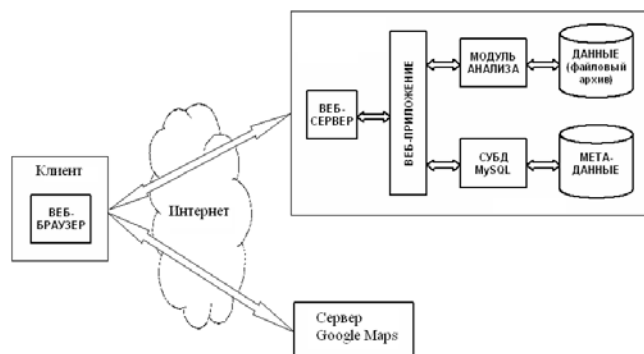


Рис. 1. Схема работы в среде Интернет

(индивидуальный ответ на конкретный клиентский запрос).

Набрав адрес ресурса <http://gis-app.ict.nsc.ru/www>, мы подключаемся к Информационному portalу Института вычислительных технологий СО РАН. В него встроена система управления ресурсами, обеспечивающая администрирование Web-портала, подсистемы управления правами пользователей и доступ к информационным ресурсам портала. Важным его ресурсом является информационная система Центра коллективного пользования регионального спутникового мониторинга окружающей среды (Шокин и др., 2008). Этот ресурс поддерживают специалисты Центра мониторинга социально-экономических процессов и природной среды ИВТ СО РАН. Заинтересованные лица могут обратиться к администрации Института и получить к нему доступ.

### Информационные Системы биологического разнообразия

На сегодняшний день в среде Интернет существует множество сайтов, на которых представлены как электронные каталоги или базы данных по коллекциям различных групп биоразнообразия, так и информационно-аналитические Системы. К ним относится сайт Миссурийского ботанического сада (<http://www.mobot.org>). На нем расположена таксономическая база со ссылками на описания видов и оценку их статуса в наиболее важных таксономических ревизиях. База гербарных образцов содержит более чем 3.5 млн записей (<http://mobot.mobot.org/Pick/Search/most.html>). В Нидерландах уже давно проводится работа по созданию открытой Интернет-информационной системы SYNBIOSYS (Syntaxonomical Biological System) (<http://www.synbiosys.alterra.nl>). Также создается и европейская биологическая информационная система по разнообразию растительности Европы — SynBioSys Europe. В США

работает AGC — Alaska Geobotany Center (<http://www.geobotany.uaf.edu/about/>). Область научных интересов этого центра направлена на изучение северных экосистем на основе использования географических информационных систем, дистанционного зондирования, полевых экспериментов, и выполнение научно-исследовательских проектов. Один из них — Арктический геоботанический Атлас (<http://www.arcticatlas.org>) — направлен на экологические исследования приполярной области Арктики и включает в себя коллекцию геоботанических карт и связанных с ними материалов. Здесь используется иерархический подход к изучению ландшафтной структуры и изменения в нескольких масштабах. Основные области исследований охватывают классификации растительного покрова, анализ растительности и ландшафтной структуры. Изучается изменение экологии через анализ геоботанических карт и динамики снежного покрова в северных регионах. Совместно с Институтом научного растениеводства Бернского университета в Швейцарии и Университетом Фэйрбенкса на Аляске в США выполняется проект по созданию базы данных флоры и растительности Арктики (Patrick Kuss & Skip Walker). Создана база данных «Атлас карт приполярной растительности и геоботанический атлас Арктики» (Skip Walker). В России с 2002 г. развивается проект информационной системы «Биоразнообразие России» <http://www.zin.ru/BioDiv/index.html>. В его создании принимают участие специалисты Зоологического института РАН и их коллеги из Института проблем экологии и эволюции РАН (ИПЭЭ), Ботанического института РАН (БИН) и Института цитологии и генетики СО РАН (ИЦиГ). Основной целью проекта «Информационная система по биоразнообразию» (ИСБР) являлось создание комплекса программных средств и баз данных (БД) для работы с классификацией животного и растительного мира. Предполагается, что они послужат основой формирования информационно-поисковой системы (ИПС) по биоразнообразию России, поддерживающей неоднородные коллекции распределенных информационных ресурсов, содержащих сведения систематического, коллекционного и экологического характера.

