
Электронный гербарий БСИ ДВО РАН

Выпуск

Dmitry E. Kislov

июл. 30, 2017

1	Администрирование электронного гербария БСИ ДВО РАН	3
1.1	Введение	4
1.2	Структура электронного гербария	5
1.3	Поля формы и их назначение	7
1.4	Персональная страница образца	12
1.5	Этикетка	12
1.6	Этикетка-Конверт	13
1.7	Штрихкодирование образцов	13
1.8	Известные виды	13
1.9	Изображения	15
2	Автоматизация поиска гербарных записей в базе данных	17
2.1	Введение	17
2.2	Параметры запросов	17
2.3	Параметры ответа сервера	19
2.4	Ограничения	23
2.5	Примеры	23
3	Правила и структура каталогов для загрузки гербарных изображений	25
3.1	Мотивация	25
3.2	Структура каталогов	25
3.3	Авторство снимков	26
3.4	Примеры размещения файлов гербарных образцов	26

Версия для печати

Администрирование электронного гербария БСИ ДВО РАН

- *Введение*
 - *Аутентификация и безопасность*
- *Структура электронного гербария*
 - *Гербарные акронимы*
 - *Подразделы гербария*
 - *Группы пользователей и права*
 - * *Куратор*
 - * *Куратор подраздела гербария*
 - * *Регулярный пользователь (наборщик)*
- *Поля формы и их назначение*
 - *Поддержка альтернативного языка*
 - *Вид*
 - *Сходство*
 - *Код образца*
 - *Полевой код*
 - *Акроним*
 - *Страна*
 - *Регион*
 - *Район*
 - *Место сбора*

- *Координаты*
- *Высота*
- *GPS-Based*
- *Собрали*
- *Начало и Конец сбора*
- *Определили*
- *Начало и конец определения*
- *Биоморфологический статус*
- *Подраздел гербария*
- *Заметки*
- *Опубликовано*
- *Переопределения*
- *Дополнительные виды*
- *Сохранить для добавления следующей записи*
- *Персональная страница образца*
- *Этикетка*
- *Этикетка-Конверт*
- *Штрихкодирование образцов*
- *Известные виды*
 - *Поля таблицы Известных видов*
 - * *Название*
 - * *Авторство*
 - * *Род*
 - * *Статус*
 - * *Синоним вида*
- *Изображения*

1.1 Введение

В документе изложены основные принципы работы с формой администрирования, ее основные функции, права пользователей, описание полей, а также рекомендации и соглашения об их заполнении.

1.1.1 Аутентификация и безопасность

Аутентификация осуществляется по предоставленному имени пользователя и паролю.

Пароль рекомендуется сменить после первого успешного входа в панель администрирования. Это можно сделать, перейдя по ссылке в правом верхнем углу окна панели администрирования.

Пароли не хранятся в открытом виде в системе, у администратора системы (или сотрудника с правами суперпользователя) не существует (кроме прямого подбора) способа, чтобы узнать пароль; однако, суперпользователь может изменить пароль любого другого пользователя.

После ста неудачных попыток авторизации IP адрес, с которого были сделаны неудачные попытки, блокируется на час.

Если пользователь в течение 16 часов работы в системе не производит никаких действий – сохранений, переходов на другие страницы, то срок валидности его сессии истекает. По истечении срока валидности сессии пользователю необходимо пройти аутентификацию заново (ввести имя пользователя и пароль), чтобы иметь возможность редактировать образцы.

1.2 Структура электронного гербария

Электронный гербарий представляет собой многопользовательское web-приложение, которое позволяет организовать хранение\изменение\добавление данных с учетом разграничения пользовательских прав, с учетом названий гербария и гербарных подгрупп.

Разграничение гербарных записей в общей таблице осуществляется по двум основным полям – принадлежности определенному гербарию (гербарному акрониму) и подразделу гербария.

Например, может потребоваться необходимость внутри данного гербария (например, VBGI), выделить биоморфологический гербарий, или коллекции грибов и пр. Для этих целей могут использоваться подразделы гербария.

Права пользователей могут распространяться как на образцы, принадлежащие данному акрониму, и\или быть ограниченными определенными подразделами гербария.

Пользовательская поисковая форма (размещаемая на странице <http://botsad.ru/herbarium>) позволяет осуществлять фильтрацию результатов поиска по гербарным акронимам и\или подразделам гербария.

Каждая гербарная запись, независимо от принадлежности гербарному акрониму или подразделу гербария, получает при своем сохранении уникальный, привязанный только к этой записи, целочисленный номер ID. Он назначается системой автоматически. Даже при удалении образца, данный номер уже не будет занят.

1.2.1 Гербарные акронимы

Гербарные акронимы — идентификаторы принадлежности тому или иному гербарию (например, гербарию БСИ ДВО РАН, или гербарию Амурского филиала БСИ ДВО РАН) хранятся в отдельной таблице акронимов, редактирование которой разрешено только суперпользователю (см. о правах суперпользователя в разделе *Группы пользователей и права*).

Пользователи, авторизуемые в системе администрирования, привязаны к тому или иному акрониму. Когда они сохраняют заполненный гербарный образец, доступное только для чтения (изначально пустое) поле акронима, заполняется автоматически в соответствии с этой привязкой.

Каждая запись таблицы акронимов (в настоящий момент таблица хранит записи — VBGI, AmBGI, SAKH) хранит информацию о физическом размещении гербария (его адрес на англ. языке), привязку к пользователям, собственно сокращенное название.

Акронимы используются для разграничения прав кураторов. Куратор гербария привязывается к тому или иному акрониму, но не имеет доступа к гербарным записям принадлежащим другому акрониму.

Информация, содержащаяся в таблице акронимов (адрес гербария, международное обозначение гербария), используется при автоматизированном создании макетов этикеток.

Если требуется внести изменения в таблицу акронимов, например, исправить AmBGI на ABGI, или что-то подобное, изменить адрес гербария, нужно обращаться к сотруднику с правами суперпользователя системы.

1.2.2 Подразделы гербария

Подраздел гербария — необязательное автоматически назначаемое поле при добавлении гербарной записи. Оно может быть пустым, либо заполняется при сохранении образца с учетом привязки конкретного пользователя к подразделу гербария.

Если пользователь с правами куратора привязан к определенному подразделу гербария, то он является куратором только данного подраздела (см. также *Группы пользователей и права*).

