

PREPRINT

Author-formatted, not peer-reviewed document posted on 25/10/2021

DOI: <https://doi.org/10.3897/arphapreprints.e77013>

Geohistorical dataset of ten plant species introduced into Occitania (France)

Morgane Claudel,  Emilie Lerigoleur, Cécile Brun, Sylvie Guillerme

Geohistorical dataset of ten plant species introduced into Occitania (France)

Morgane Claudel^{‡,§}, Emilie Lerigoleur^{‡,§}, Cécile Brun^{‡,§}, Sylvie Guillerme^{‡,§}

[‡] UMR 5602 GEODE CNRS-UT2J, Toulouse, France

[§] LTSER Zone Atelier Pyrénées-Garonne, CNRS, Université de Toulouse, Castanet Tolosan, France

| Université de Nantes, Nantes, France

Corresponding author: Morgane Claudel (claudelmorgane@orange.fr), Cécile Brun (cecile.brun@univ-tlse2.fr), Sylvie Guillerme (sylvie.guillerme@univ-tlse2.fr)

Abstract

Background

The original dataset presented here is the result of the first near-exhaustive analysis performed on historical data concerning ten plant species introduced in and around Occitania (south-western France) since 1651. Research was carried out on the following species: *Alnus incana*, *Buddleja davidii*, *Castanea sativa*, *Helianthus tuberosus*, *Impatiens glandulifera*, *Prunus cerasifera*, *Prunus laurocerasus*, *Reynoutria japonica*, *Robinia pseudoacacia* and *Spiraea japonica*.

The data file contains 199 occurrence data exclusively based on historical observations and records made between 1651 and 2004 and retrieved from 111 of the 640 literary sources consulted. All the records are associated with a year and 61% of them have associated spatial coordinates. Initially the EI2P-VALEEBEE research project focused on the introduction of these species into Occitania (95 occurrences, 47.7%), but mentions found beyond this territory - mainly in metropolitan France - are also reported.

The creation of this dataset involved five stages: (1) selection of species, (2) consultation of historical sources, (3) recording of occurrences in the dataset, (4) dataset standardization/enrichment and Darwin core mapping, (5) data publication. Quality controls were conducted at each step.

The dataset is available on the platform of the Global Biodiversity Information Facility (GBIF) at <https://doi.org/10.15468/3kvaeh>. It respects the internationally recognized FAIR Data Principles (Findable, Accessible, Interoperable and Reusable).

New information

The dataset will be progressively enriched by new data during the EI2P-VALEEBEE research project and future projects on invasive plant species conducted by the team.

Introduction

EI2P-VALEEBEE is a project co-funded by the Région Occitanie and the Maison des Sciences de l'Homme de Toulouse (cf. glossary of acronyms in Suppl. material 1). It combines two issues of major ecological, socio-cultural and economic concern: biological invasions and the decline of pollinator populations. The main objective of this project is to better identify the links between plant invasions and changes in ecosystem services (Vaz et al. 2017a) so as to better understand the potentials or constraints of invasive alien plants in connection with pollinators and try to apprehend invasive processes through the most systemic approach, by taking several dimensions into account and by taking strong interest in the practices, perceptions, and representations that the various stakeholders in the territory have manifested over time (Vaz et al. 2017b). This systemic approach is also found in the project's method itself, since it brings together several disciplines and tools, and also has a diachronic dimension over several time scales. In carrying out the project, it was decided to focus on the Occitania region in the south-west part of France and, more particularly, on two territories: the Pique valley and the Oussouet valley, both of which have not only a high level of plant diversity but also a high rate of invasion.

General description

Purpose: The data file is the result of a geo-historical study conducted on the introduction and distribution of invasive plant species. Ten plant species were selected that can all be observed in the two valleys. Some of them are considered as invasive, alien species. The study includes research on the introduction dates of the species studied, the locations of their introduction, their interest and past uses, the different human perceptions of them over time, activities that have impacted their local distribution, comments from authors and observers on their abundance, and elements of the historical context of introduction. Historical sources were consulted during 2020 in order to find the oldest elements concerning the ten species.

Additional information: Interest and use of the dataset

Without a historical analysis, it is difficult to understand the current local distribution dynamics of invasive plant species, especially since some of them were introduced into Metropolitan French territory several centuries ago. All the interest of this dataset is to provide historical depth and chronological elements for the understanding of the current

distribution of these ten species at local scale. In this perspective, the dataset is relevant to:

- identify the different periods of introduction of the ten species studied;
- have accurate data on the main introduction channels, such as the place where they were introduced (mainly ornamental gardens, thermal parks, and private gardens but also much wilder and more natural areas, such as road borders and forests, for some of them);
- identify pathways of colonization to understand the current localization of each species;
- better understand why they were introduced, particularly thanks to information on their uses;
- provide elements of analysis to understand the temporal phases of species distribution.