В основе реализации проекта — использование реляционной модели баз данных. Вся информация проекта основана на таксономических базах данных, содержащих сведения о номенклатуре и классификации всех организмов. Эти таксономические базы данных используют стандарт ZOOCOD (его ботаническая версия называется PLANTCOD), который описывает способы полноценного представления

в плоских реляционных таблицах многоуровневой иерархии биологических классификаций. В свою очередь географическое распространение организмов описывается отдельно по трем аспектам: административно-территориальному, физико-географическому и зоогеографическому. Отдельные регистрации «таксон-регион» фиксируются в базах данных только с использованием иерархических классификаторов (тезаурусов) регионов каждого из трех аспектов. Из подразделов сайта на сегодня наиболее насыщены информацией «Грибы Российской Арктики», «Лишайники Российской Арктики» и «Флора мхов Чукотки» (<http://www.binran.ru/biodiv/index.htm>).

Отечественные работы в зоологическом направлении представлены информационным проектом Зоологического института РАН — ZOOINT — зоологической интегрированной информационно-поисковой системой ([http://www.zin.ru/projects/zooint\\_r/index.html](http://www.zin.ru/projects/zooint_r/index.html)). Там была разработана информационно-поисковая система «ОКЕАН», включающая стационарную базу данных о местах сбора и поимки морских беспозвоночных и рыб (Смирнов и др., 2007). В сочетании с таксономическим классификатором, содержащим сведения о составе фауны определенного региона и коллекционной базой данных (сведения о месте и способе хранения собранного материала), стационарная база данных позволяет проводить поиск информации по многочисленным и разнообразным запросам.

Примером узкоспециализированного и информационно насыщенного отечественного Интернет-ресурса является бриологический сайт журнала *Arctoa* (<http://www.arctoa.ru>, главный редактор М.С. Игнатов (ГБС РАН)). На сайте, наряду с материалами опубликованных в журнале статей по мохообразным, организована коллекция фотографий мхов и библиотека полезных бриологических ссылок. Одновременно сайт служит для объединения усилий отечественных бриологов в работе над первым изданием «Флоры мхов России». Первый этап коллективной работы — ревизия существующих данных по видовому разнообразию мхов России и сопредельных государств был завершён созданием сводки «Список мхов Восточной Европы и Северной Азии» (Ignatov, Afonina, Ignatova et al., 2006), в которой для каждого вида приводится распределение по крупным административным выделам, снабженное литературными ссылками.

#### **Ботанические информационные ресурсы**

Одним из важнейших ресурсов ГИС-портала ИВТ СО РАН является «Электронный атлас

«Биоразнообразии животного и растительного мира Сибири» (Байков и др., 2000) (<http://www.sbras.ru/win/elbib/bio>). Атлас создавался в рамках программы «Электронная Библиотека Сибирского Отделения РАН» и при поддержке целого ряда проектов РФФИ, и интеграционных проектов СО РАН (Биоразнообразии и динамика экосистем, 2006) (<http://www.sbras.ru/win/elbib/atlas>), в которых участвовали ЦСБС СО РАН, ИЦиГ СО РАН, ИПА СО РАН, ИСЭЖ СО РАН и другие институты Сибирского отделения РАН.

Атлас включает в себя около 50 разнообразных коллекций, посвященных проблемам биоразнообразия, в том числе: «Редкие и нуждающиеся в охране растительные сообщества Сибири — Зеленая книга Сибири», «Электронный каталог растений Сибири» «База данных по медведицам (Insecta, Lepidoptera, Arctiidae) Палеарктики», атлас «Мхи России» и др. Важной компонентой является Web/Gis проект «Ландшафтные карты растительности Западно-Сибирской равнины и юга Восточной Сибири».

