Biodiversity Informatics

Максим Шашков

Екатеринбург 2020

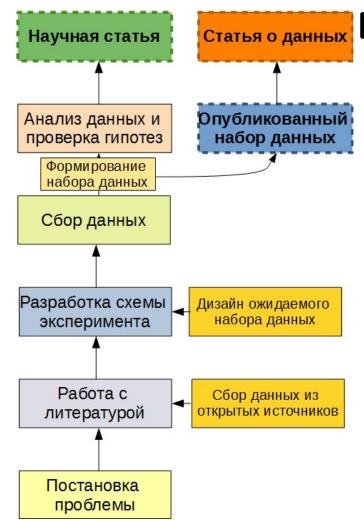
Научная публикация - исследование

Методология научного исследования

описательный метод - система процедур сбора, первичного анализа и изложения данных и их характеристик

экспериментальный метод - исследование объектов или явлений в заданных условиях





_{Статья о данных} Научная публикация - data paper

Data Paper

PhytoKeys 77: 71-80

https://doi.org/10.3897/phytokeys.77.11186 (09 Mar 2017)

Distribution of the invasive plant species *Heracleum* sosnowskyi Manden. in the Komi Republic (Russia)

▼ Ivan Chadin, Igor Dalke, Ilya Zakhozhiy, Ruslan Malyshev, Elena Madi, Olga Kuzivanova, Dmitrii Kirillov, Vladimir Elsakov

Abstract A

Occurrences of the invasive plant species *Heracleum sosnowskyi* Manden. in the Komi Republic (northeastern part of European Russia) were recorded and published in the Global Biodiversity Information Facility (GBIF http://www.gbif.org) using the RIVR information system (http://ib.komisc.ru/add/rivr/en). RIVR stands for "Rasprostranenie Invasionnyh Vidov Rastenij" [Occurrence of Invasion Plant Species]. This citizen science project aims at collecting occurrence data about invasive plant species with the help of citizen scientists. Information can be added by any user after a simple registration (concept) process. However, the data published in GBIF are provided only by professional scientists. The total study area is approximately 19,000 km². The GBIF resource contains 10894 *H. sosnowskyi* occurrence points, each with their geographical coordinates and photographs of the plants in the locus of growth. The preliminary results of species distribution modelling on the territory of European North-East Russia presented.

Keywords A

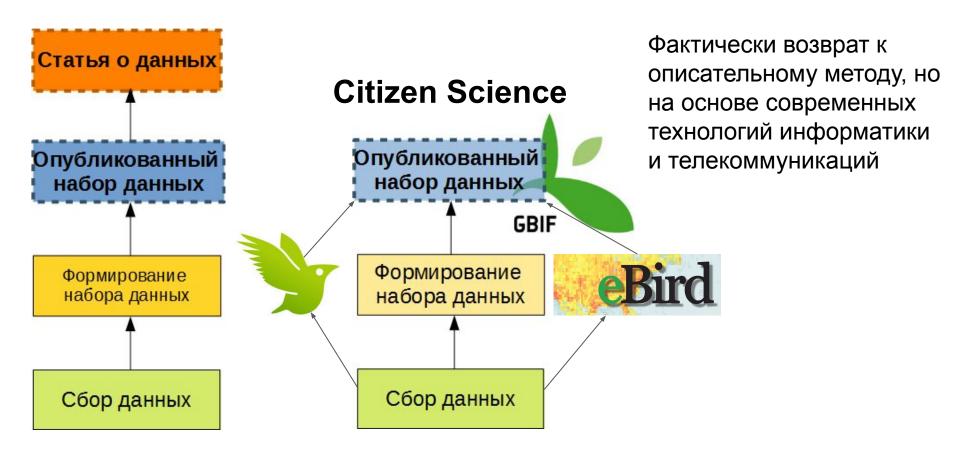
Occurrence, human observation, *Heracleum sosnowskyi*, hogweed, invasive, geotagged photographs, Komi Republic, European North-East Russia

Project details .

