

Информационные системы по фондовым коллекциям —
от таксономических классификаторов к публикации данных
в глобальных международных распределенных информационных системах



Роман Халиков, Михаил Дианов

Технический отдел системной поддержки
таксономических исследований ЗИН РАН

Задачи информационных систем по фондовым коллекциям (ИСФК)

- Инвентаризация и каталогизация коллекционных фондов
- Предоставление доступа к фондовым коллекциям для широкой научной общественности
- Полнофункциональные инструменты для повседневной работы кураторов, хранителей, лаборантов
- Сервисы внутренней статистики, отчетов, учета и т. п.
- Повышение статуса института:
публикации, метрики, датасеты, стандарты и т. п.

- Локальные базы данных, персональные СУБД
- Собственные ресурсы — выделенные серверы, публичные хостинг-площадки
- Совместные ресурсы и информационно-поисковые системы на единой площадке головной организации
- Международные глобальные распределенные информационно-поисковые системы и порталы

Преимущества глобальных платформ (внешних решений)

- «Готовое» решение от сторонних специалистов
- Наличие технической поддержки
- Наличие встроенных сервисов (статистика, аналитика, отчеты и т. п.)
- Относительная простота использования
- Относительная оперативность публикации данных
- Потенциальная возможность развивать открытые протоколы, стандарты, алгоритмы
- Широкая известность сервисов в научной среде

Недостатки глобальных платформ (внешних решений)

- Универсальные решения и алгоритмы — недостаточная гибкость для сложных типов данных и высокоспециализированных задач
- Ограниченный функционал встроенных сервисов, отсутствие возможности создать что-то «под себя» (открытая платформа \neq свободный доступ к ресурсам для разработчиков)
- Относительная сложность первоначального внедрения и дальнейшего использования
- Снижение заинтересованности в создании и развитии собственных информационных ресурсов

- Изменение критериев оценки научной деятельности подведомственных ФАНО учреждений
- Существенное увеличение значимости любых собственных ресурсов «под логотипом» учреждений
- Усугубление ситуации с недостатком ресурсов для разработки и последующей поддержки
- Выработка верной стратегии создания и развития ресурсов с оглядкой на «выживание учреждений»
- Глобальные платформы — мощный стимул развития собственных ресурсов, но не альтернатива им

- Клиент–серверная СУБД (MS SQL Server) — индустриальный стандарт структуры данных, построения запросов, средства импорта и экспорта, обеспечения целостности данных, разделяемого многопользовательского доступа и т. д.
- Импорт имеющихся разрозненных данных в единое хранилище
- Разделение ролей:
 - серверная часть — хранилище унифицированных данных для различных задач и проектов;
 - клиентская часть — различные варианты представления данных (интерфейсов пользователя)

Зоологическая иерархия и небиологические данные

- Иерархические таксономические классификаторы в стандарте ZooCod — ядро всех информационных проектов ЗИН:
Animalia — 146575 таксонов 40 рангов, включая †
<http://www.zin.ru/zodiv/animals.asp>
- ZooCod — не только стандарт для таксономии, но и универсальный инструмент для иерархических данных
- Классификаторы географических объектов, административных территорий, морских акваторий и пр. — иерархические данные в стандарте ZooCod

Преимущества иерархических классификаторов

- Классификатор — систематизированный перечень наименованных объектов, каждый из которых имеет уникальный код
- Уникальность объектов — отсутствие дублирования повторяющихся данных и атрибутов объектов
- Быстрая выборка объектов по различным критериям подчиненности как «вниз», так и «вверх» по дереву
- Возможность независимой правки/замены отдельных ветвей дерева
- Возможность сохранения истории ревизий объектов (положение в дереве, подчиненность и т. п.)

