

GOOGLE SPREADSHEETS КАК БАЗОВЫЙ ИНСТРУМЕНТ ДЛЯ УПРАВЛЕНИЯ ДАННЫМИ О БИОРАЗНООБРАЗИИ

**С.Ю. Большаков¹, Н.В. Филиппова², К.О. Потапов³, Д.В. Агеев⁴,
С.В. Волобуев¹**

GOOGLE SPREADSHEETS AS A BASIC TOOL FOR THE MANAGEMENT OF BIODIVERSITY DATA

**S.YU. Bolshakov¹, N.V. Filippova², K.O. Potapov³, D.V. Ageev⁴,
S.V. Volobuev¹**

¹*Ботанический институт им. В.Л. Комарова РАН, г. Санкт-Петербург; e-mail: sbolshakov@binran.ru*

²*Югорский государственный университет, г. Ханты-Мансийск; e-mail: filippova.courlee.nina@gmail.com*

³*Казанский (Приволжский) федеральный университет, г. Казань; e-mail: potapov_ko@mail.ru*

⁴*Грибы Новосибирской области, <http://mycology.su/>*

В настоящее время существует свободное программное обеспечение для управления данными о биоразнообразии: Biota, BRAHMS, Symbiota, Specify и др. Они позволяют развернуть собственные полноценные web-порталы, работающие на современных СУБД, таких как MySQL, SQLite и др., и включающие в себя управление коллекциями, создание чек-листов, идентификационных ключей, карт распространения видов. Однако их развёртывание и поддержка требуют наличия соответствующих знаний (или специалистов) и собственных серверов.

Альтернативным вариантом является поглощение данных уже существующими крупными порталами, таких как BioCASE, JACQ и др.; в России существуют лишь несколько удачных примеров – депозитарий «Ноев ковчег» (<https://plant.depo.msu.ru/>), ИС CRIS (<http://kpabg.ru/cris/>) и др. Преимущество такого подхода – в отсутствии необходимости в собственных технических затратах.

При этом для небольших проектов или гербариев практически повсеместно специалисты вынуждены использовать ПО компании Microsoft – Excel или Access, зачастую десктоп-версии, несмотря на наличие бесплатной онлайн-версии Excel Online.

В качестве достойной альтернативы такого базового и легкодоступного инструмента для создания и ведения простейших БД и каталогов мы предлагаем использовать Google Spreadsheets.

К преимуществам этого инструмента относятся:

– наличие SQL-подобного языка запросов в функционале формул (позволяет превратить таблицу в некоторое подобие БД с плоской струк-

турой и упрощённым использованием справочных таблиц) со всеми возможными запросами к массиву данных;

– использование JavaScript-подобного языка программирования Google Apps Script для расширения функционала (позволяет написать скрипты, которые будут выполнять некоторые функции, отсутствующие в стандартном наборе).

Коллективом специалистов-микологов из разных учреждений России была начата успешная работа по объединению данных научных публикаций по видовому составу грибов и лишайников России с использованием функционала Google Spreadsheets.

Структура таких БД представлена следующими листами:

– основная таблица Data с основными полями (в терминах DarwinCore): originalNameUsage – scientificName – bibliographicCitation – geography – stateProvince – federalProtectedArea – notes – genus – group;

– таблица с библиографическими записями (на основе полей End-Note);

– справочная таблица Species со списком названий видов и соответствующими каждому из них текущими принимаемыми названиями;

– справочные таблицы с информацией о регионах и ООПТ;

– таблица Account для запросов и анализа информации.

Наличие единой справочной таблицы Species (которая используется во всех наших подобных проектах) позволяет осуществлять в таблице Data автоподстановку значений scientificName на основе значений originalNameUsage, вносимых с помощью выпадающих списков проверки данных.

Синтакс такой формулы для автоподстановки во всех ячейках в scientificName (поле B) таблицы с данными: =query(Species!A:B;"select A where B='&A2&'"). Здесь "&A2&" – изменяемый адрес для каждой ячейки в таблице Data. Эта формула протягивается в столбце B для всех заполненных ячеек столбца A.

Следующая формула в таблице с данными позволяет извлечь название рода из названия вида (поле B): =query(split(B2;" ");"select Col1").

В таблице Account для получения уникального списка регионов (поле G), уже внесённых в таблицу Data, используется формула следующего вида: = sort (unique (query (Data!A2:M;"select G"))). Для подсчёта числа видов (поле B), которые учтены для того или иного региона (поле G) используется формула вида =countunique(query(Data!A\$2:M;"select B where G='&A2&'")). Здесь "&A2&" – изменяемый адрес для каждой ячейки в таблице Account. Эта формула протягивается в столбце B для

всех заполненных ячеек столбца А. Точно такой же синтаксис для любого другого списка значений, которое требуется подсчитать – ООПТ, роды, публикации и т.п.

Общее количество уникальных видов (или любых других значений) подсчитывается формулой вида = countunique (Data!B2:B).

Функционал Google Spreadsheets с вышеописанными структурой и формулами позволил реализовать два успешных проекта:

– базу данных по афиллофороидным грибам Европейской части России. Эта база данных объединяет данные 495 публикаций о видовом разнообразии грибов этой группы для каждого региона. Общее число выявленных видов – 1106;

– базу данных по грибам Ханты-Мансийского автономного округа – Югры (Fungal records database of Yugra – <https://fungariumysu.org/fredy>). Эта база объединяет данные около 70 публикаций о видовом разнообразии грибов всех групп, упоминаемых в этих работах, на территории округа. Общее количество выявленных видов достигло 3 тысяч.

Для совместного доступа и управления данными микологической коллекции БИН РАН (LE), так и региональных фунгариев, мы также применяем Google Spreadsheets – одновременный доступ к единой базе ряда специалистов позволяет использовать, изменять и анализировать единый массив данных, построенный по единому шаблону.

Функционал практически всех продуктов Google, включая Spreadsheets, может быть значительно расширен благодаря языку Google Apps Script. Использование кода на этом языке в скрипте, работающего с шаблоном в Google Docs, предлагается нами для создания этикеток.

Несмотря на удобство, простоту использования и доступность для всех желающих, мы рассматриваем Google Spreadsheets как инструмент для временных проектов по одновременному сбору данных от различных удалённых друг от друга специалистов. В дальнейшем данные из таких БД можно и нужно внедрять в более крупные порталы.