The interest of the dataset is directly in line with the values of the E12P-VALEEBEE project itself, whose objective is to contribute to better knowledge of plant invasion processes in a transversal way. In this perspective, it can be useful for all local, national and international organisations involved in this issue with current ecological, economic and social implications.

Project description

Title: Geohistorical dataset of ten plant species introduced in Occitania (France)

Personnel: Conceptualization, M.C. and E.L.; methodology, M.C., E.L., C.B. and S.G.; investigation, M.C.; data validation, M.C. and E.L.; writing, review and editing, M.C., E.L., C.B. and S.G.; supervision, C.B. and S.G.; project administration, S.G.; funding acquisition, S.G.

Study area description: The Oussouet valley (Pyrenean foothills, Hautes-Pyrénées) and the Pique valley (Haute-Garonne) in Occitania (South of France).

Funding:

- Région Occitanie - Appel à projets Recherche et Société(s) 2019
- Maison des Sciences de l'Homme et de la Société de Toulouse (MSH-T) APEX 2020

Sampling methods

Sampling description: The creation of this dataset involved a number of different stages: (1) selection of species, (2) consultation of historical sources, (3) recording of occurrences on the dataset, (4) dataset standardization/enrichment and Darwin core mapping, (5) data publication.

Step 1: Species selection

Current field observations were made particularly in the two valleys selected in the south-west of France: the Pique valley and the Oussouet valley. These two territories present a high rate of plant invasions. Four exotic plant species were initially observed to provide some spatial coverage in these valleys: *Buddleja davidii*, *Impatiens glandulifera*, *Reynoutria japonica* and *Spiraea japonica*. According to the scientific needs of the EI2P-VALEEBEE project, six species (*Alnus incana*, *Castanea sativa*, *Helianthus tuberosus*, *Prunus cerasifera*, *Prunus laurocerasus* and *Robinia pseudoacacia*) were added to the selection on the basis of the main relevant criteria (see glossary for acronyms in Suppl. material 1):

- a lack of reference data in the existing literature at the national and/or local level (CBNPMP, GBIF, IUCN, Catalogue of Life, INPN, Tela Botanica BDTFX, Baseflor DB- Ph. Julve, Invasive Species Compendium CABI ISC, Delivering alien invasive species in Europe, DAISIE);
- the distribution of the species in Occitania (Human observations, GBIF);
- their current status (Muller 2004, CBNPMP, GBIF, IUCN);
- the most frequently occurring species in the south-west of France (Planty-Tabacchi 1997);
- the current knowledge on their honey and nectar potentials (FranceAgriMer, the ITSAP, the Ministry of Agriculture and Food, the SNHF, Astredhor and the VAL'HOR inter-professional association, in partnership with the INRA, the CNPAIM, the GNIS and the SBF).

Step 2: Consultation of historical sources

For the consultation of historical sources, a funnel method was applied:

1. To begin, the existing sources of naturalist data were inventoried at the national level, such as flora (e.g. Flore de France, Coste 1906); seed catalogues (e.g. Andrieux and Vilmorin 1783); horticultural, botanical and beekeeping newspapers and magazines (e.g. Bulletins de la Société Centrale d'Apiculture 1856-1946, Bulletins de la Société Nationale de Protection de la Nature 1882-1888). This first national-scale inventory gave an idea of the presence of the ten species in France during the different periods of history. It also allowed the written sources containing botanical elements on a smaller scale to be identified.
2. The second stage was the analysis of literary sources at the regional level (Occitania). Local herbariums (e.g. Bordere 1871-1872), regional flora (e.g. Philippe 1859), botanical magazines and newspapers of the region (e.g. Société d'Horticulture 1854-1921) were inventoried and analysed. The seed merchants and horticulturists located in the region were also inventoried to explore their seed catalogues (e.g. Catalogue des plantes vivaces et d'extérieur, Bonamy Frères 1874) and potentially find elements on the colonization routes of the species. In addition, efforts were made to be as exhaustive as possible by consulting archival institutions at the departmental level (Departmental Archives of Hautes-Pyrénées,

- Departmental Archives of Haute-Garonne). The archived documents consulted were of all kinds: some known botanists' collections (e.g. Bouget 1901-1948), invoices related to seed orders, documents on the construction of natural parks (Natural Park of the Pyrenees collection, Aymonin 1975) and thermal gardens ("Projet d'aménagement du Parc des Quinconces", Chevalier 1902), together with herbariums, letters and notes. Examining these references for specific information on introduced exotic species and their spatial location was innovative.
3. Once enough regional historical sources had been analysed, it was possible to focus on the scale of our study sites. The main idea was to study historical sources on the scale of the municipalities located in the Pique and the Oussouet valleys. For this purpose, municipal archives (e.g. Municipal archives of Bagnères-de-Luchon, Municipal archives of Tarbes) were also consulted, taking care to identify the potential pathways of introduction on a very local scale in advance, to facilitate the search in archived articles. As local data sources, archival records enabled consultation of communal monographs that presented a chapter on the local flora (e.g. Barèges monograph, Rondou 1907). Finally, newspapers associated with the municipalities (e.g. *Revue de Comminges*, Société des Études du Comminges 1885-2004) and local herbariums (e.g. Ramond de Carbonnières 1755-1825), offered information on the flora that was much more local and precise.