- Уникальные коды всех объектов (формат данных bigint, большое целое число) + особый алгоритм расчета — абсолютная уникальность кодов любых созданных и создаваемых объектов
- Гибкая система рангов объектов с возможностью независимого использования различных рангов в разных ветвях древа
- Поддержка синонимии объектов и сохранения истории ревизий объектов
- Алгоритмы сборки полных названий таксонов согласно положениям кодекса номенклатуры

Примеры реализации стандарта ZooCod

UID	Parent_ID	AltParent_ID	Basionim_ID	AltGen_ID	Status	Taxon	LatNam	Author	Year	RanCod	Syn	Seq	Reference
1101950461143927	1101950453890847	1101950453890847	<i>NULL</i>	<i>NULL</i>	1	Euphorbianaee	Euphorbianaee	<i>NULL</i>	<i>NULL</i>	38	<i>NULL</i>	4	<i>NULL</i>
1101950461207617	1101950461143927	1101950461143927	<i>NULL</i>	<i>NULL</i>	1	Euphorbiales	Euphorbiales	<i>NULL</i>	<i>NULL</i>	40	<i>NULL</i>	1	<i>NULL</i>
1101950461270424	1101950461207617	1101950461207617	<i>NULL</i>	<i>NULL</i>	1	Euphorbiaceae	Euphorbiaceae	Juss.	<i>NULL</i>	50	<i>NULL</i>	1	<i>NULL</i>
1101950461830005	1101950461270424	1101950461270424	<i>NULL</i>	<i>NULL</i>	1	Euphorbia	Euphorbia	L.	<i>NULL</i>	70	<i>NULL</i>	9	<i>NULL</i>
1345743303860951	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia borealis	borealis	Baikov	<i>NULL</i>	90	<i>NULL</i>	2	<i>NULL</i>
1345743303920086	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia esula	esula	L.	<i>NULL</i>	90	<i>NULL</i>	3	<i>NULL</i>
1193074896560509	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia fauriei	fauriei	H. Lév. et Vaniot	1908	90	<i>NULL</i>	1	Repert. Spec. Nov. Regni Veg. 5: 281
1345743303983083	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia helioscopia	helioscopia	L.	<i>NULL</i>	90	<i>NULL</i>	4	<i>NULL</i>
1345743304047072	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia humifusa	humifusa	Willd.	<i>NULL</i>	90	<i>NULL</i>	5	<i>NULL</i>
1345743304110740	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia jolkini	jolkini	Boiss.	<i>NULL</i>	90	<i>NULL</i>	6	<i>NULL</i>
1345743304170386	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia komaroviana	komaroviana	Prokh.	<i>NULL</i>	90	<i>NULL</i>	7	<i>NULL</i>
1345743304233831	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia lucorum	lucorum	Rupr.	<i>NULL</i>	90	<i>NULL</i>	8	<i>NULL</i>
1345743304297663	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia pekinensis	pekinensis	Rupr.	<i>NULL</i>	90	<i>NULL</i>	9	<i>NULL</i>
1345743304360527	1101950461830005	1101950461830005	<i>NULL</i>	<i>NULL</i>	1	Euphorbia savaryi	savaryi	Kiss	<i>NULL</i>	90	<i>NULL</i>	10	<i>NULL</i>