Таблица гербарных подразделов с их наименованиями и привязкой к конкретным пользователям редактируется сотрудником с правами суперпользователя.

Если текущий пользователь, редактирующий гербарий, не привязан ни к какому гербарному подразделу, при сохранении поле поле, отвечающая за подраздел, остается пустым.

1.2.3 Группы пользователей и права

Суперпользователь — имеет права на все; за исключением удаления\изменения опубликованных гербарных образцов. Тем не менее, он может изменить статус опубликованного образца на неопубликованный.

Куратор

Куратор гербария — осуществляет мониторинг всех гербарных записей в рамках акронима, к которому он привязан.

Кроме того:

- куратор может быть привязан только к одному гербарному акрониму;
- куратор осуществляет публикацию (и\или снятие с публикации) гербарных образцов;
- куратор может редактировать и просматривать любые гербарные образцы в рамках своего акронима;
- куратор не может удалять виды, рода или семейства, но может их добавлять;
- куратор может исключить вид из результатов поиска в всплывающих подсказках, присвоив ему статус «Deleted»;
- куратор может добавлять\изменять названия уже добавленных видов (при этом изменения скажутся сразу на всех образцах, у которых указан данный вид: например, если вид *Betula mandshurica* в таблице видов переименовать на *Betula davurica*, то все образцы, где был указан до переименования вид *Betula mandsurica*, после переименования будут иметь вид *Betula davurica*);
- куратор может удалять любые образцы в рамках своего акронима

Куратор подраздела гербария

Если имя пользователя с правами куратора привязано к какому-либо гербарному подразделу, то все права куратора ограничиваются только данным подразделом. Остальные права куратора гербарной группы идентичны правам куратора гербария.

Регулярный пользователь (наборщик)

Осуществляет набор данных от своего имени. Акроним гербарной записи при этом назначается автоматически, исходя из привязки пользователя определенному гербарному акрониму.

Регулярный пользователь может быть дополнительно привязан к подразделу гербария. В этом случае, сохранение образца влечет автоматическое назначение подраздела данной гербарной записи.

- регулярный пользователь может просматривать и редактировать\удалять только собственно созданные записи;
- может создавать новые виды;
- не может создавать рода и семейства;
- не может изменять статус видов;
- не может публиковать\снимать с публикации гербарные записи;

Существует еще дополнительный набор прав — «Редактор мультивидовых сборов», который подразумевает, что форма редактирования образца включает поля, отражающие присутствие дополнительных видов в данном гербарном сборе; в обычной гербарии сосудистых растений такая функция не нужна, поэтому она включается только при присвоении пользователю данного набора прав.

1.3 Поля формы и их назначение

Любые поля формы, выделенные жирным шрифтом, обязательны для заполнения.

В форме редактирования гербарных образцов такое поле одно — это поле **Вид**.

1.3.1 Поддержка альтернативного языка

Заполняя данные в цифровой гербарий, прежде всего следует ориентироваться на международное научное сообщество. Таким образом, предпочтительным языком заполнения является английский.

Тем не менее, некоторые поля, а именно *Страна*, *Регион*, *Район*, *Место сбора*, *Высота*, *Заметки*, поддерживают эмуляцию двуязычного заполнения при помощи спецсимвола «|».

Эмуляция двуязычности работает следующим образом. Разграничение языков осуществляется при помощи специального символа «|». Например, поле *Регион* может содержать: «Дальний Восток России|Russian Far East» (*кавычки в форме администрирования не ставятся*). В этом случае, система автоматически будет использовать русский вариант (Дальний Восток России), если пользовательский язык просмотра страницы «русский», и английский вариант (Russian Far East) — в случае, если язык просмотра «английский».

Кроме того, при выборе языка в строках с символом «|» учитываются следующие правила:

- если в строке символ «|» встречается более одного раза, система отображает строку как есть (никакого выбора языка не происходит);
- если в строке символ отделяет пустую подстроку (например, «Владивосток|»), то строка отображается как есть (никакого выбора языка не происходит);
- порядок русско- и англоязычных вариантов в строке относительно символа «|» не имеет значения: т.е. «Дальний Восток России|Russian Far East» и «Russian Far East|Дальний Восток России» являются эквивалентными записями с точки зрения системы;
- система считает русскоязычной ту подстроку относительно символа «|», в которой больше встретилось кириллических символов.

Выбор языка для полей – **Дат сбора\определения** осуществляется автоматически системой, исходя из текущего языка браузера пользователя.

Поля **Определили** и **Собрали** будут использовать автоматическую транслитерацию на английский язык, если они заданы на русском, а язык просмотра пользователем страницы отличен от русского. Данные поля не нужно переводить, если они заполнены на русском, но можно сразу заполнять на английском.

1.3.2 Вид

Заполнить поле вид можно только элементом из всплывающей подсказки. Подсказка формируется по уже добавленным видам в базу, а также более 500k (по состоянию на конец 2016 г) видам из базы данных theplantlist.org.

Поиск выполняется как только набрано 3 и более символов в поле **Вид**; при этом полагается, что первые символы — должны состоять в названии рода, а последующие, если они идут через пробел – видового эпитета. Иными словами, если мы вводим в поисковое поле **Вид**: *Tra*, то в поиске появятся все виды с родами, начинающиеся на *Tra*, при этом количество видимых вариантов будет ограничено 50 вариантами; если в поисковом поле **Вид** будет, например, *Tra ps*, то найдутся все виды, рода у которых начинаются на *Tra*, и, кроме того, видовой эпитет содержит *ps*, т.е., например, *Trapa pseudoincisa* и т.п.

Если требуемый вид отсутствует, нужно нажать рядом кнопку в виде «зеленого плюса» и добавить недостающий вид. Если отсутствует необходимый род и\или семейство, необходимо обратиться к сотруднику с правами куратора и\или суперпользователя, чтобы добавить недостающий род и\или семейство.

У вида можно задать синоним. Добавление синонимов доступно на странице редактирования вида. Подробно об этом можно прочитать в разделе *Известные виды*.

1.3.3 Сходство

Характеризует степень уверенности в определении вида данного образца. Возможные значения «affinis» (aff.), «confertum» (cf.).

Данное поле присутствует также во вкладках *Переопределения* и *Дополнительные виды*.

