


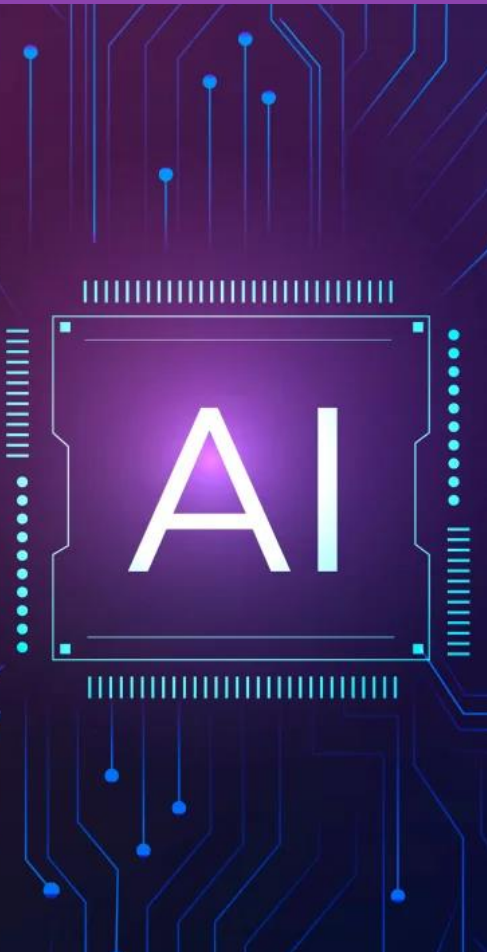
Automating Herbarium Sheet Data Extraction with AI