Project title

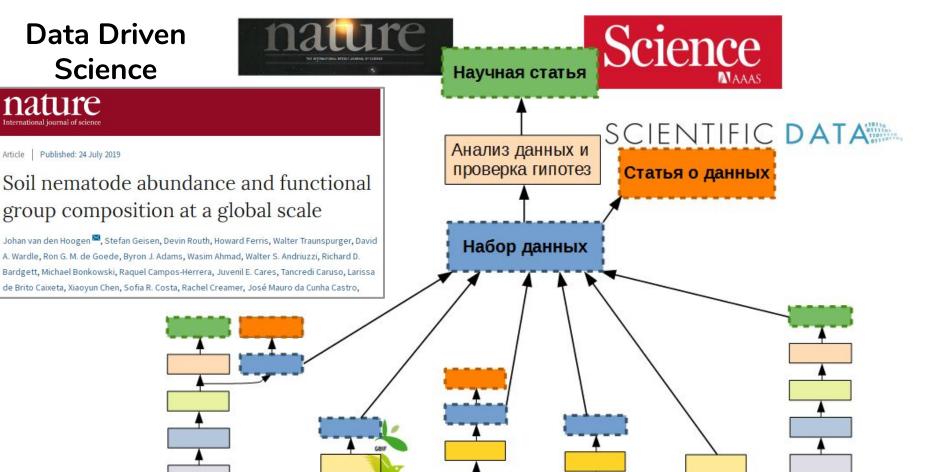
"Ecophysiological modelling of invasive plant species distribution. The case of *Heracleum sosnowskyi* in the taiga zone of the European part of Russia"

Набор данных - новый вид научного продукта



Data Driven Science

Published: 24 July 2019



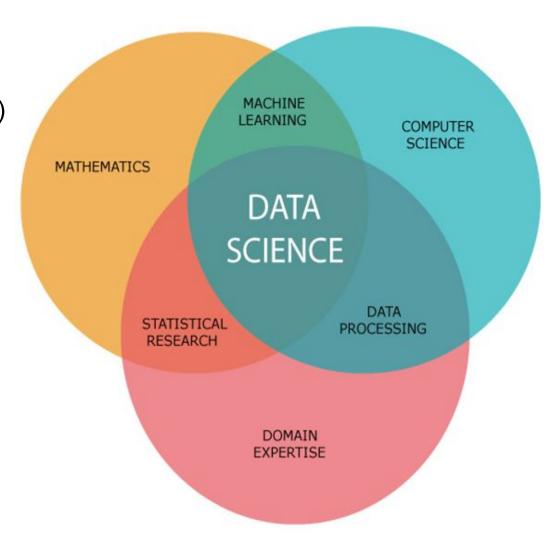
Data Science

Наука о данных (Data Science)

сплав математической статистики, информатики и предметной области.

Междисциплинарная область, направленная на извлечение знаний из наборов данных, обычно большого объема (Big Data)

Machine Learning



Задачи Data Science

Сбор данных (Data Mining)

Обработка данных (Data Processing)

Визуализация данных (Data visualization)

Верификация и корректировка (Data Cleaning)

Управление данными (Data Management)

Анализ данных (Data analysis)

Моделирование (Modelling)

Средства Data Science

Средства систематизации и хранения данных: СУБД и Big Data

Языки программирования - R и Python

Средства визуализации данных.

Средства анализа данных

Рабочие среды (Frame Works)

R - язык статистического программирования

с 1993 года, фактически общепринятая среда для статистических вычислений. Все новые методы в первую очередь появляются в R

мощные средства построения графиков, обеспечивающие качество типографского уровня.

Python - скриптовый язык высокого уровня

с 1990 года, в настоящее время написаны тысячи открытых библиотек.