1.3.4 Код образца

Уникальный в рамках данного акронима числовой код. Этот код опциональный и может не заполняться. Однако, в некоторых публикациях могут быть ссылки на гербарные образцы с указанием этого кода, поэтому он может быть важен.

Если данный код задан, то он отражается на этикетке; если не задан — вместо него используется символ «*».

Регулярный пользователь (наборщик) не имеет прав заполнять данное поле, однако его права могут быть расширены на этот случай.

По умолчанию, права на заполнение данного поля имеют кураторы и суперпользователи.

1.3.5 Полевой код

Опциональный код. Его назначает сборщик гербария; его максимальная длина 20 символов, при этом допустимо использовать любые символы в рамках кодировки utf-8 (т.е. там могут быть и японские иероглифы).

Также может быть важным для ссылок.

Если существует, то отражается на этикетке.

1.3.6 Акроним

Автозаполняемое поле. Оно доступно только для чтения для всех пользователей, за исключением суперпользователя. Суперпользователь может самостоятельно назначить принадлежность образца любому акрониму. Автозаполнение осуществляется на основе привязки пользователей к акронимам.

Поле используется при формировании заголовка этикетки.

1.3.7 Страна

Рекомендуемое к заполнению поле. Необходимо выбрать страну происхождения гербарного сбора. Поиск осуществляется по русскоязычным и англоязычным общепринятым в рамках стандарта ISO перечнем стран (вместо «Российская Федерация» для компактности формы при создании электронных макетов этикеток используется «Россия»).

Данное поле отображается на английском языке на этикетке.

1.3.8 Регион

Отражается на этикетке. Это поле с возможным автозаполнением из того, что уже было введено в базу.

Поддерживает двуязычное заполнение, описанное в разделе *Поддержка альтернативного языка*.

1.3.9 Район

Не отражается на этикетке. Опциональное поле.

Поддерживает двуязычное заполнение, описанное в разделе *Поддержка альтернативного языка*.

1.3.10 Место сбора

Максимальная длина этого поля 600 символов. В этом поле следует также размещать важную информацию об экологических особенностях места сбора.

Поддерживает двуязычное заполнение, описанное в разделе *Поддержка альтернативного языка*.

1.3.11 Координаты

Для заполнения можно использовать флажок на прилагаемой карте google. При изменении позиции флажка, автоматически изменяются и координаты. В правом верхнем углу карты есть и поисковое поле, в которое можно ввести здесь название населенного пункта и флажок переместится в центр этого пункта, если, конечно, такой будет найден (т.е. если название известно google).

1.3.12 Высота

Высота над уровнем моря в метрах.

Поддерживает двуязычное заполнение, описанное в разделе *Поддержка альтернативного языка*.

1.3.13 GPS-Based

Отмечается, если координаты сбора были получены при помощи GPS; это характеристика точности позиционирования сбора; поскольку координаты сбора могут быть получены исходя из описания сбора ручным указанием положения флажка на google-карте.

1.3.14 Собрали

Поле-автоподсказка. Автоподсказка формируется из уже известных уникальных записей, внесенных в базу.

1.3.15 Начало и Конец сбора

Для заполнения может быть использован всплывающий календарик (кнопка справа). Начало и конец указываются если не известна точная дата сбора, но известны, например, даты проведения экспедиции, в ходе которой был осуществлен сбор.

Если дата известна точно, то можно заполнить только одно поле — начало сбора; также можно указать конец сбора, таким же как и начало сбора, либо оставить пустым.

Дата сбора отражается на этикетке, в виде, например, таком: 15 Jul 1998.

Если известен только месяц сбора, то этот факт следует отражать указав начало сбора — первое число месяца, а конец сбора — последнее число месяца. Например, если сбор выполнен в марте, 1999 года, то начало сбора будет 1 марта 1999 г, а конец сбора — 31 марта 1999 г.

Если время сбора указано с точностью до года, следует поступать аналогичным образом — указать первое и последнее числа года — 1 января и 31 декабря.

1.3.16 Определили

Поле-автоподсказка. Работает по аналогии с полем «Собрали». Отражается на этикетке. Если ученых, участвующих в определении много, на этикетке будет указан сокращенный вариант — первые одна, две фамилии (сколько удастся автоматически разместить).

1.3.17 Начало и конец определения

Аналогично началу и концу сбора. Поле не отражается на этикетке.

1.3.18 Биоморфологический статус

Отражается на этикетке, если не пусто. Возможные значения «Dev.stage partly» или «life form». Эти словосочетания и печатаются на этикетке. Специально для биоморфологического гербария БСИ ДВО РАН.

Вполне возможно, оно будет строго привязано к гербарному подразделу «Биоморфологический гербарий», и не будет появляться у пользователей, не привязанных к этой группе.

1.3.19 Подраздел гербария

Автоматически назначаемое поле и доступное только для чтения для регулярных пользователей и кураторов. Может быть пустым. Суперпользователь может редактировать данное поле и указывать подраздел гербария явно. В других случаях поле назначается исходя из привязки пользователя подразделу. Назначается при сохранении образца. До сохранения – не определена.

1.3.20 Заметки

Все что еще мы хотим сообщить о сборе. Для этого здесь доступно 1000 символов.

Поддерживает двуязычное заполнение, описанное в разделе *Поддержка альтернативного языка*.

1.3.21 Опубликовано

Если отмечено, то образец опубликован. Публиковать образцы (как и снимать их с публикации) могут только кураторы гербария\гербарной группы, а также суперпользователь.

1.3.22 Переопределения

Переопределения заполняются, если первоначально определенный вид, потом переопределили. На этикетке, однако, при этом сохраняется первоначальные данные. История переопределений не отражается на этикетке.

В разделе «Переопределения» можно добавить несколько определений, указав соответственно начало (и при необходимости конец) валидности определения. Последним полем блока «Переопределения» является вид, то на что текущий вид был переопределен.

Если этот раздел заполнен, то он отображается на персональной странице образца.

Если поле «Определили» пусто, а история переопределений имеется, то на этикетке будет отображена последняя запись из истории переопределений.

1.3.23 Дополнительные виды

Раздел доступен для редактирования только пользователям с правами «Редактор мультивидовых сборов» (пользователей со специальными правами, у которых в сборах может быть больше одного вида).

Дополнительные виды заполняются по аналогии с полем *Переопределения*.