Throughout the research process, key informants having good knowledge of the study areas and their backgrounds contributed information and advice on valuable literary sources, allowing the research to best fit the study areas and to be as complete as possible. In order to consult sources of all kinds, it should also be noted that, as soon as we felt that a literary source could potentially contribute elements on one of the ten species, we consulted it even if, at first sight, it had no connection with botany (e.g. the recipe book of La Varenne 1651).

As a result, during this step, 640 literary sources were consulted. Among these, 111 (17.3%) provided information on the introduction and colonization of the ten species over time (Suppl. material 3). It must be understood that consulting historical sources was one of the most important steps, not only in the creation of this dataset but also for its future updates, because this work made it easier to identify potential sources of information and distinguish them from blind alleys.

Step 3: Occurrences recorded in the data file

Each time a species was mentioned in the historical literature consulted, it was recorded in an occurrence data file created in LibreOffice Calc (spreadsheet program). The file format is OpenDocument Spreadsheet (.ods). When recording an occurrence, attention was paid to the vernacular and scientific synonyms used in the historical literature. We recorded the occurrences for which the historical name is currently identified as a synonym in the Catalogue of Life, INPN, and ISSG but also according to the number of elements in the bibliographic source that allowed the taxon to be identified as such: photographic representation, image or plate of the species, precise description, mention of other known scientific and vernacular names. In the data file, a maximum of elements mentioned by the

author were recorded: the synonym cited; the reference code from the French taxonomic referential TAXREF (<https://inpn.mnhn.fr/programme/referentiel-taxonomique-taxref?lg=en>); the bibliographic reference in which the mention of the species was found; the date of observation of the species or, failing that, the date of its mention; the names of the observers and authors; the type of source; the description of the location as soon as it was mentioned; its spatial coverage and abundance; its minimum and maximum altitudes; any comments by the author about the species, the location of observation or mention of the species; the nearest town/village; and the latitude and longitude coordinates. In addition, each element concerning the literature source was recorded in the same software (LibreOffice Calc): bibliographic reference number (identifier), name(s) of the author(s), title, year of publication, collection and publisher if they were mentioned, as well as the city of publication, the name of the journal (if the source was an article), the URL if accessible, or the reference of the archive document with the location of the archive institution, the call number and the series.

Step 4: Dataset standardization/enrichment and Darwin Core mapping

The Darwin Core Standard (Wieczorek et al. 2012) "offers a stable, straightforward and flexible framework for compiling biodiversity data from varied and variable sources. (...) This standardization not only simplifies the process of publishing biodiversity datasets, it also makes it easy for users to discover, search, evaluate and compare datasets as they seek answers to today's data-intensive research and policy questions." (source: <https://www.gbif.org/darwin-core>).

Each column header of the occurrence spreadsheet was searched for an equivalent term in the Darwin Core quick reference guide (<https://dwc.tdwg.org/terms/>). We also chose the Identification History extension (<https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Identification>) to manage synonyms of taxon names as they were cited in the literature consulted.

The geographic data (longitude, latitude, WGS84 datum) were structured in two ways: 1) geographic coordinates with an accuracy of 500 metres for occurrences whose locality was precisely identified in the literary source, 2) geographic coordinates with an accuracy of 10,000 metres for occurrences whose literary source mentioned only the name of the municipality (<https://www.geonames.org/>). For each occurrence, the Darwin core terms "country", "province", "county", "municipality" and "locality" were assigned as far as possible from the elements of the literary sources.

Step 5: Data publication

For the publication in the GBIF platform, the Integrated Publishing Toolkit (IPT (gbif.org)) of the GBIF was used to fill out the metadata and to generate the Darwin Core Archive. The dataset is available on the GBIF platform at <https://doi.org/10.15468/3kvaeh> (Claudel et al. 2021). The dataset now respects the FAIR Data Principles (Findable, Accessible, Interoperable and Reusable) defined by Wilkinson et al. (Wilkinson 2016). Table 1 summarizes the FAIRness assessment criteria used to make the dataset FAIR.

Quality control: Several quality controls have been implemented. First, data cleaning and corrections were performed with proofreading by a third party and the use of pivot tables to check data integrity. Harmonization and standardization of content were also necessary to allow and facilitate the mapping with Darwin Core terms. The latter made it possible to identify new additional information such as nomenclatural code, coordinate uncertainty, geodetic datum, georeference source, license, etc. As many standards as possible were chosen to describe country codes (ISO 3166-1-alpha-2), municipality names and their geographic coordinates (geonames.org), taxon scientific names (TAXREF v13.0 - 2019-12-06), and taxon ID from several sources (GBIF, IUCN, Catalogue of Life, IPNI, INPN, Tela Botanica BDTFX). Finally, we used a GIS tool (QGIS 3.10 LTR) to check the geographic coordinates of occurrences.