При разработке Атласа были поставлены и решены несколько важных задач:

1. Создание электронной библиотеки «Биоразнообразии животного и растительного мира Сибири» (Федотов и др., 1998) предназначалось для сохранения и организации широкого доступа к информации о животных, растениях и экологических характеристиках регионов Сибири. Основные источники информации — данные, собранные исследователями в течение нескольких десятилетий в биологических институтах СО РАН, научные журналы, монографии, учебники, материалы, хранящиеся в зоологических музеях и гербариях, а также результаты и полевые журналы экспедиций и экспериментальных исследований.
2. Создание в рамках библиотеки мультимедийных баз данных со средствами поиска по важнейшим группам животных и растений. В частности, баз данных «Редкие виды животных и растений Сибири», «Охраняемые природные территории», «Определители таксономической принадлежности», «Лекарственные растения», «природно-очаговые болезни», «Генофонд с.-х. растений и животных» и т.д.
3. Создание автоматически актуализируемых баз данных биологических видов и на основе системы классификаторов и оптимизированных определительных (таксонометрических) таблиц. Поддержка распределенной работы и кооперации исследователей, находящихся в разных регионах.

В настоящее время Электронный атлас содержит максимально полную связанную и формализованную информацию о видах растений и животных, необходимую для исследований в экологии, биологии, ботанике, зоологии и генетике и практического применения в работах по охране биосферы. В Атласе собраны данные о видах животных и растений, экологических характеристиках территорий и обеспечения доступа к этим данным через Internet.

Электронный атлас и базы данных разрабатываются в виде распределенной информационной системы, использующей серверы, расположенные в Новосибирске (СО РАН, ИВТ СО РАН и ИЦиГ СО РАН), Якутске (ИКФИА СО РАН) и в других научных центрах СО РАН.

В базы данных включены уникальные коллекции: гербарий, состоящий из описаний, карт, рисунков и фотографий более чем 310 тыс. видов растений Сибири; фитоценотека, содержащая более чем 23 тыс. описаний и фотографий растительных сообществ Сибири; базы данных генофонда сельскохозяйственных и лекарственных (ценных) растений и животных.

#### **База данных «Зеленая книга Сибири»**

База данных содержит информацию о редких и нуждающихся в охране растительных сообществах Сибири в виде унифицированных паспортов описания сообществ. Результаты, представленные в базе данных, являются плодами работ большого количества исследователей (список авторов «Зеленой книги Сибири» (<http://www.sbras.ru/win/elbib/bio/green/db/authors.dhtml>) состоит более чем из 60 человек). Основой для создания базы данных послужила монография «Зеленая книга Сибири», изданная в 1996 г. в Новосибирском отделении издательства «Наука» под редакцией акад. И.Ю. Коропачинского (Зеленая книга Сибири, 1996). Представленные в базе данных сообщества классифицированы по типу сообществ, по названиям, авторам составителям описаний, а так же по ключевым словам. Предусмотрен просмотр списка всех паспортов сообществ в базе данных в алфавитном порядке и поиск по названиям сообществ, авторам, ключевым словам и классификационным признакам.

#### **Атлас «Мхи России»**

В настоящее время активно разрабатывается и пополняется коллекция «Мхи России», (<http://gis-app.ict.nsc.ru/bio>), в которой объединенным коллективом отечественных бриологов ведется аккумуляция данных по листовстельным мхам России.

Создание базы было инициировано в 2004 г. как основа для обобщения и ревизии гербарных материалов к подготовке первого издания «Флоры мхов России» ([www.arctoa.ru](http://www.arctoa.ru)) (рис. 2).

В Атласе аккумулируются данные по листовидным мхам России. Список видов взят из опубликованного списка мхов Восточной Европы и Северной Азии (Ignatov, Afonina, Ignatova et al. 2006. Check-list of mosses of East Europe and North Asia.). В базе данных хранятся гербарные этикетки

«Мхи России». Пополнение ведется коллективными усилиями российских бриологов. Реализованы возможности: поиск информации по любому полю/сочетанию полей или по интервалу географических координат; просмотр найденной информации по отдельным записям или в виде сводной таблицы; отображение результатов запроса на карте Google (рис. 3).