1.3.24 Сохранить для добавления следующей записи

Если отметить данную позицию, то следующий добавляемый гербарный объект будет иметь уже заполненные поля, как у текущего. Данная функция удобна, когда необходимо добавить несколько гербарных объектов, имеющих однотипное описание: собранных в одинаковом месте, в одинаковых условиях и т.п.

Прежде чем отмечать позицию «Сохранить для добавления следующей записи», важно предварительно сохранить заполненную форму. Поэтому, рекомендуется всегда сначала нажимать «Сохранить и продолжить редактирование», а уже потом отмечать «Сохранить для добавления следующей записи».

Отмечая позицию «Сохранить для добавления следующей записи» пользователь сообщает системе, чтобы она запомнила ID текущего образца и использовала данные его полей для добавления следующих записей.

При добавлении последующих образцов надпись позиции «Сохранить для добавления следующей записи» дополнится фразой «(не этот образец)», которая указывает, что информация для предварительного заполнения полей берется из какой-то другой записи (на которой позиция была отмечена).

Снятие галочки с данной позиции выключает данную функциональность.

Функциональность работает в рамках данной сессии пользователя. Если пользователь завершит работу с системой и потом снова авторизуется, функциональность запоминания полей будет отключена.

Следует иметь в виду, что в текущей реализации данная функциональность не запоминает поля форм *Переопределения* и *Дополнительные виды* (вполне возможно, что в будущем это будет изменено).

1.4 Персональная страница образца

Детальная информация об опубликованном образце доступна по адресу: <http://botsad.ru/hitem/ID>, где «ID» это уникальный код образца, назначаемый системой.

Также, в целях соответствия общим стандартам, валидными ссылками на персональную страницу образца являются ссылки вида: <http://botsad.ru/hitem/ACRONYMXXXXXX>, где ACRONYM – акроним гербария, XXXXX – уникальный числовой код. Примеры: <http://botsad.ru/hitem/VBGI133>, <http://botsad.ru/hitem/VBGI120>

Аналогичный формат доступа к персональным страницам используется в электронном гербарии KEW: <http://apps.kew.org/herbcat/gotoCiteUs.do>

На этой странице указывается история определений, заметки и прочая информация, не вошедшая на этикетку.

Адрес персональной страницы не зависит от акронима и\или гербарной группы.

1.5 Этикетка

Для генерации этикеток в общем списке гербарных записей панели администрирования необходимо выделить образцы, выбрать действие – «Создать этикетки» и нажать «Выполнить».

Этикетка генерируется только для уже опубликованных образцов.

За один запрос можно сгенерировать не более 100 этикеток; они автоматически размещаются оптимально на странице формата А4.

QR-код, размещаемый на этикетке, представляет собой URL персональной страницы образца, также указываемый мелким шрифтом сразу под изображением QR-кода.

Если среди опубликованных образцов для генерации этикетки будут выбраны и неопубликованные – последние будут проигнорированы, а этикетки будут созданы только для опубликованных записей.

Пример этикетки

1.6 Этикетка-Конверт

Некоторые гербарные сборы предпочтительно хранить в конвертах. С этой целью система предлагает специальный формат этикетки «Этикетка-Конверт».

Этикетка-Конверт представляет собой лист формата А4, в нижней трети которого, представляющей лицевую сторону конверта, располагается подробная информация о сборе, а также штрих-код, оформленный в соответствии с правилами, описанными в разделе *Штрихкодирование образцов*; в центре страницы, соответствующей оборотной стороне конверта, размещается QR-код, кодирующий ссылку на персональную страницу образца;

Для генерации этикеток в общем списке гербарных записей панели администрирования необходимо выделить образцы, выбрать действие – «Создать этикетки-конверты» и нажать «Выполнить».

За один запрос можно сгенерировать не более 100 этикеток-конвертов.

Первый вид, указываемый на этикетке-конверте – основной вид сбора, далее, каждый на новой строке, идут дополнительные (сопутствующие) виды.

После располагается блок информации о сборе; в этом блоке отображается содержание полей: Начало сбора, Широта, Долгота, Высота, Собрали, Страна, Регион, Район, Место сбора, Определили.

Пример этикетки-конверта

1.7 Штрихкодирование образцов

Назначение инвентаризационных номеров важный этап систематизации гербарных накоплений. Штрихкод размещается на гербарном листе перед его сканированием и представляет собой уникальный идентификатор данной гербарной записи, формат которого в текущий момент принимается мировым научным сообществом.

Для генерации страницы штрихкодов в общем списке гербарных записей панели администрирования необходимо выделить образцы, выбрать действие – «Создать штрихкоды» и нажать «Выполнить».

За один запрос можно сгенерировать не более 100 штрихкодов; они размещаются автоматически оптимальным образом на странице формата А4. Если штрихкоды не помещаются на одной странице, генерируется многостраничный pdf-документ.

Штрихкоды можно генерировать для всех (не обязательно опубликованных) образцов, внесенных в базу.

В качестве алгоритма для создания штрихкодов используется CODE39, а кодируемая строка имеет вид ACRONYMXXXX, где XXXX – уникальный числовой код (ID) образца внутри данного акронима. Под штрихкодом дублируется кодируемая им строка (размер шрифта фиксирован). Над штрихкодом приводится название организации (размер шрифта динамический, выбирается так, чтобы название организации не выходило за границы штрихкода). Оформление штрихкода сделано по образцу электронного гербария KEW.

Пример документа со штрихкодами

1.8 Известные виды

Все известные виды представлены в трех таблицах – таблице семейств, таблице родов и, собственно, названий видов. Названия видов с авторами привязаны к таблице родов, записи таблицы родов – привязаны к таблице семейств.

Таблица названий видов используется для формирования подсказок при заполнении поля **Вид** формы гербарного образца. Каждая запись таблицы видов имеет дополнительный статус – «From plantlist» (из базы theplantlist.org), «Approved» (проверенный), «Deleted» (удаленный) и «Recently added» (новый, недавно добавленный).

Названия видов, имеющие статус «From plantlist» или «Approved» считаются доверенными, и образцы, в которых участвуют такие виды, могут быть беспрепятственно опубликованы куратором гербария.

В случае, если название вида имеет статус «Recently added», и оно участвует в гербарной записи, такую гербарную запись опубликовать не получится.