UID	Parent_ID	AltParent_ID	Basionim_ID	AltGen_ID	Status	Taxon	LatNam	Author	Year	RanCod	Syn	Seq	Reference
1101016989400944	1101016978410441	1101016978410441	<i>NULL</i>	<i>NULL</i>	1	Pezizales	Pezizales	J. Schröt.	1894	40	<i>NULL</i>	1	in Engler & Prantl, Die Natürliche...
1103153456570181	1101016978410441	1101016978410441	<i>NULL</i>	<i>NULL</i>	1	Pezizales	Pezizales	C. Bessey	1907	40	=	<i>NULL</i>	Nebraska Univ. Stud. 7: 25
1101017684340575	1101016989400944	1101016989400944	<i>NULL</i>	<i>NULL</i>	1	Pezizales incertae sedis	Pezizales incertae sedis	<i>NULL</i>	<i>NULL</i>	50	<i>NULL</i>	6	<i>NULL</i>
1101017684277624	1101016989400944	1101016989400944	<i>NULL</i>	<i>NULL</i>	1	Pezizaceae	Pezizaceae	Dumort.	1829	50	<i>NULL</i>	7	<i>NULL</i>
1101017367657489	1101017684277624	1101017684277624	<i>NULL</i>	<i>NULL</i>	1	Peziza	Peziza	Dill. ex Fr.	1822	70	<i>NULL</i>	21	<i>NULL</i>
1101017377847609	1101017684420391	1101017684420391	<i>NULL</i>	<i>NULL</i>	1	Peziza	Peziza	Fuckel	1870	70	=	<i>NULL</i>	<i>NULL</i>
1101017820763933	1101017691957318	1101017691957318	<i>NULL</i>	<i>NULL</i>	1	Peziza	Peziza	L.	1753	70	=	<i>NULL</i>	<i>NULL</i>
1103418715407592	1101017367657489	1101017362037386	<i>NULL</i>	<i>NULL</i>	1	Peziza acetabulum	acetabulum	L.	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1103418716437907	1101017367657489	1101017634003898	<i>NULL</i>	<i>NULL</i>	1	Peziza acutipila	acutipila	P. Karst.	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1103418716797421	1101017367657489	1101017655977176	<i>NULL</i>	<i>NULL</i>	1	Peziza acuum	acuum	Alb. et Schwein.	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1103418717780128	1101017367657489	1101017615870068	<i>NULL</i>	<i>NULL</i>	1	Peziza aeruginosa	aeruginosa	Vahl	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1188770512183398	1101017367657489	1101017636490586	<i>NULL</i>	<i>NULL</i>	1	Peziza albipes	albipes	Wallr.	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1103418720087460	1101017367657489	1101017639520047	<i>NULL</i>	<i>NULL</i>	1	Peziza albocitrina	albocitrina	Cooke	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1103418720453176	1101017367657489	1101017727853908	<i>NULL</i>	<i>NULL</i>	1	Peziza albiviolascens	albiviolascens	<i>NULL</i>	<i>NULL</i>	90	=	<i>NULL</i>	<i>NULL</i>
1102386687783434	1101017367657489	1101017367657489	<i>NULL</i>	<i>NULL</i>	1	Peziza alcis	alcis	Harmaja	1986	90	<i>NULL</i>	1	Karstenia 26(2): 44

Отсутствие классификатора — что делать?

- Импорт структурированных списков (Excel, Access) в среду СУБД MS SQL Server
- Подготовка и последующий импорт внешних данных — GBIF, Catalogue of Life и т. п.
- Проблема качества данных внешних классификаторов, в том числе классификаторов небиологических объектов
- Проблема выбора подходящих/привычных систем

- Инфологическая структура оригинальной информационно-поисковой системы по образцам фондовых коллекций
- Типовые экземпляры — основа и приоритет
- Модельные группы — 15 различных коллекций
- Библиотека изображений (дигитализация)
- Картирование и геореференсирование
- Публичный веб-сайт — *<http://www.zin.ru/collections/>*

- Унификация, единые алгоритмы хранения и обработки данных при максимальном сохранении принципов нормализации реляционных баз данных
- Коллекционный модуль — каталог коллекции + таксономический регистратор
- Коллекционный образец — первичная сущность, таксономические определение/ревизии — вторичны
- Глобальные модули:
библиотека изображений, регистрация в GenBank
- Двухязычность модулей, словарей, классификаторов

Коллекционные данные — галереи и карты

- Изображения коллекционных образцов и сопутствующих материалов (оригинальные этикетки, каталожные карточки, конверты, рентгеновские снимки, единицы хранения и т. п.)
- Сохранение аутентичных сведений для последующей разносторонней интерпретации первичных данных специалистами

Asteronux niger Djakonov, 1954

Коллекционный номер: 14205-1

Оригинальная этикетка: "Татария", 1948, 14.09, Сп. №2, 303-275
№, 01 (фрагмент на СС)

Место сбора: Тихий океан, запад Камчатки, остров
Итуру, запад от залива

Исследователь: ЮН - ЮНФЮ

Институция: Курно-Самойлова институт

Страна-организатор: СССР

Станция: ИО

Координаты: 44.525000°, 148.244447° — 44.530000°,
148.238333°

Глубина, м: 275-300

Грунт:

Орудие лова: трап промышленный

Номер орудия лова: 24

Номер профиля: 81

Сборщик: Лелюга, Леонид, 14.09.1949

Номер в каталоге ГИНбанка:

