

Образец рыб – первый зоологический образец в информационной системе L.

Мелехин А.В.¹, Федоров Р.К.², Зубова Е.М.³, Кашулин Н.А.³

¹ ПАБСИ КНЦ РАН, Апатиты, melihen@yandex.ru

² ИСТДУ СО РАН, Иркутск, fedorov@icc.ru

³ ИППЭС КНЦ РАН, Апатиты, seelewolf84@yande.ru; kashulin@mail.ru

Аннотация. В рамках развития информационной системы L., основанной на фреймворке Geoservices, для ИППЭС КНЦ РАН (ИНЕР) в 2019-2020 гг. были проведены работы по включению в ИС рыб. Образец рыб стал первым зоологическим образцом в L. – ранее, мы имели дело только с ботаническими материалами.

Ключевые слова: информационная система, зоологический образец, биология рыб.

Fish sample – the first zoological sample in the L. information system

Melekhin A.V.¹, Fedorov R.K.², Zubova E.M.³, Kashulin N.A.³

¹ PABGI KSC RAS, Apatity, melihen@yandex.ru

² ISDCT SB RAS, Иркутск, fedorov@icc.ru

³ INEP KSC RAS, Apatity, seelewolf84@yande.ru; kashulin@mail.ru

Abstract. As part of the development of the L. information system, based on the Geoservices framework, work was carried out for the inclusion of fish in IP for INEP in 2019–2020. The fish specimen became the first zoological specimen in L. – previously, we only dealt with botanical materials.

Key words: information system, zoological sample, fish biology.

Введение

Информационная система (ИС) L. – это инструмент для внесения, хранения, организации, вывода и анализа всех типов данных (данные по наблюдениям, образцам коллекций, литературе и так далее) по всем группам живых организмов. Целями L. является обеспечение полного цикла работы с биологическими данными: от работы в полевых условиях и в коллекциях до организации и анализа данных по любой группе организмов, по всем возможным источникам. Она предоставляет фактические сведения по видовому составу различных территорий на основе материалов коллекций, данных наблюдений, литературных указаний, коллекций культур. Также в L. можно найти фотографии сотен видов различных групп организмов. Исторически L. является результатом перехода ИС «CRIS» на новую техническую платформу (с Drupal на Geoservices – создан Федоровым Р.К. на NodeJS). Исчерпывающую информацию по данным проектам можно узнать на странице ИС L. (<https://isling.org>).

Анализ имеющейся «Базы данных рыб ИНЕР» формирующуюся с 80-х годов прошлого века (сделанной вначале в СУБД «РЕБУС», затем в локальных файлах «Excel», от которого отказались в пользу «Google Документы»), показал необходимость перехода на более сложные и мощные инструменты организации данных. Группой ихтиологов ИНЕР было принято решение использовать в качестве такого инструмента ИС L.

В рамках развития ИС L. для Лаборатории водных экосистем ИППЭС КНЦ РАН (ИНЕР), в 2019-2020 гг. были проведены работы по включению в ИС данных по рыбам, что стало первым опытом размещения зоологических данных в L., где ранее размещались только ботанические материалы.

Результаты и обсуждение

Одной из важнейших черт ИС L., является направленность на сближение «образца» в ИС (дерева данных об образце) с реальными физическими объектами. Ввиду того, что в подавляющем большинстве случаев мы имели дело с многовидовыми (физически неделимыми на одновидовые) образцами, вся архитектура ИС L. была построена с расчетом на работу с такими образцами.

На этапе анализа предметной области нами было выявлено, что образец рыб уже можно сформировать без внесения изменений в код L., во всяком случае, для решения задач группы ихтиологов INEP.

Хранимый физический образец любой коллекции представлен в нашей ИС как класс, наследующий от класса «Все образцы». Такой образец коллекции называется «Фиксированный образец», он отличается от «Все образцы» наличием ссылок на место хранения, локальным номером хранения и т.п. Для удовлетворения нужд различных групп исследователей были созданы классы пользовательских образцов (наследующих от «Фиксированный образец»), отличающихся от базового для них класса наличием экологических или географических дополнительных полей, относящихся ко всему образцу в целом (т.е. ко всем видам в образце). Виды в образце – это таблица класса «Базовый вид», включающий поля: вид, дату и имя определения, картинка, комментарий.

В ходе реализации проекта выяснилось, что зоологический образец имеет ряд особенностей, с которыми мы сталкиваемся впервые.

1. Большое динамическое дерево биологических данных (далее здесь для краткости – «био-хвост» – совокупность дополнительных данных, помещенных в группу полей таблицы, отсутствующих в базовой таблице).

Рыбный образец как первый представитель действительно биологического образца в нашей системе (можно сказать, что остальные образцы в ИС – флористические и экологические) вскрыл проблему «био-хвоста» – ранее небольшие «био-хвосты» встречались лишь в образцах печеночников и мхов, но они уместились в стандартный комментарий к виду. Дерево биоданных для вида рыб оказалось не только в десятки полей, но и весьма ветвистым и неоднородным по содержанию. Эти данные могут включать морфологические, размерно-весовые, физиологические, возрастные характеристики, состав питания, состояния внутренних органов, содержание химических элементов в различных органах и тканях и др. Появление большого количества биоданных заставило использовать для видов рыб не «Базовый вид», а новую таблицу «Виды рыб» (наследующую от «Базовый вид»). В связи с тем, что «Виды рыб» содержат массу полей, специфическую для данной биогруппы, для других организмов эту видовую таблицу использовать нецелесообразно.