Чтобы опубликовать такую запись необходимо, чтобы куратор (или суперпользователь) изменил статус вида (проверил вид) на «Approved».

Ни куратор, ни регулярный пользователь не может полностью удалить вид из таблицы известных видов. Вместо этого, куратор может изменить статус вида на «Удаленный» («Deleted»). Виды, имеющие статус «Deleted», не участвуют во всплывающих списках-подсказках при заполнении полей формы гербарного образца. В таблице видов отображаются все виды, в том числе и имеющие статус «Deleted». Таким образом, статус «Deleted» должен использоваться чтобы ограничить результаты поиска во всплывающих подсказках, что может быть полезным чтобы исключить устаревшие и/или неправильные названия видов.

Регулярный пользователь (наборщик гербария) не может изменять статус вида.

На странице редактирования вида можно задать его синоним. К данному виду можно присоединить только один вид-синоним. Если вид имеет несколько синонимов, можно поступить следующим образом. Допустим вид А имеет синонимы В, С, D. Тогда, редактируя страницы видов В, С, D можно в них указать, что они являются синонимами вида А. При выполнении поиска с учетом синонимов, информация об образуемых этим способом классах эквивалентности (синонимичности) видов будет использована при формировании запроса к базе гербарных образцов.

1.8.1 Поля таблицы Известных видов

Название

Название вида – это видовой эпитет. Значение поля хранится в нижнем регистре. Если Вы введете в данное поле, например, **Davurica**, значение будет автоматически переведено в нижний регистр, т.е. **davurica**. Недопустимо включать в данное поле авторов вида (хотя бы потому, что авторы записываются с учётом регистра). Данное поле может включать информацию о вариациях вида или подвидах, например, **yokogurensis subsp. fragilifolia**.

Авторство

Авторство вида. Примеры (через точку с запятой): Maxim; L.; Kom.; (Moench) Mold.; Stephani Данное поле хранится с учетом регистра.

Род

Название рода. Должно выбираться из выпадающего списка предложенных названий. Если нужное наименование рода отсутствует, его необходимо добавить в систему нажав «Добавить/add» (кнопка «зеленый плюс» справа от поля; кнопка может отсутствовать, если Ваш уровень прав не позволяет добавлять **Рода**);

Статус

Значения полей описано выше.

Синоним вида

Если синоним у вида отсутствует, поле должно быть пустым. Синоним выбирается из выпадающего списка известных системе видов. Если требуемый вид отсутствует, его можно добавить используя кнопку «Добавить/add» («зеленый плюс» справа от поля);

1.9 Изображения

Подготовка изображений для привязки их к гербарным образцам регламентируется отдельным *документом*.

Автоматизация поиска гербарных записей в базе данных

- *Введение*
- *Параметры запросов*
- *Параметры ответа сервера*
 - *Структура массива данных*
 - * *История переопределений и дополнительные сборы*
- *Ограничения*
- *Примеры*

2.1 Введение

Данный документ представляет собой описание HTTP-API (программного интерфейса по протоколу HTTP) для работы с электронным гербарием, работающим под управлением данного web-приложения. HTTP-API не предназначено для изменения данных, а может использоваться только для их чтения.

2.2 Параметры запросов

При формировании поисковых запросов к серверу допустимы только GET-запросы (протокол HTTP(S)).

Одновременное задание полей при формировании GET-запроса приводит к поисковому запросу типа «И»: т.е. если задать одновременно *colstart=2016-01-01* и *collectedby=Бар*, то найдутся все записи, для которых начало сбора позже 1 января 2016 г и у которых в поле «Собрали» встречается последовательность символов „бар“ (без учета регистра).

Формирование запросов типа „ИЛИ“ не реализовано в текущем HTTP-API, но может быть эмулировано серией дополнительных запросов к базе данных.

Возможны следующие параметры GET-запроса, по которым осуществляется поиск:

- **family** — название семейства; название семейства с точностью до регистра записи должно совпадать с запрашиваемым;
- **genus** — название рода; название рода с точностью до регистра записи должно совпадать с запрашиваемым; в случае противоречивости значений параметров **family** и **genus** выводится сообщение об ошибке – отсутствии образцов, удовлетворяющих условиям поиска;
- **species_epithet** — видовой эпитет; поиск осуществляется без учета регистра; условие поиска считается выполненным, если запрашиваемая строка включается в поле «видовой эпитет» основного вида гербарного сбора;
- **place** — место сбора; выполнение условия поиска: включение запрашиваемой строки без учета регистра хотябы в одно из полей данных о гербарном сборе, поля: место сбора, регион, район, примечание.
- **collectedby** — коллекторы; выполнение условия поиска: включение запрашиваемой строки без учета регистра (поиск также автоматически выполняется по транслитерированной на английский язык строке);
- **identifiedby** — определители; выполнение условия поиска: включение запрашиваемой строки без учета регистра (поиск также автоматически выполняется по транслитерированной на английский язык строке);
- **country** — название страны, места сбора образца; включение запрашиваемой строки без учета регистра в название страны на русском или английском языках (используются названия стран, принятые в стандартах **ISO3166-1-ru** или **ISO3166-1-en**; для Российской Федерации используется сокращенное название Россия);
- **colstart** — дата начала сбора; задается в формате **yyyy-mm-dd**
- **colend** — дата окончания сбора; формат аналогичен параметру **colstart**;
- **acronym** — название гербарного акронима; выполнение условия поиска – точное совпадение акронима гербария без учёта регистра;
- **subdivision** — название гербарного подраздела; выполнение условия поиска: включение запрашиваемой строки без учета регистра;
- **latl** — минимальное значение широты сбора; допустимый диапазон значений (-90.0, 90.0); используется для поиска по прямоугольной области сбора;
- **latu** — максимальное значение широты сбора; допустимый диапазон значений (-90.0, 90.0); используется для поиска по прямоугольной области сбора;
- **lonu** — максимальное значение долготы сбора; допустимый диапазон значений (-180.0, 180.0); используется для поиска по прямоугольной области сбора;
- **lonl** — минимальное значение долготы сбора; допустимый диапазон значений (-180.0, 180.0); используется для поиска по прямоугольной области сбора;
- **synonyms** — возможные значение false или true; отсутствие параметра в GET-запросе приравнивается к false; индикатор поиска с учетом синонимов; при поиске с учетом синонимов важно задать поля **genus** и **species_epithet**; true – означает выполнять поиск по синонимам;
- **additional** — возможные значения false или true; отсутствие параметра в GET-запросе приравнивается к false; индикатор поиска по дополнительным видам внутри гербарных сборов; true – означает искать совпадения в дополнительных видах;