Geographic coverage

Description: Within the framework of this geohistorical study, we consulted the information on the introductions and distributions of target species existing in archival sources regarding the Oussouet valley (Pyrenean foothills, Hautes-Pyrénées) and the Pique valley (Haute-Garonne), in Occitania (95 occurrences, 47.7%). When information concerning other French or European territories: other parts of France (50.8%), Belgium (1%) and the UK (0.5%), was found in the documents consulted, it was also recorded. This explains the European geographical coverage. Of the 199 occurrences, only 122 (61%) occurrences are precisely geolocated (municipality or 500 metre buffer). Fig. 1 shows the geolocation of these 122 occurrences and Fig. 2a and Fig. 2b focus on the occurrences found in the Oussouet valley and the Pique valley.

Coordinates: 42°5'52.8"N and 59°15'57.6"N Latitude; 8°36'46.8"W Longitude and 8°26'16.8"E Longitude.

Taxonomic coverage

Description: Ten plant species were studied: *Alnus incana* (L.) Moench, 1794; *Buddleja davidii* Franch., 1887; *Castanea sativa* Mill., 1798; *Helianthus tuberosus* L., 1753; *Impatiens glandulifera* Royle, 1833; *Prunus cerasifera* Ehrh., 1784; *Prunus laurocerasus* L., 1753; *Reynoutria japonica* Houtt., 1777; *Robinia pseudoacacia* L., 1753, and *Spiraea japonica* L.f., 1782. The scientific names of the ten species comply with the national taxonomic and nomenclatural reference system for fauna, flora and fungi in metropolitan France and overseas: TAXREF v13 (Gargominy et al. 2019). The database contains 199 records of these ten plant species. They are all angiosperms and belong to 7 orders, 8 families and 9 genera (Table 2), classified according to APG IV (2016).

For the purposes of the historical study, all the Latin synonyms identified and validated by the national taxonomic and nomenclatural reference frame TAXREF v13 related to the ten taxa studied were considered. During the analysis of historical documents, we also collected all vernacular synonyms as soon as they were associated with a validated Latin

synonym (Suppl. material 2); they will help to enrich the current vocabulary designating these species.

Temporal coverage

Data range: 1651-1-01 - 2004-1-01.

Notes: The geohistorical database includes observation and record data for the ten species from 1651 until 2004, collected in 2020. The objective was to cover the different periods of introduction of these ten species in Occitania. We consulted literature dating from the 17th century, a period during which some of the species studied seem to have been introduced in France (*Helianthus tuberosus*, *Prunus cerasifera*, *Prunus laurocerasus* and *Robinia pseudoacacia*).

The number of occurrences collected increased from around 1800 until 1950, which can be explained by the introduction of four of the exotic species into Metropolitan France in the 19th century: *Buddleja davidii*, *Impatiens glandulifera*, *Reynoutria japonica* and *Spiraea japonica*, and also by an increase in the number of historical sources relating to botany and horticulture, which facilitated the identification of mentions or observations of the species studied. Therefore, 75% of the historical data collected is from about 1800 to 1950 (Fig. 3). The lack of data for the 1950-1974 period could reflect the significant decrease in the number of naturalists' records during this post-war period.

Usage licence

Usage licence: Other

IP rights notes: [Creative Commons Attribution \(CC-BY\) 4.0 License](#)

Data resources

Data package title: dwca-ei2p_exotic_plants_geohistorical_occurrences_database-v1.4.zip

Resource link: <https://doi.org/10.15468/3kvaeh>

Number of data sets: 2

Data set name: Darwin Core Archive Geohistorical plants occurrences database (occurrence.txt)

Character set: UTF-8

Download URL: http://ipt.gbif.fr/archive.do?r=e_i2p_exotic_plants_geohistorical_occurrences_database&v=1.4

Data format: Darwin Core Archive format

Data format version: 1.4

Description: The Darwin Core Standard (DwC) was used to offer a "stable, straightforward and flexible framework for compiling biodiversity data from varied and variable sources" (<https://www.gbif.org/en/darwin-core>). All columns labels and descriptions are from <https://dwc.tdwg.org/terms/>.