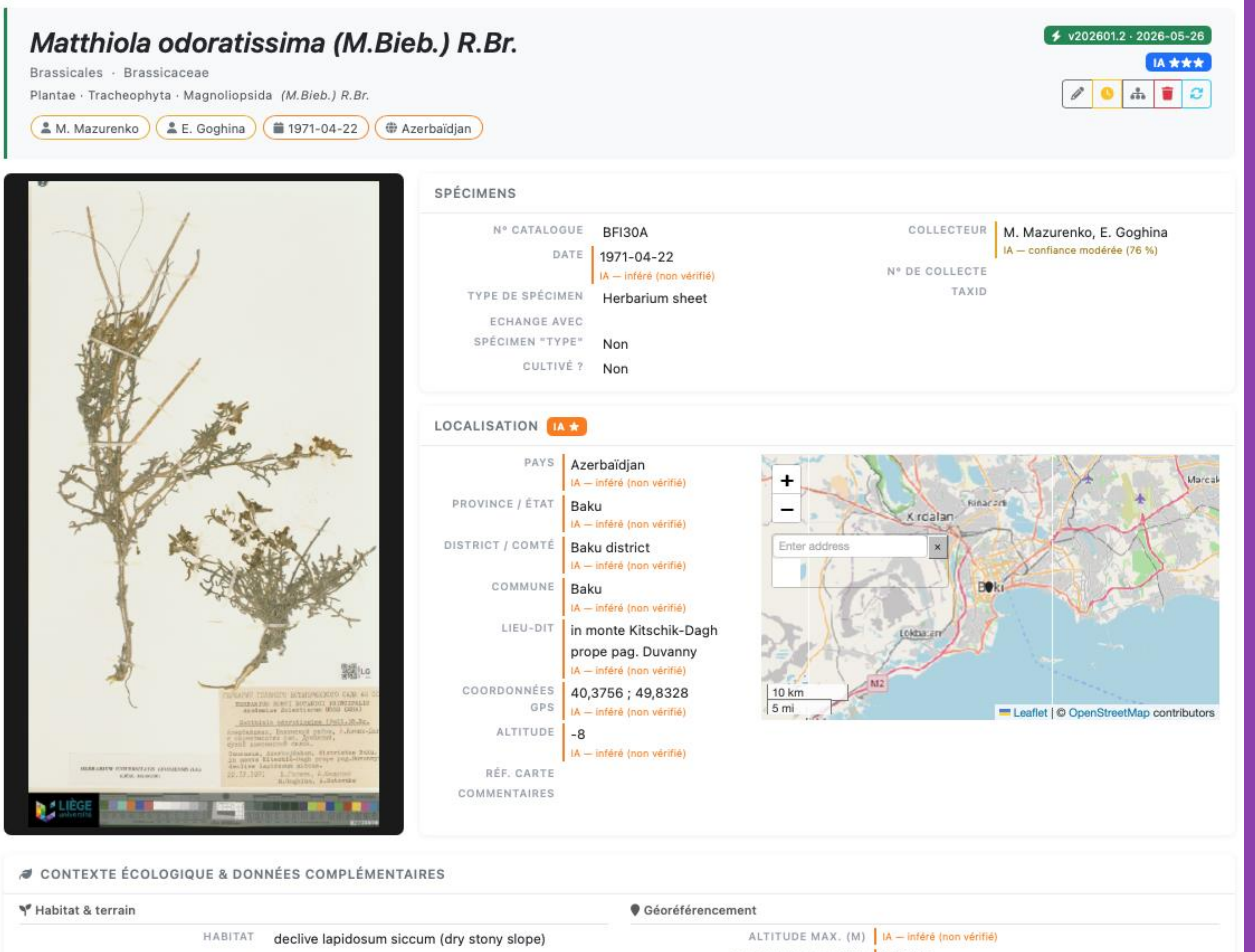
Laurent Gohy & Pierre Tocquin

University of Liege









Matthiola odoratissima (M.Bieb.) R.Br.


Brassicales · Brassicaceae
Plantae · Tracheophyta · Magnoliopsida (M.Bieb.) R.Br.

M. Mazurenko E. Goghina 1971-04-22 Azerbaïdjan

SPÉCIMENS

N° CATALOGUE	BFI30A	COLLECTEUR	M. Mazurenko, E. Goghina
DATE	1971-04-22	N° DE COLLECTE	IA — confiance modérée (76 %)
IA — inféré (non vérifié)		TAXID	
TYPE DE SPÉCIMEN	Herbarium sheet		
ECHANGE AVEC			
SPÉCIMEN "TYPE"	Non		
CULTIVÉ ?	Non		

LOCALISATION IA ★

PAYS	Azerbaïdjan	
PROVINCE / ÉTAT	Baku	
DISTRICT / COMTÉ	Baku district	
COMMUNE	Baku	
LIEU-DIT	in monte Kitschik-Dagh prope pag. Duvanny	
COORDONNÉES	40,3756 ; 49,8328	
GPS	IA — inféré (non vérifié)	
ALTITUDE	-8	
	IA — inféré (non vérifié)	

CONTEXTE ÉCOLOGIQUE & DONNÉES COMPLÉMENTAIRES

Habitat & terrain	Géoréférencement		
HABITAT	decive lapidosum siccum (dry stony slope)	ALTITUDE MAX. (M)	IA — inféré (non vérifié)
NOTES DE TERRAIN		INCERTITUDE GPS (M)	108765

The Challenge: Overcoming variability

Deciphering printed, handwritten, multilingual, and inconsistently structured herbarium labels...

...to produce structured, verified, **extended** and database-ready information



ГЕРБАРИЙ ГЛАВНОГО БОТАНИЧЕСКОГО САДА АН СО
HERBARIUM HORTI BOTANICI PRINCIPALIS
Academiae Scientiarum URSS (MHA)
Matthiola odoratissima (Pall.)R.Br.
Азербайджан, Бакинский район, г.Кичик-Даг
в окрестностях пос. Дуванный,
сухой каменистый склон.
Caucasus, Azerbajdzhan, districtus Baku,
in monte Kitschik-Dagh prope pag.Duvanny
declive lapidosum siccum.
22.IV.1971 Е.Гогина, А.Маценко
E.Goghina, A.Matsenko

ГЕРБАРИЙ ГЛАВНОГО БОТАНИЧЕСКОГО САДА АН СО
HERBARIUM HORTI BOTANICI PRINCIPALIS
Academiae Scientiarum URSS (MHA)
Matthiola odoratissima (Pall.)R.Br.
Азербайджан, Бакинский район, г.Кичик-Даг
в окрестностях пос. Дуванный,
сухой каменистый склон.
Caucasus, Azerbajdzhan, districtus Baku,
in monte Kitschik-Dagh prope pag.Duvanny
declive lapidosum siccum.
22.IV.1971 Е.Гогина, А.Маценко
E.Goghina, A.Matsenko

```
{
  "input": "т BF130A LG 58268 HERBARIUM UNIVERSITATIS LEODIENSIS LG
  LIEGE, BELGIQUE ГЕРБАРИЙ ГЛАВНОГО БОТАНИЧЕСКОГО САДА ... ",
  "specimen": {
    "scientificName": "Matthiola odoratissima (M.Bieb.) R.Br.",
    "canonicalName": "Matthiola odoratissima",
    "genus": "Matthiola",
    "family": "Brassicaceae",
    "order": "Brassicales",
    "kingdom": "Plantae",
    "phylum": "Tracheophyta",
    "class": "Magnoliopsida",
    "scientificNameAuthorship": "(M.Bieb.) R.Br.",
    "gbifTaxonKey": 7850308,
    "species_synonym": "",
    "recordedBy": [
      "Goghina, E.",
      "Matsenko, A."
    ],
    "recordedByID": [
      "1794"
    ]
  }
}
```