Интерфейс Google Maps API для доступа к сервису из JavaScript программ позволяет добавлять

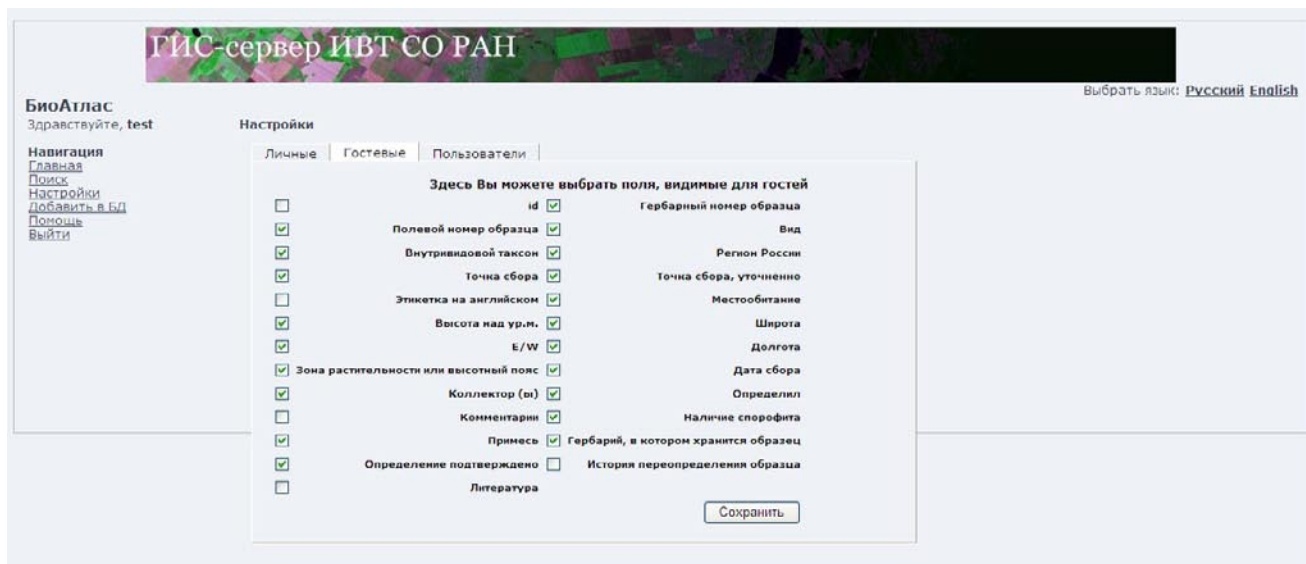


Рис. 2. Страница формирования запросов в Атласе «Мхи России»