Column label	Column description
id	Same as occurrenceID: An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique. This is the primary key of this table.
type	The nature or genre of the resource.
modified	The most recent date-time on which the resource was changed.
language	A language of the resource.
license	A legal document giving official permission to do something with the resource.
rightsHolder	A person or organization owning or managing rights over the resource.
institutionID	An identifier for the institution having custody of the object(s) or information referred to in the record.
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.
datasetName	The name identifying the data set from which the record was derived.
ownerInstitutionCode	The name (or acronym) in use by the institution having ownership of the object(s) or information referred to in the record.
basisOfRecord	The specific nature of the data record. Recommended best practice is to use the standard label of one of the Darwin Core classes.
occurrenceID	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.
occurrenceRemarks	Comments or notes about the Occurrence.
recordedBy	A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first.
occurrenceStatus	A statement about the presence or absence of a Taxon at a Location. Recommended best practice is to use this controlled vocabulary: http://rs.gbif.org/vocabulary/gbif/occurrence_status.xml

associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the Occurrence.
eventDate	The date-time or interval when the event was recorded. Not suitable for a time in a geological context. Recommended best practice is to use an encoding scheme, such as ISO 8601:2004(E).
year	The four-digit year in which the Event occurred, according to the Common Era Calendar.
eventRemarks	Comments or notes about the Event.
locationID	An identifier for the set of location information (data associated with dcterms:Location). May be a global unique identifier or an identifier specific to the data set.
continent	The name of the continent in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
countryCode	A unique (preferably globally-unique) identifier for the taxon represented in the row. Recommended best practice is to use ISO 3166-1-alpha-2 country codes: http://rs.gbif.org/vocabulary/iso/3166-1_alpha2.xml
stateProvince	The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
county	The full, unabbreviated name of the next smaller administrative region than stateProvince (county, shire, department, etc.) in which the Location occurs. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
municipality	The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs. Do not use this term for a nearby named place that does not contain the actual location. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
locality	The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description.
verbatimLocality	The original textual description of the place.
minimumElevationInMeters	The lower limit of the range of elevation (altitude, usually above sea level), in meters.
maximumElevationInMeters	The upper limit of the range of elevation (altitude, usually above sea level), in meters.
locationAccordingTo	Information about the source of this Location information. Could be a publication (gazetteer), institution, or team of individuals.

decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.
geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based. Recommended best practice is use the EPSG code as a controlled vocabulary to provide an SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value "unknown".
georeferenceSources	A list (concatenated and separated) of maps, gazetteers, or other resources used to georeference the Location, described specifically enough to allow anyone in the future to use the same resources.
georeferenceRemarks	Notes or comments about the spatial description determination, explaining assumptions made in addition or opposition to the those formalized in the method referred to in georeferenceProtocol.
identifiedBy	A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the subject. Recommended best practice is to separate the values in a list with space vertical bar space ().
dateIdentified	The date on which the subject was determined as representing the Taxon. Recommended best practice is to use a date that conforms to ISO 8601-1:2019.
taxonID	An identifier for the set of taxon information (data associated with the Taxon class). May be a global unique identifier or an identifier specific to the data set.
scientificNameID	An identifier for the nomenclatural (not taxonomic) details of a scientific name.
scientificName	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the identificationQualifier term.
nameAccordingTo	The reference to the source in which the specific taxon concept circumscription is defined or implied - traditionally signified by the Latin "sensu" or "sec." (from secundum, meaning "according to"). For taxa that result from identifications, a reference to the keys, monographs, experts and other sources should be given.
kingdom	The full scientific name of the kingdom in which the taxon is classified.
phylum	The full scientific name of the phylum or division in which the taxon is classified.

class	class
order	The full scientific name of the order in which the taxon is classified.
family	The full scientific name of the family in which the taxon is classified.
genus	The full scientific name of the genus in which the taxon is classified.
taxonRank	The taxonomic rank of the most specific name in the scientificName.
vernacularName	A common or vernacular name.
taxonRemarks	Comments or notes about the taxon or name.

Data set name: Darwin Core Archive Geohistorical plants occurrences database - Identification History Supplementary file (identification.txt)

Character set: UTF-8

Download URL: http://ipt.gbif.fr/archive.do?r=ei2p_exotic_plants_geohistorical_occurrences_database&v=1.4

Data format: Darwin Core Archive format

Data format version: 1.4

Description: The Darwin Core Identification History is an extension allowing multiple identification/determinations of species occurrences, particularly name spellings found in each original text. All identifications including the current one are listed, while the current should also be repeated in the occurrence core for simple access (Source: <http://tools.gbif.org/dwca-validator/extension.do?id=dwc:Identification>). All columns labels and descriptions are from <https://dwc.tdwg.org/terms/>.

Column label	Column description
id	An identifier for the Occurrence linked to the occurrence.txt file (same as occurrenceID). It can be repeated as a foreign key here.
identificationID	A unique identifier corresponding to the name spelling reported as found in the original text. This is the primary key of this table.
dateIdentified	The date on which the subject was determined as representing the Taxon. Recommended best practice is to use a date that conforms to ISO 8601-1:2019. The date format here is YYYY (eg. 1694).
scientificName	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the identificationQualifier term.

nameAccordingTo	The reference to the source in which the specific taxon concept circumscription is defined or implied - traditionally signified by the Latin "sensu" or "sec." (from secundum, meaning "according to"). For taxa that result from identifications, a reference to the keys, monographs, experts and other sources should be given.
vernacularName	A common or vernacular name.
taxonRemarks	Comments or notes about the taxon or name.