Step 1: Text recognition



ГЕРБАРИЙ ГЛАВНОГО БОТАНИЧЕСКОГО САДА АН СС
HERBARIUM HORTI BOTANICI PRINCIPALIS
Academiae Scientiarum URSS (MHA)

Matthiola odoratissima (Pall.)R.Br.

Азербайджан, Бакинский район, г.Кичик-Даг
в окрестностях пос. Дуванный,
сухой каменистый склон.

Caucasus, Azerbajdzhan, districtus Baku,
in monte Kitschik-Dagh prope pag.Duvanny
declive lapidosum siccum.

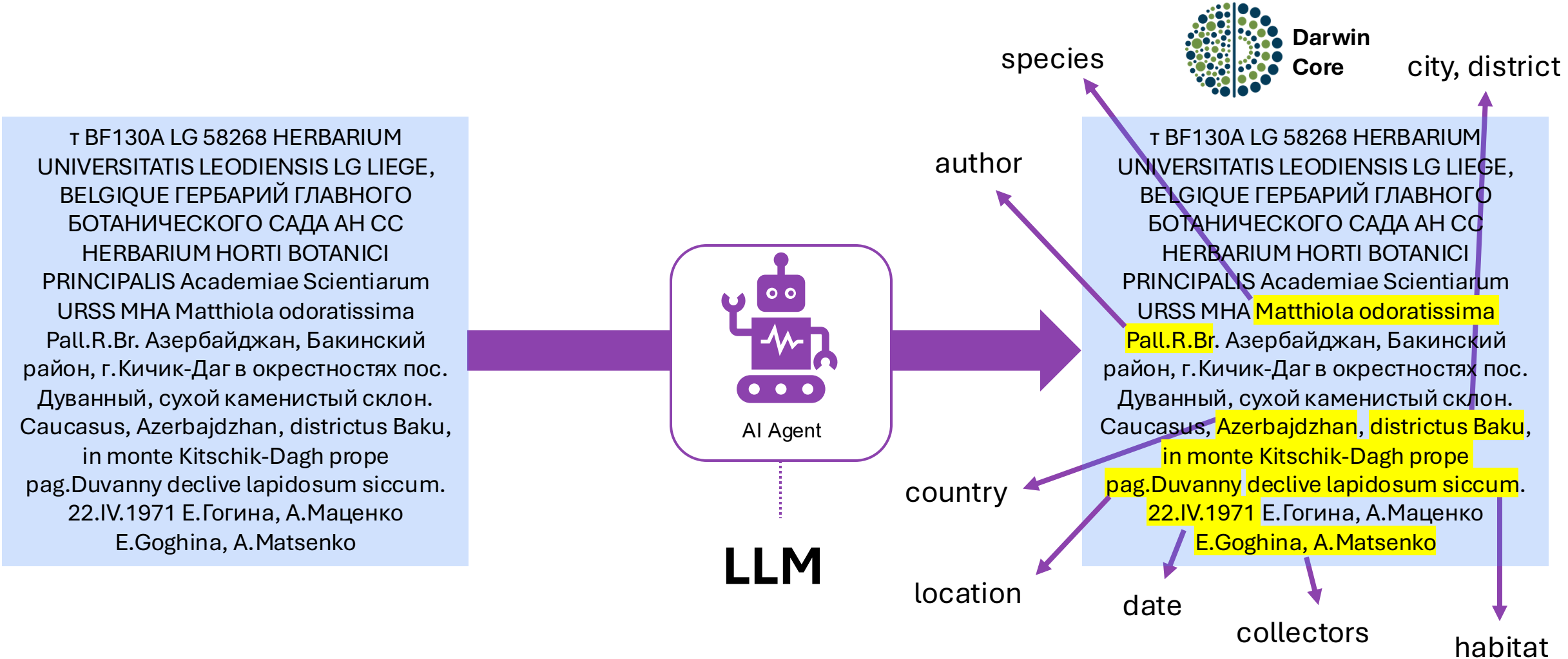
22.IV.1971 Е.Гогина, А.Маценко
E.Goghina, A.Matsenko