Requests - запрос и получение данных через HTTP. Например через API iNaturalist или GBIF

Matplotlib - визуализация данных в виде разнообразных графиков

SciPy - научные и инженерные расчеты, статистический анализ

NumPy - работа с многомерными массивами

Pandas - надстройка для NumPy для удобства работы

Pillow - обработка изображений

TensorFlow - машинное обучение, работа с нейросетями

Exiftool - работа с метаданным изображений



Biodiversity Informatics

Часть пространства Data Science, связанная с живой природой, на популяционном и экосистемном уровне

TDWG

Biodiversity Information Standards

Virtual Conference 2020



Biodiversity Data Quality



Biodiversity Informatics Curriculum



Biodiversity Services



Biological Interactions Data



Citizen Science



Collection Descriptions



Darwin Core



Earth Sciences and Paleobiology



Genomic Biodiversity



Machine Observations



Observations & specimens

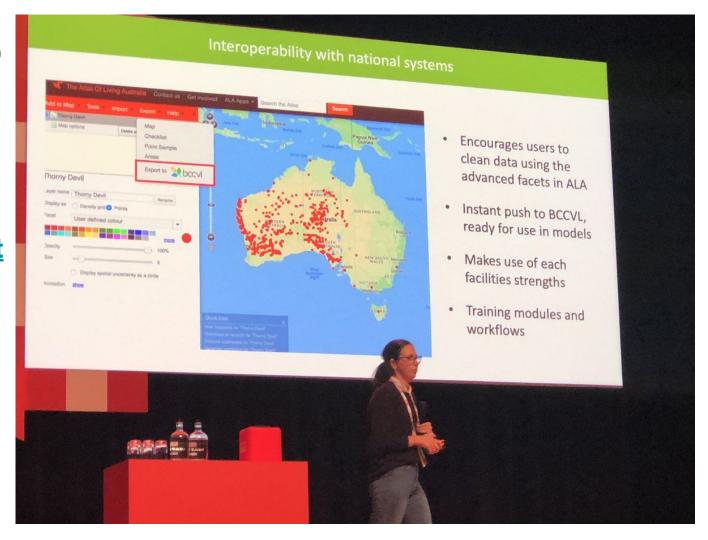


Process

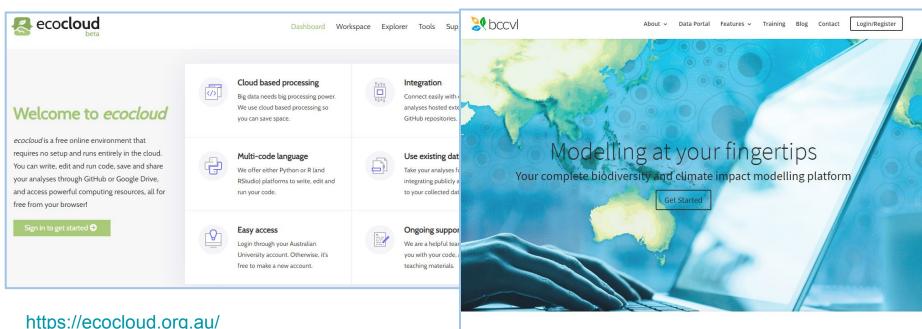
TDWG 2019

Leiden, Netherlands

Biodiversity_next



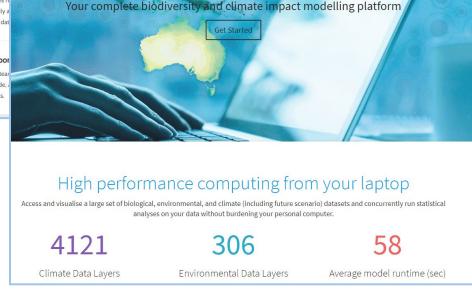
Системы облачных вычислений **ecocloud** и **bccvl**



http://bccvl.org.au/

Hallgren et al., 2016, The Biodiversity and Climate Change Virtual Laboratory: Where ecology meets big data.

DOI: 10.1016/j.envsoft.2015.10.025



Species Distribution Modelling (SDM)