Additional information

Maintenance and future work

All data will be maintained by their creators. They will be progressively enriched by new data during the current EI2P-VALEEBEE research project, and also during the projects that the team will continue to conduct on invasive plant species thereafter.

The dataset is already archived and published through GBIF:

<https://www.gbif.org/dataset/345820cc-a0a8-4d76-b7eb-fba85b21ad08>

It will be regularly updated and versioned through GBIF.

Acknowledgements

Project title: EI2P - Espèces invasives et pollinisateurs, entre contraintes et potentiels | VALEEBEE - VALorisation des Espèces exotiques Envahissantes et Abeilles

Funding: Région Occitanie - Appel à projets Recherche et Société(s) 2019 | Maison des Sciences de l'Homme et de la Société de Toulouse (MSH-T) APEX 2020

This work is endorsed by the CNRS/INEE Zone Atelier Pyrénées Garonne (ZA PYGAR). The Zones Ateliers network (RZA) is recognized by ALLENI, as an ESFRI eLTER (European Long-Term Ecological Research) "Integrated European Long-Term Ecosystem, Critical Zone & Socio-Ecological System Research Infrastructure".

We appreciated the support and precious help of Sophie Pamerlon from GBIF-France who gave her time to enlighten us on Darwin Core mapping and the use of the Integrated Publishing Toolkit (<https://www.gbif.org/ipt>). We also warmly thank the technical evaluator of the data for his valuable advice.

References

- Andrieux P, Vilmorin P (1783) Catalogue des plantes, arbres, arbrisseaux et arbustes dont on trouve des graines, des bulbes et du plant chez les sieurs Andrieux et Vilmorin,

- Marchand Grainier-Fleuriste & Botaniste du Roi & Pépiniériste. Vilmorin-Andrieux, Paris. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k6209244b.texteImage>
- APG IV (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181 (1): 1-20. <https://doi.org/10.1111/boj.12385>
 - Aymonin G (1975) Itinéraire phyto-écologique des Landes du Marensin aux montagnes de Soule. Contribution à l'étude floristique et écologique du Parc National des Pyrénées. Fonds du Parc National des Pyrénées. 1863 W 61. Departmental Archives of the Hautes-Pyrénées, France. Accessed on August 19, 2020.
 - Bonamy Frères (1874) Catalogue général Bonamy Frères. Construction, plan et vues aquarellées du projet de jardin paysager et installation de la serre du Casino. 5 M 28. Municipal Archives of Bagnères-de-Luchon, France. Accessed on September 17, 2020.
 - Bordere H (1871-1872) Herbiers de Henri Bordere, membre de la Société Botanique de France. 1 J 592; 593; 594; 595. Departmental Archives of the Hautes-Pyrénées, France. Accessed on September 14, 2020.
 - Bouget J (1901-1948) Fonds du botaniste Joseph Bouget. 149 J 109; 9; 11; 176; 182; 235; 285. Departmental Archives of the Hautes-Pyrénées, Tarbes, France. Accessed on September 14, 2020.
 - Chevalier J (1902) Aménagement du parc des Quinconces. Lettre à destination du maire de Luchon, Paris, le 2 novembre 1902. 1 O 19. Municipal Archives of Bagnères-de-Luchon, France. Accessed on September 17, 2020.
 - Claudel M, Lerigoleur E, Brun C, Guillerme S (2021) Geohistorical plants occurrences database. Occurrence dataset accessed via GBIF.org on 2021-01-16. UMR 5602 Géographie de l'Environnement (CNRS/Université Toulouse 2). <https://doi.org/10.15468/3kvaeh>
 - Coste H (1906) Flore descriptive et illustrée de la France, de la Corse et des contrées limitrophes. 3. Librairie des Sciences Naturelles Paul Klincksiek, Paris.
 - Gargominy O, Terceire S, Régnier C, Ramage T, Dupont P, Daszkiewicz P, Poncet L (2019) TAXREF v13, référentiel taxonomique pour la France: méthodologie, mise en œuvre et diffusion. Muséum National d'Histoire Naturelle, Paris. Rapport Patrinat. 1–63. URL: <https://inpn.mnhn.fr/telechargement/referentielEspece/taxref/13.0/menu>
 - La Varenne (1651) Le cuisinier françois, enseignant la manière de bien apprester et assaisonner toutes sortes de viandes grasses et maigres, légumes, pâtisseries et autres mets qui se servent tant sur les Tables des Grands que des particuliers. Chez Pierre David, Paris, 326 pp. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k114423k>
 - Muller S, et al. (2004) Plantes invasives en France. Patrimoine Naturels, 62. Muséum National d'Histoire Naturelle, Paris, 168 pp.
 - Philippe X (1859) Flore des Pyrénées. 1. Bagnères-de-Bigorre, P. Plassot, 605 pp. URL: <https://www.biodiversitylibrary.org/page/12445645#page/8/mode/1up>
 - Planty-Tabacchi A (1997) Invasions des corridors fluviaux du Sud-Ouest par des espèces végétales exotiques. *Bulletin Français de la Pêche et de la Pisciculture* 344/345: 427-439. URL: <https://www.kmae-journal.org/articles/kmae/pdf/1997/01/kmae199734434534.pdf>
 - Ramond de Carbonnières L (1755-1825) Herbarium des Pyrénées. Herbarium Ramond, CBNPMP collection. URL: <https://cbnpmp.blogspot.com/p/acces-sites-et-outils.html>