- **id** — ID образца; при указании в запросе данного поля, все остальные поля игнорируются; возвращается информация только об образце с заданным ID, если такой был найден, в противном случае – выводится сообщение об ошибке;
- **fieldid** — полевой номер образца;
- **itemcode** — инвентарный номер образца (используется в гербарном хранилище);
- **authorship** — авторство основного вида; условие выполняется, если строка авторства основного вида гербарного сбора содержит в себе значение передаваемое в данном параметре (сравнение производится без учета регистра записи);

Замечание При поиске по полям **collectedby** и **identifiedby** выполняется только односторонняя транслитерация: т.е. если, например, задано условие поиска ****collectedby** = Бакалин**, то это эквивалентно поиску ****collectedby** = Бакалин** или ****collectedby** = Bakalin** (найдутся все записи на русском и английском языках). Если задано условие поиска ****collectedby** = Bakalin**, то найдутся только записи, где в поле «сборщики» встречается строка «bakalin» (без учёта регистра).

2.3 Параметры ответа сервера

Ответ сервера на поисковый запрос представляет собой **JSON-форматированный** текст, передаваемый по протоколу HTTP, и имеющий следующие параметры:

- **errors** — массив ошибок, возникших при обработке поискового запроса.
- **warnings** — массив предупреждений, возникших при обработке поискового запроса. Предупреждениями являются различные поисковые ситуации: например, отсутствие данных, удовлетворяющих текущему поисковому запросу, игнорирование тех или иных поисковых параметров, при их противоречивости и т.п.
- **data** — массив структурированных данных гербарных записей, удовлетворяющих текущему поисковому запросу.

2.3.1 Структура массива данных

Параметр **data** представляет собой массив данных, удовлетворяющих текущему поисковому запросу.

Он имеет следующую структуру, описывающую текущий гербарный сбор:

- **family** — название семейства (заглавными буквами, на латыни);
- **family_authorship** — автор семейства;
- **genus** — название рода;
- **genus_authorship** — автор рода;
- **species_epithet** — видовой эпитет;
- **species_id** — ID вида образца; не путать с ID текущей гербарной записи. ID текущей гербарной записи однозначно характеризует данную оцифрованную гербарную запись. ID вида образца, только вид. Гербарных записей, содержащих какой-либо вид может быть много.
- **species_authorship** — автор вида;
- **species_status** — текущий статус вида; определяет степень признанности данного вида, точнее триплета (род, видовой эпитет, авторство вида) в научном сообществе на настоящее время. Возможные значения данного параметра 1) «Recently added» — вид недавно добавлен и, скорее, не проверялся специалистом; название вида с таким статусом может быть устаревшим, либо

содержать ошибки; 2) «Approved» — название вида подтверждено специалистом; 3) «Deleted» — вид имеет ошибку в записи, или его название устарело и не используется; 4) «From plantlist» — название импортировано из базы <http://theplantlist.org>.

- **species_fullname** — полное название вида, т.е. Род + видовой эпитет + авторство.
- **id** — уникальный идентификатор данной гербарной записи; всегда целое число;
- **gpsbased** — получены ли данные о географической привязки места сбора образца с помощью GPS (значение **true**), либо другим способом (**false**); следует иметь ввиду, что у многих образцов, даже при **gpsbased** равном **false**, координаты, если таковые заданы, были получены при помощи GPS; это связано с тем, что не все отмечают соответствующее поле (**gpsbased**) при заполнении электронной формы образца;
- **latitude** — широта, градусы; географическая координата места сбора в системе WGS-84;
- **longitude** — долгота, градусы; географическая координата места сбора в системе WGS-84;
- **fieldid** — полевой номер образца;
- **itemcode** — инвентаризационный номер, используемый в гербарном хранилище;
- **acronym** — гербарный акроним, которому принадлежит данная гербарная запись (для большинства записей поле имеет значение **VBG1**);
- **branch** — подраздел гербария внутри акронима; иногда удобно иметь подразделы внутри общей гербарной базы: например, «гербарий грибов», «биоморфологический гербарий» и т.п.;
- **collectors** — текстовая строка: сборщики образца;
- **identifiers** — текстовая строка: те, кто определил вид гербарного сбора;
- **devstage** — стадия развития; определена для биоморфологического гербария; возможные значения: Development stage partly, Life form, или пустое поле;
- **updated** — дата последнего изменения записи в базе данных;
- **created** — дата создания записи (т.е. занесения её электронную базу данных);
- **identification_started** — дата начала определения вида;
- **identification_finished** — дата окончания определения вида; дата определения вида задана в виде интервала, поскольку не всегда может быть указана точная дата, а например, только месяц, или время проведения какой-либо экспедиции;
- **country** — название страны сбора образца;
- **country_id** — числовой идентификатор страны сбора образца;
- **altitude** — высота над уровнем моря места сбора образца; значение представляется собой строку, не всегда однозначно определяющую реальную высоту сбора. Возможны, например, варианты: 100-300 м, 120 м, 400, 300-400 и т.п.
- **region** — регион сбора;
- **district** — район сбора;
- **details** — экологические условия сбора, дополнительные уточнения не вошедшие в поля регион и район;
- **note** — примечание; может содержать информацию о месте сбора, экологических условиях и т.п.;
- **dethistory** — представляет собой массив — историю переопределений вида гербарного сбора;
- **additional** — некоторые гербарные сборы могут содержать более одного вида; данный массив описывает характеристики каждого из них.

- **images** — перечень изображений, относящихся к гербарной записи; список имеет следующий формат: [] — пустой список — означает отсутствие изображений; [(„http://someresource.com/path/to/image1“, „image1 type“, „meta information1“), („http://someresource.com/path/to/image2“, „image2 type“, „meta information2“), ...]

- *http://...* — первое поле записи об изображении, указывает на путь, по которому можно загрузить изображение через Интернет;
- *image type* — тип изображения, символ, „p“ или „s“, что означает „p“ = „place“ — изображение является фотографией места сбора (относится к месту сбора);

„s“ = „sheet“ — изображение является фотографией (сканом) гербарного листа;

- *meta information* — json-форматированная строка, содержащая дополнительную информацию об изображении, например, авторство снимка.

