INATURALIST VS. PL@NTNET: ACCELERATING DATA COLLECTING ON ALIEN PLANTS OF RUSSIA IN REAL-TIME MODE

A. P. Seregin
Lomonosov Moscow State University, 1 Leninskie Gory, 119991 Moscow, Russia
(*botanik.seregin@gmail.com*)

Modern technologies make it possible to collect information on the distribution of living organisms with unprecedented speed and efficiency. Thus, 60 % of records in Global Biodiversity Information Facility (https://www.gbif.org/) on dated records of living organisms were made in the last ten years (2011–2021) during field research. In most cases, researchers are now collecting data in electronic form directly in the field using various gadgets. This information is rapidly entering the global data circulation, using GBIF platform as a rule.

A special role in the collection of big data on biodiversity is assigned to non-professional researchers who generate electronic datasets on records of various plants, animals and fungi using simple public platforms like ebird, iNaturalist, Pl@ntNet, Naturgucker and many others. Over the past two years, both iNaturalist website and app have gained unprecedented popularity in Russia, largely due to the emergence of our project "Flora of Russia" in early 2019 (https://www.inaturalist.org/projects/flora-of-russia). Currently, it has over 910,000 observations based on photos from 11,000 users.

Large datasets from non-professional researchers are of particular value when compared with other available e-data on plants of Russia, i.e. published sources and herbarium collections. At the moment, data from iNaturalist and Pl@ntNet form the main source of information (as of January 27, 2021) on alien species which rapidly expanding their ranges.

Number of GBIF-mediated records for ten most widely distributed invasive species of Russia for all years are given below.

1) *Heracleum sosnowskyi* – GBIF: 14267; iNaturalist: 3027; Pl@ntNet: n/a; both: 21.2 %.
2) *Acer negundo* – GBIF: 6989; iNaturalist: 5032; Pl@ntNet: 1200; both: 89.2 %.
3) *Erigeron canadensis* – GBIF: 3732; iNaturalist: 2445; Pl@ntNet: n/a; both: 65.5 %.
4) *Impatiens glandulifera* – GBIF: 3694; iNaturalist: 1785; Pl@ntNet: 1654; both: 93.1 %.
5) *Matricaria discoidea* – GBIF: 3328; iNaturalist: 1594; Pl@ntNet: 184; both: 53.4 %.
6) *Impatiens parviflora* – GBIF: 3053; iNaturalist: 1944; Pl@ntNet: 719; both: 87.2 %.
7) *Lupinus polyphyllus* – GBIF: 2896; iNaturalist: 1882; Pl@ntNet: 475; both: 81.4 %.
8) *Bidens frondosa* – GBIF: 1868; iNaturalist: 1289; Pl@ntNet: 2; both: 69.1 %.
9) *Erigeron annuus* – GBIF: 1828; iNaturalist: 1454; Pl@ntNet: 9; both: 80.0 %.
10) *Solidago canadensis* – GBIF: 1699; iNaturalist: 1399; Pl@ntNet: 26; both: 83.9 %.

The same figures restricted to the data collected in 2019–2021 are as follows.

1) *Heracleum sosnowskyi* – GBIF: 2748; iNaturalist: 2744; Pl@ntNet: n/a; both: 99.9 %.
2) *Acer negundo* – GBIF: 5697; iNaturalist: 4572; Pl@ntNet: 1080; both: 99. 2 %.
3) *Erigeron canadensis* – GBIF: 2331; iNaturalist: 2302; Pl@ntNet: n/a; both: 98.8 %.
4) *Impatiens glandulifera* – GBIF: 3165; iNaturalist: 1551; Pl@ntNet: 1562; both: 98.4 %.
5) *Matricaria discoidea* – GBIF: 1724; iNaturalist: 1522; Pl@ntNet: 179; both: 98.7 %.
6) *Impatiens parviflora* – GBIF: 2557; iNaturalist: 1811; Pl@ntNet: 682; both: 97.5 %.
7) *Lupinus polyphyllus* – GBIF: 2082; iNaturalist: 1616; Pl@ntNet: 426; both: 98.1 %.
8) *Bidens frondosa* – GBIF: 1186; iNaturalist: 1164; Pl@ntNet: 2; both: 98.3 %.
9) *Erigeron annuus* – GBIF: 1409; iNaturalist: 1372; Pl@ntNet: 9; both: 98.0 %.
10) *Solidago canadensis* – GBIF: 1318; iNaturalist: 1284; Pl@ntNet: 26; both: 99.4 %.