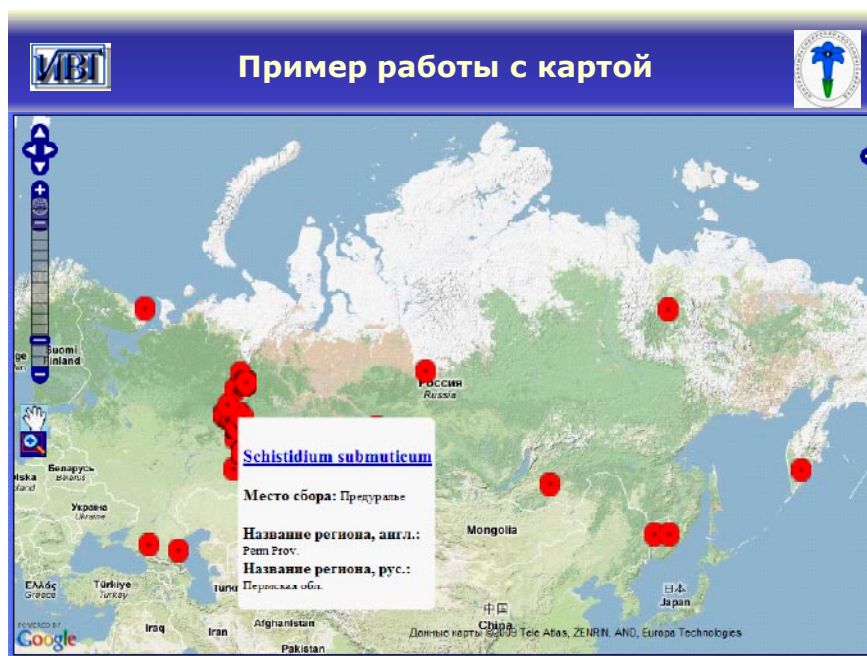


Рис. 3. Пример работы с картой

на карту собственные данные и публиковать ее на веб-странице. Данные пользователя могут включать маркеры, контуры, интерактивные подсказки, всплывающие окна и пользовательские слои в спе-

циальном для Google Maps формате KLM (Keyhole Markup Language). Система управления поведением карты включает стандартные возможности навигации и анимации.

## ЗАКЛЮЧЕНИЕ

Важнейшей задачей современного научного сообщества является сохранение информации об объектах биологического разнообразия. Одним из возможных путей для ее решения является перевод информации с традиционных носителей в электронный вид. Информационные ресурсы, переведенные в электронную форму, приобретают новое качество, обеспечивая их более широкое распространение и эффективное использование. Описаны существующие базы данных по ботанике и биологии, представленные в сети. На примерах Интернет-ресурсов показаны идеи формирования электронных ботанических коллекций на основе

материалов полевых работ. Прототип информационно-поисковой системы позволяет хранить связанную и формализованную информацию о видах растений (<http://gis-app.ict.nsc.ru/bio>).

Содержимое базы данных отображается в интерактивном режиме на ГИС-картах по запросу пользователя. В исходном массиве данных использована ботаническая база данных «Мхи России» объединяющая около 30 тыс. документов.

Работа выполнена при частичной поддержке РФФИ (проекты 09-07-00277, 08-07-00229, 07-07-00271), президентской программы «Ведущие научные школы РФ» (грант НШ-931.2008.9) и интеграционных проектов СО РАН.

## ЛИТЕРАТУРА

Байков К.С., Ермаков Н.Б., Коропачинский И.Ю., Федотов А.М., Хорев А.Г., Шокин Ю.И. К вопросу создания электронной библиотеки «Биоразнообразие растительного мира Сибири» // Вычислительные технологии. 2000. Т. 5. № 2. С. 134–147.

Биоразнообразие и динамика экосистем: информационные технологии и моделирование / Отв. ред. акад. РАН В.К. Шумный и акад. РАН Ю.И. Шокин. Новосибирск. 2006. 648 с.

Вестник КазНУ им. аль-Фараби. Серия: Математика, механика, информатика. Совместный вып. по материалам междунар. конф. «Вычислительные и информационные технологии в науке, технике и образовании». № 4 (59). Ч. III. С. 371–376.

Зеленая книга Сибири. Редкие и нуждающиеся в охране растительные сообщества. Новосибирск. 1996. 397 с.

Смирнов И.С., Пугачев О.Н., Лобанов А.Л., Алимов А.Ф., Воронина Е.П. Электронные коллекции Зоологического

института по морским животным и метаданные // Тр. 9-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Ярославль, 2007. С. 122–127.

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. 1998. Т. 3. № 5.

Флора Сибири. Новосибирск. 1988–2003. Т. 1–14.

Шокин Ю.И., Добрецов Н.Н., Пестунов И.А., Молородов Ю.И., Смирнов В.В., Синявский Ю.Н. Система сбора, хранения и обработки спутниковых и наземных данных Новосибирского научного центра СО РАН // Вычислительные технологии. 2008. Т. 13.