Поиски решения (или обхода) задачи динамического «био-хвоста» позволили сделать следующие выводы:

– «био-хвост» должен реализовываться именно через прикрепление к виду даже в «одновидовом» образце рыб – нельзя положить «био-хвост» в сам образец, то есть нельзя «закрывать глаза» на потенциальную множественность видов образца (например, появление в образце неотделимого физически паразита). Принятая нами реализация логически непротиворечива и не повлечет за собой неожиданных проблем в будущем;

– вопрос «био-хвоста» не решается использованием формата DarwinCore или подобных подходов. Даже при искусственном разделении образца на «одновидовые образцы», динамичность «био-хвоста» не позволяет разумно его реализовать внутри строки DarwinCore – эта строка становится потенциально бесконечной и неудобной для разных групп исследователей – специфические поля каждой биогруппы, различные еще и для каждой группы исследователей, будут все добавляться и добавляться, в результате чего «био-хвост» вырастет до неразумных размеров.

На данный момент смена биогруппы «на лету» в списке видов внутри одного образца не реализована, но запланирована как важнейшая перспективная задача, решение которой сделает систему L. «всеядной» даже на уровне образца.

2. Ярко-выраженная биологическая неполнота физического образца.

В случае с образцом рыб (и, вероятно, образцами других животных) данные часто ссылаются на несуществующие объекты: органы рыбы, которые давно утилизированы – хранится только чешуя или плавник и т.п.

Таким образом, образец должен быть виртуальным контейнером, реалистично и при этом, лаконично отражающий состав и структуру физического образца.

Если посмотреть внимательнее на образцы растений, особенно криптогамных, неполнота образца так же характерна для них: мы храним небольшой кусок лишайника длиной 5 см, но в «образце» (в ИС) указан его размер в 30 см. То есть налицо такая же ситуация, как с животными, но выраженная меньше – образец в ИС здесь так же является виртуальным контейнером физической сущности, которой на самом деле нет, как целого объекта (все остальные 25 см лишайника остались в природе).

Технически, неполнота не является проблемой, но именно в процессе анализа образца рыб нам удалось выявить эту теоретическую проблему, которую мы не замечали ранее, что важно для понимания границ образца в процессе проектирования.

3. Направленность зоологических исследований на обработку материала в полевых условиях.

Изучение растений происходит главным образом по гербарным образцам, которые могли пролежать в гербариях десятки лет. Почти все показатели зоологического образца снимаются в природе. При этом, зачастую, регистрируются индивидуальные параметры не по единой для всех образцов схеме, а определяются текущими задачами (например, массовый промер длин рыб) или различными источниками получения образцов. Это не только приводит к неполноте и гетерогенности данных хранимого образца, как указывалось выше, но и задает повышенные требования к ошибкоустойчивости и надежности ИС. Ошибкоустойчивость обеспечивается проверкой полей на типы данных (нельзя ввести буквы в числовое поле, например) и невозможность ввода новых значений в списки выбора (например, нельзя «придумать» новый вид рыбы – его можно только выбрать из имеющихся). Надежность хранения данных обеспечивается регулярными резервными копированиями и самой клиент-серверной архитектурой ИС (данные могут быть потеряны только в случае какого-то глобального катаклизма).

С учетом вышеизложенных моментов, мы построили из имеющихся элементов L. (без внесения изменений в код) образец рыб, который отвечает следующим требованиям (помимо уже указанных надежности и ошибкоустойчивости):

- разделение прав на чтение, внесение и изменение (возможно ограничение на чтение, внесение данных по образцам INEP разрешено только группе ихтиологов INEP, возможно изменение только «своих» образцов обычными пользователями и правка всех образцов модератором группы);
- возможность экспорта данных в формате csv;
- возможность печати этикеток для физических образцов;
- есть история изменений;
- возможность массовой правки (модератором) образцов;
- прикрепление изображений и других файлов, как к образцу, так и к видам в образце;
- возможность связывания с другими образцами, в том числе из других групп и другими (внешними) источниками данных;
- включение видов и образцов в общие списки и карты распространения всех групп организмов, содержащихся в L.

Образец рыб содержит около 200 полей, по этой причине его схема не приводится здесь – все они собраны в связанные таблицы:

1. Main page. География и экология, идентификаторы хранимого образца.
2. Species. Содержит вид рыбы с датой и именем детерминатора. К виду внутри образца прикреплены следующие таблицы:
 - 2.1. General characteristics. Основные морфологические показатели.
 - 2.2. Food. Данные о питании (видовой состав и количественные характеристики).
 - 2.3. Morphology. Все данные о морфологии.
 - 2.4. Pathologies. Данные о патологиях.
 - 2.5. Chemistry. Данные о содержании различных металлов в органах и тканях.Более подробно с образцом рыбы можно познакомиться на странице системы L.

Помимо внесения данных по хранимым образцам рыб, ихтиологи могут пользоваться другими возможностями ИС: внесение данных из литературных источников, наблюдений, данных из сто-

ронних ИС; поиск и анализ открытых данных других исследовательских групп; формирование проектов типа «рыбы региона...». Так же использование L. для организации своих данных позволяет исследователям ссылаться на свободно-доступные первичные данные из публикаций.

Результаты тестирования показывают, что ИС L. может быть использована для организации данных по образцам рыб, а также по любой другой биогруппе. Использование ИС L. обеспечивает высокую степень интеграции данных о рыбах и связанных с ними организмов и позволяет организовывать все необходимые данные по своей группе (без ограничений на географию или источники) внутри одного инструмента – ИС L. Выход из тестового режима, поглощение существующей базы данных и переход группы ихтиологов INER на использование L. планируется на осень 2020 года.

Работа выполнена в рамках темы НИР № 0226-2019-0045 и частично поддержана из средств гранта РФФИ 18-05-60125 Арктика.