- Rondou J (1907) Monographie de la vallée de Barèges par Jean-Pierre Rondou, instituteur à Gèdre. Tome 2 Faune, flore, géologie. 339 pp. 1 J 513 /2. Departmental Archives of the Hautes-Pyrénées, France. URL: http://www.archivesenligne65.fr/arkotheque/navigation_facette/index.php?f=ad65&reset_facette=1&crit1=6&v_6_1=Monographie+communale&v_6_2=Monographie+communale
- Société Centrale d'Apiculture (1856-1946) L'Apiculteur: bulletin de la Société Centrale d'Apiculture. Société Centrale d'Apiculture, Paris. URL: <https://gallica.bnf.fr/ark:/12148/cb34378120q/date>
- Société des Études du Comminges (1885-2004) Revue de Comminges. Société des Études du Comminges, Saint-Gaudens. URL: <https://gallica.bnf.fr/ark:/12148/cb34426879z/date>
- Société d'Horticulture (1854-1921) Annales de la Société d'Horticulture de la Haute-Garonne. Toulouse. URL: <https://gallica.bnf.fr/ark:/12148/cb34413079t/date>
- Société Nationale de Protection de la Nature (1882-1888) Bulletin de la Société Nationale d'Acclimatation de France. Société Nationale d'Acclimatation de France, Paris. URL: <https://gallica.bnf.fr/ark:/12148/cb345084375/date>
- Vaz A, Kueffer C, Kull C, Richardson D, Vicente J, Kühn I, Schröter M, Hauck J, Bonn A, Honrad J (2017a) Integrating ecosystem services and disservices: insights from plant invasions. *Ecosystem Services* 23: 94-107. <https://doi.org/10.1016/j.ecoser.2016.11.017>
- Vaz A, Kueffer C, Kull C, Richardson D, Schindler S, Jesús Muñoz-Pajares A, Vicente J, Martins J, Hui C, Kühn I, Honrado J (2017b) The progress of interdisciplinarity in invasion science. *Ambio* 46: 428-442. <https://doi.org/10.1007/s13280-017-0897-7>
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (2012) Darwin Core: an evolving community-developed biodiversity data standard. *PLOS One* 7 (1): e29715 <https://doi.org/10.1371/journal.pone.0029715>
- Wilkinson M, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018 <https://doi.org/10.1038/sdata.2016.18>

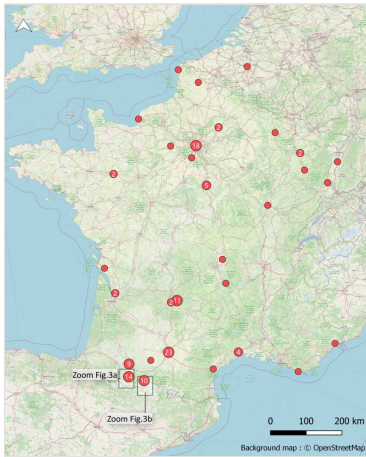


Figure 1.

Spatial location of the 122 occurrences having data associated with coordinates.

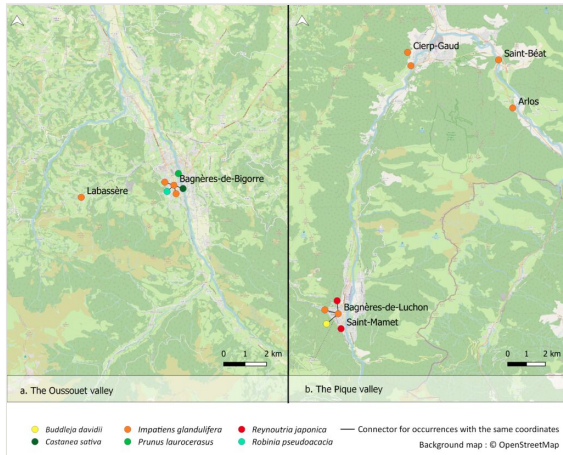


Figure 2.
Focus on the **a** Oussouet valley and **b** Pique valley.