В случае авторства снимка, например, строка может иметь вид: «{„photographer“: „Pavel V. Krestov“, „organization“: „VBGI“}»; Мета информация об изображении может включать произвольные поля, например, разрешение снимка, тип камеры и т.п.

Поля **region**, **district**, **details**, **note**, **altitude** могут быть заполнены с поддержкой двуязычности с использованием спецсимвола «|». Например, строка, возвращаемая в поле **region**, может быть такой «Russian Far East|Дальний Восток России». Это означает, что относительно символа «|» даётся русско- и англоязычный варианты строки. Дальнейшая обработка значений таких строк ложится на пользователя системы, которому решать какую из компонент строки относительно символа «|» оставить, а какую — удалить. Система HTTP-API не принимает таких решений.

Структура массивов **dethistory** и **additional** приводится ниже.

История переопределений и дополнительные сборы

История переопределений

Каждый элемент массива «История переопределений» (**dethistory**) представляет собой описание попытки определения (переопределения) вида в текущем гербарном сборе и имеет следующие поля (значения полей, характеризующих вид, аналогично описанным выше):

- **valid_from** — дата валидности определения;
- **valid_to** — дата окончания валидности определения; поле может быть не задано, что означает, что предполагает, что определение актуально в настоящее время;
- **family** — название семейства;
- **family_authorship** — авторство семейства;
- **genus** — название рода;
- **genus_authorship** — автор рода;
- **species_epithet** — видовой эпитет;
- **species_id** — ID вида образца;
- **species_authorship** — автор вида;
- **species_status** — текущий статус вида;
- **species_fullname** — полное название вида;

Сроки валидности вида (**valid_from**, **valid_to**) позволяют корректно описать любые его последующие переопределения.

Примечание Если в гербарном сборе представлен не один вид, то массив «История переопределений» представляет собой историю переопределений основного вида.

Дополнительные виды

Каждый элемент массива «Дополнительные виды» (**additional**s) представляет собой описание вида, находящегося в данном гербарном сборе. Каждое из таких описаний имеет поля, аналогичные записям из **Истории переопределений**:

- **valid_from** — дата валидности определения;
- **valid_to** — дата окончания валидности определения; поле может быть не задано, что означает, что предполагает, что определение актуально в настоящее время;
- **family** — название семейства;
- **family_authorship** — авторство семейства;
- **genus** — название рода;
- **genus_authorship** — автор рода;
- **species_epithet** — видовой эпитет;
- **species_id** — ID вида образца;
- **species_authorship** — автор вида;
- **species_status** — текущий статус вида;
- **species_fullname** — полное название вида;

Таким образом, массив «Дополнительные виды» позволяет хранить информацию о видах в гербарном сборе, сопутствующих данному основному виду (выделенному из экспертных соображений в качестве основного), а указание валидности позволяет описать переопределения (если таковые имеются) каждого из таких видов.

Пояснение и интерпретация

Рассмотрим для примера следующий массив «Дополнительных видов» (для краткости выписаны не все поля):

```
[
  {'genus': 'Quercus', 'species_epithet': 'mongolica', ... , 'valid_from': '2015-05-05', 'valid_to': '2016-01-01'},
  {'genus': 'Quercus', 'species_epithet': 'dentata', ... , 'valid_from': '2016-01-01', 'valid_to': ''},
  {'genus': 'Betula', 'species_epithet': 'manshurica', ... , 'valid_from': '2015-05-05', 'valid_to': ''},
  {'genus': 'Betula', 'species_epithet': 'davurica', ... , 'valid_from': '2015-05-05', 'valid_to': ''}]
```

Если сегодня, например, 1 сентября 2015 года (2015-09-01), то массив дополнительных видов состоит из *Quercus mongolica*, *Betula manshurica* и *Betula davurica*, а *Quercus dentata* является неактуальным определением на данный момент времени.

Если сегодня 2017 год, например, 2017-01-01, то неактуальным оказывается *Quercus mongolica*, и, таким образом, актуальными видовыми составляющими сбора являются *Quercus dentata*, *Betula manshurica* и *Betula davurica*.

Примечание Массив «Дополнительные виды» предназначен только для описания дополнительных видов; основной вид не указывается в дополнительных видах.

2.4 Ограничения

Поскольку поисковому запросу пользователя может удовлетворять большой объём данных, для формирования ответа сервера может потребоваться значительное время.

Чтобы снизить нагрузку на сервер, вызванную вероятно долгими keep-alive HTTP-соединениями, действуют ограничения.

Количество одновременно возможных соединений для сервиса автоматизированного опроса гербарной базы определяется текущим значением параметра `JSON_API_SIMULTANEOUS_CONN`.

По превышении этого количества, сервер не обрабатывает поисковые запросы, а возвращает сообщение об ошибке.

На запросы, содержащие **id**, данное ограничение не действует, поскольку получение информации об объекте по его **ID** — не ёмкая в плане ресурсов операция.

Информация о **неопубликованных** образцах не выводится; при попытке получить информацию о неопубликованном образце по его **ID** выводится ошибка.

2.5 Примеры

Для проверки работы системы и получения json-ответа сервера достаточно передать поисковый запрос в url браузера.

Например, переход по ссылке

<http://botsad.ru/hitem/json/?genus=riccardia&collectedby=bakalin>

приведет к появлению на экране браузера json-ответа, содержащего информацию о всех сборах — представителей рода *Riccardia*, в строке, содержащей информацию о сборщиках которых встречается *bakalin*.

При указании **id** в **GET** запросе, все остальные поисковые поля игнорируются и выводится информация о гербарном образце с указанным **id**:

<http://botsad.ru/hitem/json?id=500>

<http://botsad.ru/hitem/json?id=44>

<http://botsad.ru/hitem/json?id=5>

Правила и структура каталогов для загрузки гербарных изображений

3.1 Мотивация

Поскольку каждое гербарное изображение предполагается представлять в высоком разрешении, что повлечет за собой увеличение требуемого места на сервере, хранение таких изображений будет отдельной задачей, не связанной с заполнением электронного гербария.

Выделение процесса создания гербарных изображений в независимый процесс, позволит также организовать обработку таких изображений (предположительно объемом более 600 GB), исключив дополнительную нагрузку на основной сервер.