AI powered
OCR

т BF130A LG 58268 HERBARIUM UNIVERSITATIS LEODIENSIS LG LIEGE,
BELGIQUE ГЕРБАРИЙ ГЛАВНОГО БОТАНИЧЕСКОГО САДА АН СС
HERBARIUM HORTI BOTANICI PRINCIPALIS Academiae Scientiarum URSS
MHA Matthiola odoratissima Pall.R.Br. Азербайджан, Бакинский район,
г.Кичик-Даг в окрестностях пос. Дуванный, сухой каменистый склон.
Caucasus, Azerbajdzhan, districtus Baku, in monte Kitschik-Dagh prope
pag.Duvanny declive lapidosum siccum. 22.IV.1971 Е.Гогина, А.Маценко
E.Goghina, A.Matsenko

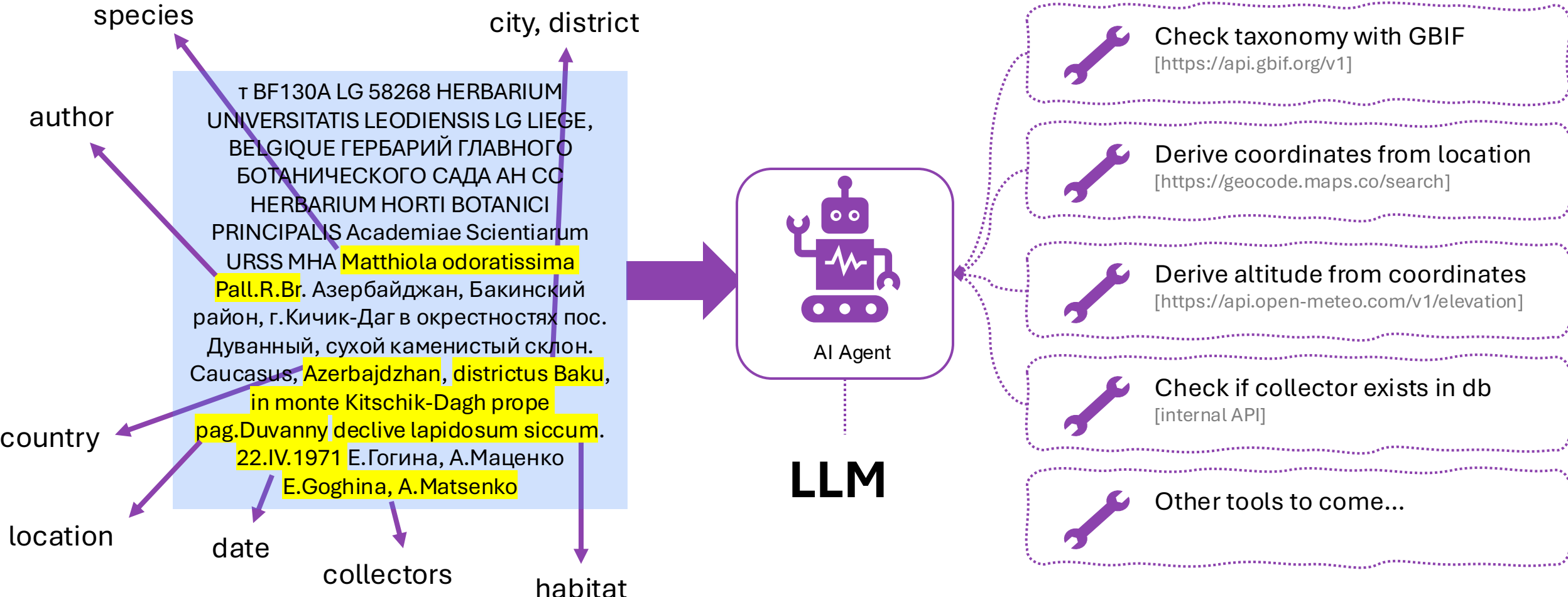
Step 2: Data processing with Agentic AI

Action 1: make sense from unstructured text using LLM



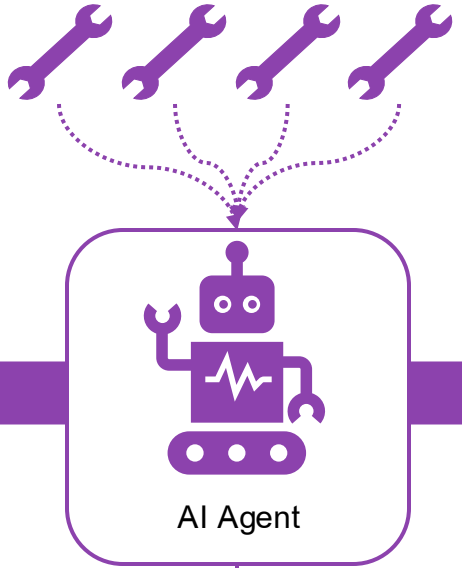
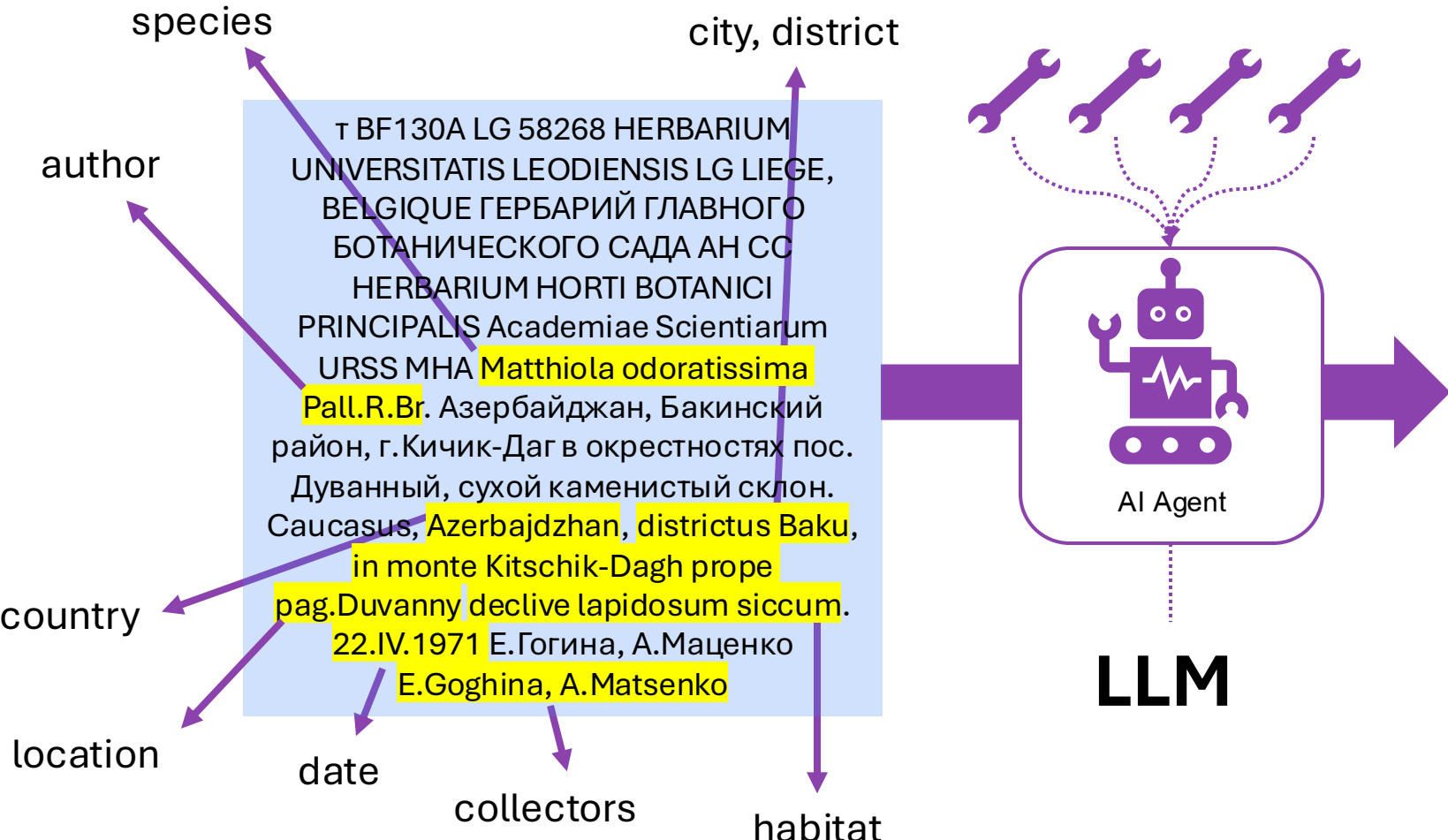
Step 2: Data processing with Agentic AI