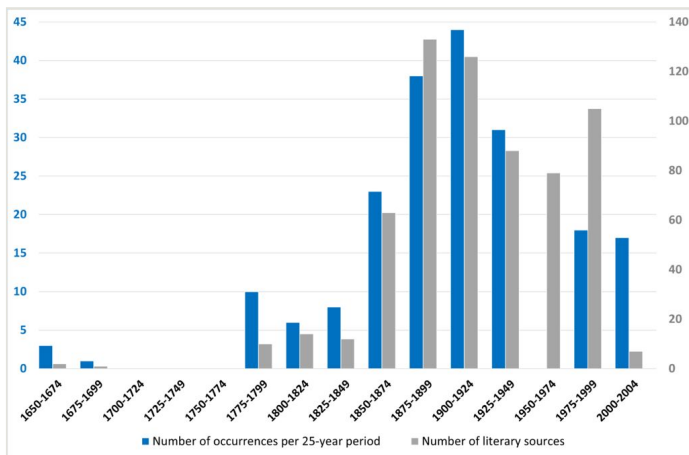


Figure 3.
 Number of occurrences found and number of literary sources consulted per 25-year period.

Table 1.

FAIRness assessment criteria used for this dataset.

FAIR principles	FAIRness assessment criteria used
FINDABLE	<ul style="list-style-type: none"> • Use a DOI for the dataset attributed by GBIF. • Use unique identifiers (UUID) for each observation occurrence. • Make persistent metadata and datasets thanks to the deposit on the GBIF platform. • Use the Ecological Metadata Language (EML) internationally recognized standard to describe the database metadata and its associated projects, including standardized search keywords. • Use a versioning system to allow future updates.
ACCESSIBLE	<ul style="list-style-type: none"> • Data storage in GBIF in accordance with the guidelines for quality standards (e.g. use of EML). • The GBIF repository provides efficient, rich services for various uses and users.
INTEROPERABLE	<ul style="list-style-type: none"> • Standard vocabularies used as far as possible for some fields. • Thesaurus used to search keywords from international thesauruses such as GEMET including INSPIRE themes. • Exclusive use of Darwin Core terms. • A Darwin Core Archive offers a stable, straightforward and flexible framework for compiling biodiversity data from varied and variable sources (source: https://www.gbif.org/en/darwin-core).
REUSABLE	<ul style="list-style-type: none"> • The Darwin Core Archive facilitates the reusability of the dataset because it enables publication in the GBIF. This compact package (a ZIP file) contains interconnected text files and enables users to share their data using a common terminology (source: https://www.gbif.org/en/darwin-core). • Use an open format for the dataset (OpenDocument.ods) and open source software to reuse it. • EML metadata includes provenance for raw and derived data. • This data paper explains the data processing steps, curation protocol, quality assurance processes, methods and tools that permit long-term integrity and understandability of data. • The spatial/temporal/taxonomic coverage is clearly mentioned in the EML metadata and in this data paper, as well as the CC-BY license and rules for large reuse.

Table 2.

Phylogenetic classification of the studied taxa ordered by order, family, genus and species with their occurrences number and their hyperlink to the subsample by species on gbif.org.

Order	Family	Genus	Species	Common name	Occurrences number (%)
Asterales	Asteraceae	<i>Helianthus</i>	Helianthus tuberosus	Jerusalem artichoke	17 (8.5)
Caryophyllales	Polygonaceae	<i>Reynoutria</i>	Reynoutria japonica	Japanese knotweed	21 (10.6)
Ericales	Balsaminaceae	<i>Impatiens</i>	Impatiens glandulifera	Indian balsam	53 (26.6)
Fabales	Fabaceae	<i>Robinia</i>	Robinia pseudoacacia	False-acacia	26 (13.1)
Fagales	Betulaceae	<i>Alnus</i>	Alnus incana	Grey alder	7 (3.5)
	Fagaceae	<i>Castanea</i>	Castanea sativa	Sweet chestnut	7 (3.5)
Lamiales	Scrophulariaceae	<i>Buddleja</i>	Buddleja davidii	Butterfly-bush	17 (8.5)
Rosales	Rosaceae	<i>Prunus</i>	Prunus cerasifera	Cherry plum	20 (10.1)
			Prunus laurocerasus	Cherry laurel	20 (10.1)
		<i>Spiraea</i>	Spiraea japonica	Japanese spiraea	11 (5.5)

Supplementary materials

Suppl. material 1: Glossary of acronyms

Authors: Morgane Claudel, Emilie Lerigoleur, Cécile Brun, Sylvie Guillaume

Data type: Glossary

[Download file](#) (1.97 kb)

Suppl. material 2: The synonymy of the ten species (TAXREF v13 and historical sources)

Authors: Morgane Claudel, Emilie Lerigoleur, Cécile Brun, Sylvie Guillaume

Data type: Vernacular name identified in historical literature

[Download file](#) (4.50 kb)

Suppl. material 3: List of historical sources consulted from which the dataset was produced

Authors: Morgane Claudel

Data type: Historical sources consulted

[Download file](#) (24.52 kb)