Количество особей в единице хранения: 4

Метод фиксации: спирт

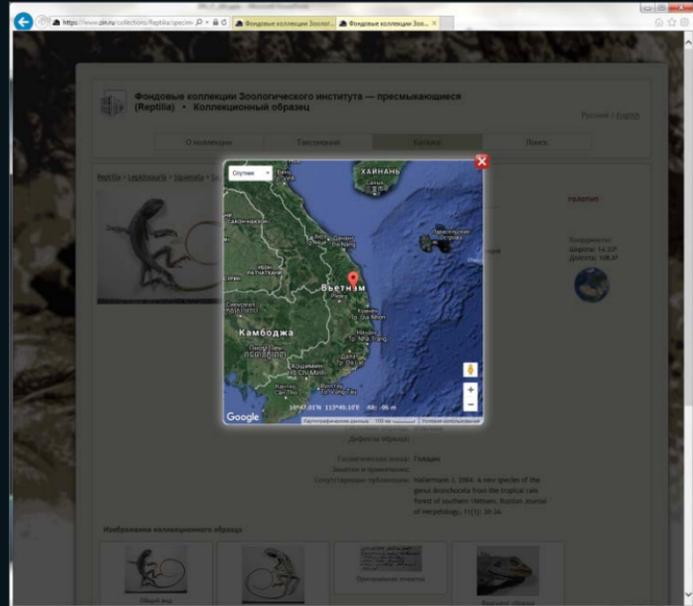
Наименование образца: в коллекции

Сопутствующие публикации, заметки и
примечания:

Изображения коллекционного образца

- Общая фото
- Общий вид
- Оригинальная этикетка
- Каталожная карточка
- Каталожная карточка
- Единица хранения

- Онлайн картирование точек находок на Google Maps
- Перспективы использования международных стандартов геоференсирования National Science Foundation



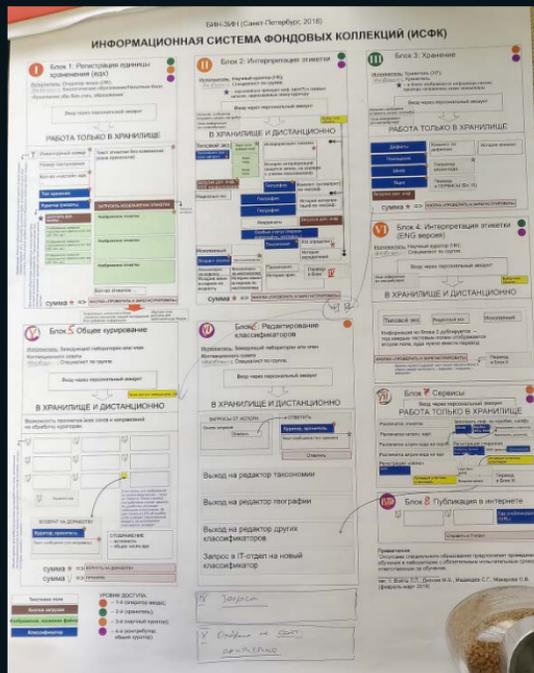
Коллекционный веб-сайт — решение мирового уровня

- Современный лаконичный стиль без сложных форм поисковых запросов
- Унифицированная структура разделов и каталожных карточек коллекционных образцов
- Постоянные прямые ссылки на индивидуальные страницы коллекционных экземпляров и выборок из каталога по таксономической группе
- Полноценная двуязычность контента и функций
- Оптимизация в соответствии со стандартами протокола Open Graph (<http://ogp.me>)

Гибкость и многообразие представления данных

- Интранет ресурсы и приложения — административный интерфейс, средства для работы кураторов по вводу и редактированию данных, любые внутренние служебные инструменты (этикетки, каталоги, инвентарные карточки и т. п.)
- Публичные веб-сайты:
 - различные, в том числе совместные, проекты;
 - разнообразная интерпретация/экспорт данных из единого хранилища
- Выборочная публикация данных в глобальные распределенные информационные системы

Инtranet приложения для кураторов — начало пути



- Ограниченность ресурсов — «делать, чтобы не переделывать»
- Современный уровень технической реализации — аутсорсинг профессиональными программистами
- Подробная документация и детальное техническое задание — договоры/акты с подрядчиками, сохранение всей технической документации для последующих расширения и развития проектов

Инtranет приложения для кураторов — начало пути

- Авторизация для всех операций редактирования (на основе учетных записей Active Directory)
- Собственные средства редактирования для каждой коллекции с разделением ролей пользователей
- Ступенчатые «утверждения» на выполненные части работ
- Полный учет произведенных операций над данными с обязательным указанием оператора
- Распечатка данных на основе шаблонов, сводки по таксономической группе, зданиям–кабинетам–шкафам–коробкам, использование штрих–кодов, журналы выдачи экземпляров специалистам и т. п.