Action 2: provide tools and let the agent decide which tool(s) to use to check or complete information



Step 2: Data processing with Agentic AI

Action 3: the agent produces a Darwin Core structured response and add it to the database



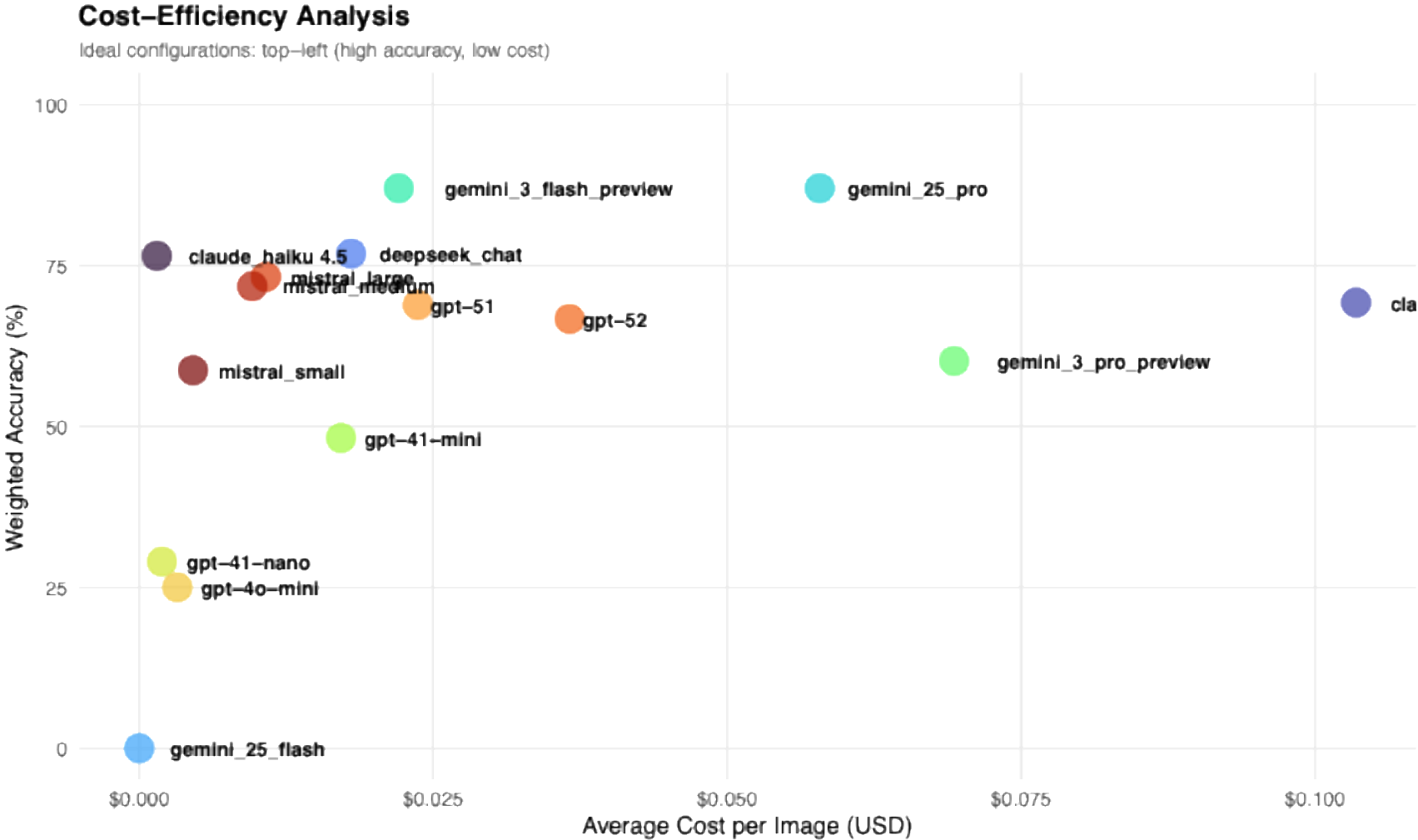
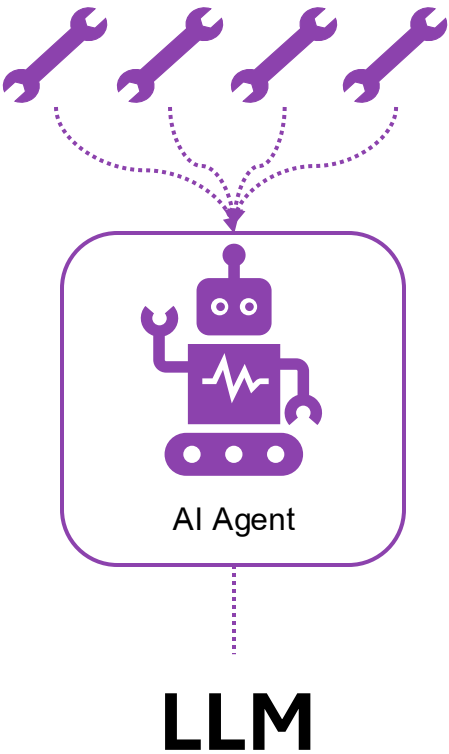
```

"specimen": {
  "scientificName": "Matthiola odoratissima (M.Bieb.) R.Br.",
  "canonicalName": "Matthiola odoratissima",
  "genus": "Matthiola",
  "family": "Brassicaceae",
  "order": "Brassicales",
  "kingdom": "Plantae",
  "phylum": "Tracheophyta",
  "class": "Magnoliopsida",
  "scientificNameAuthorship": "(M.Bieb.) R.Br.",
  "gbifTaxonKey": 7850308,
  "species_synonym": "",
  "recordedBy": ["Goghina, E.", "Matsenko, A." ],
  "recordedByID": ["1794", "1471"],
  "eventDate": "1971-04-22",
  "recordNumber": "",
  "identifiedBy": [],
  "identifiedByID": [],
  "country": "AZ",
  "stateProvince": "Baku",
  "county": "Baku district",
  "municipality": "Baku",
  "locality": "in monte Kitschik-Dagh prope pag. Duvanny",
  "decimalLatitude": 40.3755885,
  "decimalLongitude": 49.8328009,
  ....
}

