

«ФЛОРА РОССИИ» НА INATURALIST: ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПОЛЕВОЙ БОТАНИКЕ

Серегин А.П.

МГУ имени М.В. Ломоносова, биологический факультет, кафедра экологии и географии растений, Гербарий, Ленинские Горы, д. 1, 119991, Москва

Стремительный рост данных о биоразнообразии – новая реальность, с которой мы столкнулись в последние годы. Четверть данных об отдельных местонахождениях живых организмов была собрана за последние три года, а половина – за семь лет. На этой основе стала складываться новая научная дисциплина «информатика биоразнообразия». Основа этого лавинообразного роста – доступность и рост производительности электронных устройств (особенно смартфонов) и высокоскоростной интернет. Это явление получило массовое распространение и стало неожиданным драйвером существовавшего и прежде научного волонтерства (citizen science).

Научная проблема, которой занимается наша группа, – сбор и анализ больших данных о разнообразии сосудистых растений России. Ранее мы организовали на международной платформе iNaturalist проект «Флора России» (<https://www.inaturalist.org/projects/flora-of-russia>) для сбора данных о разнообразии сосудистых растений страны в форме привязанных к карте фотографий, определенных и верифицированных экспертами. Проект стартовал 9 января 2019 г., а спустя 22 месяца при участии 10,8 тыс. волонтеров-наблюдателей и 4 тыс. экспертов мы собрали 848 тыс. фотонаблюдений о 7 тыс. видах. Сейчас iNaturalist обеспечивает четверть электронных данных по флоре России.

Сейчас мы создали самый большой массив данных о биоразнообразии России, это единственный источник данных о современном состоянии разнообразия флоры страны. Полученный массив – совершенно новое явление, фундаментальные характеристики которого нам ясны лишь в общих чертах. Растущее сообщество участников проекта в сентябре 2020 г. преодолело отметку в 10 тыс. участников, и сейчас это самый масштабный проект в области научного волонтерства в России. При реализации проекта происходит внедрение в классическую науку новейших технологий (методы сбора, машинное обучение, анализ больших данных). Методы проекта хорошо применимы для сосудистых растений, как доступного объекта наблюдений. Важно и то, что при привлечении волонтеров происходит перераспределение затрат, в т.ч. за счёт трудового вклада сообщества. Рост данных проекта приводит к бутылочному горлышку экспертизы, также имеется проблема неравномерности данных. Диспропорции в данных проекта самые разные – пространственные, таксономические, сезонные, временные, социально-экономические. Диспропорциональность выборки требует разработки методов и технологий выравнивания данных, их калибровки и дальнейшего регулярного пересмотра моделей. Темпы накопления данных показывают, что анализ, который будет проводиться через год, будет основан на сведениях, которые пока не собраны. Мы впервые получаем возможность оценить современные изменения во флоре страны. Проект дает уникальную возможность прямого общения натуралистов со всей страны с экспертами из ведущих научных центров.