ZIN + BIN в глобальном мире — GBIF Integrated Publishing Toolkit

- Специальный сервис публикации данных GBIF ZIN IPT <http://ipt.zin.ru>
- ZIN IPT + Komarov Botanical Institute RAS Endorsement

Type	Name	Organization	Type	Subtype	Records	Last modified	Last publication	Next publication
Herbarium	Collection of the Herbarium of Botanical and Zoological Institute, Russian Academy of Sciences, St. Petersburg	Zoological Institute, Russian Academy of Sciences, St. Petersburg	Checklist	Inventory Thematic	22	2017-03-30	2017-10-30	2017-11-29 16:58:12
Herbarium	Collection of the Herbarium of Compositae-Labiatae-Gesneriaceae-Myrsinaceae of the Zoological Institute, Russian Academy of Sciences	Zoological Institute, Russian Academy of Sciences, St. Petersburg	Checklist	Inventory Thematic	33	2017-03-30	2017-10-30	2017-11-29 16:58:13
Herbarium	Collection of the Herbarium of Gramineae-Poleaceae-Fragaceae-Myricaceae-Eurycorymbaceae of the Zoological Institute, Russian Academy of Sciences	Zoological Institute, Russian Academy of Sciences, St. Petersburg	Checklist	Inventory Thematic	33	2017-03-30	2017-10-30	2017-11-29 17:00:00
Herbarium	Collection of the Herbarium of Fabaceae-Celastraceae-Scrophulariaceae-Trochodendraceae of the Zoological Institute, Russian Academy of Sciences	Zoological Institute, Russian Academy of Sciences, St. Petersburg	Checklist	Inventory Thematic	58	2017-03-30	2017-10-30	2017-11-29 16:58:08
Herbarium	Herbarium collection of the Zoological Institute, Russian Academy of Sciences	Zoological Institute, Russian Academy of Sciences, St. Petersburg	Occurrence	Specimen	6383	2017-03-30	2017-10-30	2017-11-15 16:45:14

CONTACTS

Zoological Institute, Russian Academy of Sciences, St. Petersburg
Russian Federation
190120, Russia

Wladimir G. Pliginskii
Deputy Head of Institute
E-mail: pliginskii@zin.ru
Phone: +7 812 262 01 00

Natalia B. Anisimova
Administrative Head of Institute
E-mail: anisi@mail.zin.ru
Phone: +7 812 262 01 00

5185 Occurrences

- Качественные специализированные датасеты (наборы данных) «Каталог типовых экземпляров <taxonomic group> фондовых коллекций ЗИН РАН» на основе данных коллекционной платформы
- Опубликовано 5 датасетов — электронных публикаций с DOI:
 - Ophiuroidea, <https://doi.org/10.15468/ej3i4f>
 - Polycestinae, <https://doi.org/10.15468/c3eork>
 - Cosmopterigidae, <https://doi.org/10.15468/sbga6b>
 - Pogonophora, <https://doi.org/10.15468/1mlkdp>
 - Bufonidae, Megophryidae, <https://doi.org/10.15468/crgfcq>

- Catalogue of the type specimens of Pogonophora (Annelida; seu Polychaeta: Siboglinidae) from research collections of the Zoological Institute, Russian Academy of Sciences
<https://doi.org/10.15468/1mlkdp>

Catalogue of the type specimens of Pogonophora (Annelida; seu Polychaeta: Siboglinidae) from research collections of the Zoological Institute, Russian Academy of Sciences

Published by Zoological Institute, Russian Academy of Sciences, St. Petersburg
 80 Roman Smirnov • Aleksey Golikov • Roman Khalikov