The proportion of these two crowdsourcing platforms for the ten most common invasive species of the Russian flora ranges from 21 to 89 %. At the same time, the latest data on localities of these species (collected in 2019–2021) are by 97.5–99.9 % supported by records from iNaturalist and Pl@ntNet.

Pl@ntNet is the main competitor to iNaturalist among automatic plant recognition mobile apps and as a GBIF publisher of citizen science data on plants. Below we discuss the main differences between the two platforms.

1) In Pl@ntNet (France), you can upload photos of vascular plants only, whereas iNaturalist (USA) enables to upload all living organisms. Four photos per single observation is the limit of Pl@ntNet, and iNaturalist has twenty photo limit.

2) In Pl@ntNet, one can create projects, but only by contacting the head office. In iNaturalist, creation and administration of projects is available to any user through a web interface. The strength of iNaturalist is its huge community of experts who are checking identifications.

3) Pl@ntNet does not have a common entry point to a single base, i.e. the "Explore" button is available on project pages only. iNaturalist has a common access point for all observations (https://www.inaturalist.org/observations).

4) In Pl@ntNet, any data on geography is hidden on the observation page, but available through GBIF, if the observation is there. To what extent this location is coarse remains unclear. In iNaturalist, all geo data are available, except when they are hidden on purpose by the user in the settings (or coarsened by iNaturalist for vulnerable species).

5) In general, observation on Pl@ntNet contains very little information: photo, taxon name, function of anonymous identification, rough assessment of data quality (for/against), author, and date. Also, there is a tab with a discussion of the records. The functionality and details of the observations available on iNaturalist are much wider.

6) The web version of Pl@ntNet is available in three languages, with no plant names in Russian. iNaturalist has a web interface available in dozens of languages and has very detailed databases of national names. In the apps, the choice of languages for both platforms is approximately equal.

7) To identify something on Pl@ntNet, one needs to select a species and slide through the photos. The correct name can be inserted on observation page only. There are no geographic filters in searching the records. iNaturalist has a special interface for identification with a detailed filter system.

8) Cultivated and wild plants in Pl@ntNet are mixed, and it is impossible to separate one from the other. On the contrary, a clear distinction between cultural and natural species is one of the fundamental requirements of iNaturalist.

9) Pl@ntNet data transferred to GBIF as two datasets: https://doi.org/10.15468/gtebaa (794K observations with photos) and https://doi.org/10.15468/mma2ec (9.6M automatically identified snapshots without photos). iNaturalist has a single dataset in GBIF available at https://doi.org/10.15468/ab3s5x (23.4M records with photos, incl. 8.7M records of vascular plants).

10) In Pl@ntNet, the probability of automatic suggestions (as a percentage) is available for the user. If it's high, no other options are offered. On iNaturalist, for instance, the following disclaimer is available: "This is most likely the genus *Adoxa*" with *Adoxa moschatellina* as the first option, and in addition seven other unnecessary species. This point is definitely better worked out on Pl@ntNet. However, the suggestion of a genus, family or tribe on iNaturalist is a very strong feature useful for identification of species unknown to AI. Full taxonomic names on Pl@ntNet contain taxonomic authors, whereas iNaturalist does not provide this data.

Additionally, Pl@ntNet has no project with journals and discussions, logs (dashboards), forum, user's statistics, csv uploads, shapefile uploads, calendar, profile, notification and dashboard, opportunity for bulk uploads, bioblitzes. All this is on iNaturalist.

Citizen science platforms like iNaturalist, Pl@ntNet and others as well as GBIF with freely available data rapidly change the way we study plants, their geography, phenology and ecology. For many alien species – a case in point is Russia – they are merely the only source of reliable modern data.