Таким образом, это не только позволит сэкономить место на основном сервере, но и позволит независимо наполнять базу изображениями, что важно в плане ее скорейшего ввода в рабочий режим.

Формирование финальных web-страниц гербарных образцов будет в этом случае осуществляться с подгрузкой соответствующих изображений, которые должны быть строго структурированными и привязанными к конкретным гербарным образцам. С целью такой привязки, а также удобства последующей автоматизированной обработки изображений (создания малых их копий для предпросмотра и т.п.) ниже приводится система соглашений о размещении и наименовании изображений.

3.2 Структура каталогов

Изображения, привязанные к гербарным образцам, должны иметь имена определенного формата и размещены по каталогам следующим образом:

Общее правило названия файлов и структуры каталогов:

ACRONYM/[ID, CODE]/[SET/OF/NESTED/FOLDERS]/dddddd_dd<p>.ext

Пояснения:

- ACRONYM – папка, указывающая на принадлежность изображений гербарии данного акронима (у нас это VBGI).

- [ID, CODE] – ID и/или CODE; внутри папки акронима гербария находятся папки ID, CODE; Кроме ID и CODE – другие названия недопустимы, если внутри папки акронима, кроме этих двух (или какой-либо одной из этих двух папок) имеются посторонние папки/файлы, структура каталогов считается неправильной. В папке ID размещаются изображения, привязанные к гербарным образцам по полю ID электронной базы данных (это поле назначается автоматически системой при сохранении гербарных образцов); В папке CODE размещаются изображения, привязанные к гербарным образцам по полю CODE (это поле назначает куратор гербария, оно уникально внутри данного акронима);
- [SET/OF/NESTED/FOLDERS] – множество вложенных папок с изображениями; вложения делаются лишь для удобства человека, создающего гербарные изображения; например, если сегодня 16 февраля 2017 года кто-либо создал некий набор изображений гербарных образцов, привязанных по ID, он может создать в папке ID вложенную папку, например, 16.02.2017, и уже туда поместить соответствующие изображения. В следующий день, он также может создать папку 17.02.2017 и т.д. Можно также выполнять разделение по семействам/родам. Наличие вложенных папок не обязательно, но они позволяют структурировать расположения файлов, – это прежде всего важно для сканирующего (создающего) изображения гербарных листов человека.
- dddddd_dd<p>.ext – представляет собой имя файла изображения, где:
 - до символа _ идет числовой ID либо CODE гербарного образца, к которому привязывается изображение;
 - после символа _ идет номер изображения данного гербарного образца; это делается для того, чтобы отличить имена файлов гербарных изображений в случае, если одному гербарному образцу соответствует несколько изображений; нумерация изображений строго идет от 1.
 - <p> – опциональный символ, указывающий, что данное изображение является изображением места сбора, а не гербарного образца.
 - .ext – расширение файла (обычно, .jpg)
 - ПРИМЕРЫ ВАЛИДНЫХ ИМЕН ФАЙЛОВ: 348234_1.jpg, 12345_1.jpg, 12345_1p.jpg, 83727_1.jpg, 83727_2.jpg

3.3 Авторство снимков

Если возникает необходимость указать авторство фотографии, наряду с файлом снимка должен быть подготовлен файл метаданных.

Если имя файла снимка, допустим, 12345_1.jpg, то если необходимо назначить авторство этой фотографии, в том же каталоге должен быть сохранен текстовый файл с расширением .meta, т.е. 12345_1.meta должен быть, например, следующего содержания:

```
{«photographer»: «Dmitry E. Kislov»}
```

Файл должен содержать только одну строку. Кодировка при сохранении файла должна быть utf-8.

Аналогично со снимками места сбора, если имя файла снимка 123456_1p.jpg, то имя файла метаданных будет: 123456_1p.meta.

Пример meta-файла можно загрузить по ссылке.

3.4 Примеры размещения файлов гербарных образцов

- размещение без вложенной структуры каталогов (в куче): VBGI/ID/132432_1.jpg, VBGI/ID/132432_1.jpg, ...; VBGI/CODE/13212_1.jpg, VBGI/CODE/13211_1.jpg, ...

- размещение с пользовательской структурой каталогов: VBGI/ID/17.02.2017/63723_1.jpg, VBGI/CODE/18.02.2017/65123_1.jpg, VBGI/CODE/18.02.2017/65123_1p.jpg, ...

3.4.1 Доступ к файлам

Обработанные файлы изображений будут доступны для чтения по протоколу HTTP по адресам:

- http://BASE_URL/ss/ACRONYM/{ID, CODE}/dddddd_dd<p>.ext
- http://BASE_URL/ms/ACRONYM/{ID, CODE}/dddddd_dd<p>.ext
- http://BASE_URL/fs/ACRONYM/{ID, CODE}/dddddd_dd<p>.ext

где BASE_URL – адрес сервера, через который будут доступны изображения (возможно botsad.ru, возможно, какой-нибудь herbstatic.botsad.ru); далее, символы ss, ms, fs – указывают на пути к изображениям различного разрешения: ss (small size) – максимальная ширина или высота 100 px, ms (medium size) – максимальная ширина или высота изображения 2000 px; fs (full size) – максимальная ширина или высота ограничивается возможностью сканирующего устройства.

Пользовательская структура каталогов (SET/OF/NESTED/FOLDERS) при размещении на сервере не сохраняется.

Изображения размера full size – не будут доступны для просмотра на странице отображения гербарных образцов, но их можно будет скачать по прямой ссылке.

3.4.2 Калибровочные ячейки

Крайне рекомендуется на сканах гербарных образцов представлять калибровочные квадраты, и желательно, разных цветов. Калибровочные квадраты используются для автоматизированного пересчета соответствия между «количеством пикселей» и принятыми единицами длины.

Калибровочные квадраты должны быть строго одинаковыми для всех изображений, 1 см x 1 см.

Установленный формат калибровочных ячеек (квадратов) можно найти по [ссылке](#)

Выполненные в ярких разных цветах калибровочные квадраты позволят потенциальным пользователям электронного гербария идентифицировать где находится калибровочный квадрат, а где элемент растительности в автоматизированном режиме при помощи соответствующих программных решений.

Пример гербарного изображения с калибровочными квадратами можно посмотреть ниже.