155 Occurrences
 73 Accepted names

100% With taxon match
 2 Synonyms

100% With coordinates
 91% Overlap with GBIF Backbone

81% With year
 85% Overlap with Catalogue of Life

155 GSDIFFERENCED RECORDS

Occurrence Detail 1457096

Record	Status Of Record: PRESERVED_SPECIMEN Collection Code: ZSP Dynamic Properties: Original Label: 300n, sect. AH Pogonophora №70 Polasterium australe Smirnov Orig. №8 Catalogue, 2005 Стор "Земельный Материал", 08.03.89, с.1. 4097, m. 3090-3100 м, water Corals Institution Code: ZIN Institution ID: http://gbif.org/ocid/zj9h-zj9h Owner Institution Code: ZIN	Occurrence	Associated References: Cseplová P. B., 2005. Holcové měřítko Polasterium (Pogonophora) ze země Čechy a jejich související skupiny... Catalogue Number: page54-3 Disposition: in collection Individual Count: 3 Occurrence ID: http://catalog.gbif.org/occurrence/1457096 Occurrence Remarks: collection of types Preparations: formalin Record Number: 122 Recorded By: Romanov V.N.
Event	Day: 9 Month: 3 Year: 1989 Event Date: 1989-03-09T00:00:00+0000 Habitat: mud, sand, gravel Sampling Protocol: Slugsby travel	Identification	Date Identified: 2009-01-01T00:00:00+0000 Identified By: R.V. Smirnov Type Status: PARATYPE
Taxon	Kingdom: Animalia Phylum: Annelida Class: Polychaeta Order: Sabellida Family: Siboglinidae Genus: Polasterium Specific Epithet: australe Accepted Name Usage: Polasterium australe Smirnov, 2005	Location	Country: Antarctica Country Code: AQ Decimal Latitude: -60.686111 Decimal Longitude: -54.683333 Depth: 3080 Depth Accuracy: 20 Genetic Datum: WGS84 Higher Geography: South Ocean Locality: Scotia Sea, to the north off

Darwin Core — английский язык для данных по биоразнообразию

- Darwin Core (DwC) — международный стандарт описания данных по биоразнообразию. Со своими ограничениями, жесткой структурой и т. п.
- Основная задача Darwin Core — представление и обмен данными в унифицированном формате
- Экспорт собственных данных в Darwin Core — сугубо техническая задача, не имеющая никакого отношения к инфологической структуре ИСФК и алгоритмам ее работы
- ИСФК ЗИН РАН — публикация данных в GBIF в формате Darwin Core + экспорт выборок в XML-файлы Darwin Core на коллекционном веб-сайте

ЗИН + БИН — соглашение о сотрудничестве

- 8 ноября 2017 г. БИН РАН и ЗИН РАН подписали Соглашение о научно-исследовательском и научно-информационном сотрудничестве
- «Создание межинститутской научно-технической лаборатории, основной целью которой будет развитие информационных систем по фондовым коллекциям (ИСФК) обоих Институтов на основе совместных разработок и технических решений, используя технические мощности, финансовые и интеллектуальные ресурсы каждого из них.»
- Использование аутсорсинга для выполнения технических работ, требующих специальной подготовки и квалификации

- Актуализация и ревизия таксономического классификатора *Vegetabilia* — 38368 таксонов 19 рангов
- Модельные группы коллекций:
 - гербарий сосудистых растений;
 - микологический гербарий;
 - коллекция культур базидиомицетов;
 - палеоботаническая коллекция;
 - ??????
- Выборочная публикация коллекционных данных в GBIF посредством IPT ЗИН РАН

- Внедрение, тестирование и доработка Интранет приложений для работы кураторов с ИСФК на модельных группах коллекций
- Расширение сферы применения актуализированных иерархических классификаторов:
 - региональные чек-листы;
 - конспекты флоры;
 - ??????
- Выборочная публикация данных в GBIF посредством IPT ЗИН РАН

- Глобальная поддержка на уровне дирекций институтов
- Финансирование работ на регулярной основе, безотносительно текущих грантовых проектов и/или государственных заданий институтов
- Заинтересованность во внедрении, тестировании и доработке решений со стороны кураторов коллекций
- Рабочие руки пассионарных личностей на стыке фундаментальной науки и IT-технологий