```

Step 2: Data processing with Agentic AI

Choosing the right LLM: beyond size and cost...



Step 3: Data Curation

Pre-completed and AI pre-processed data greatly accelerates the human curation process



Information à propos du déterminateur ⌵

Données taxonomiques ⌴

CLASSIFICATION

ESPÈCE	GENRE	FAMILLE	ORDRE
Matthiola odoratissima (M.Bieb.) R.Br. ↻	Matthiola	Brassicaceae	Brassicales

AUTEUR

(M.Bieb.) R.Br.

IDENTIFIANTS & STATUT

TAXID	TYPE DE SPÉCIMEN	ÉCHANGE AVEC	N° CATALOGUE	RÉFÉRENCE
⌵	Herbari... ⌵	⌵	BFI30A //	<input type="text"/>

_istype _iscultivated vérifié ?

SUGGESTIONS TAXONOMIQUES

RANG	NOM CANONIQUE	NOM SCIENTIFIQUE	GENRE	FAMILLE
------	---------------	------------------	-------	---------

Localisation ⌴



Picture (4096 x 7542)

Information à propos du déterminateur

Données taxonomiques

CLASSIFICATION

ESPÈCE IA ★★★ GENRE IA ★★★ FAMILLE IA ★★★ ORDRE IA ★★★
Primula auricula L. Primula Primulaceae Ericales

AUTEUR
L.

IDENTIFIANTS & STATUT

Table with columns: TAXID, TYPE DE SPÉCIMEN, ÉCHANGE AVEC, N° CATALOGUE, RÉFÉRENCE. Includes a dropdown for 'Herbariu...' and a text field with 'WEI30A'.

istype iscultivated vérifié ?

SUGGESTIONS TAXONOMIQUES

Table with columns: RANG, NOM CANONIQUE, NOM SCIENTIFIQUE, GENRE, FAMILLE

Localisation

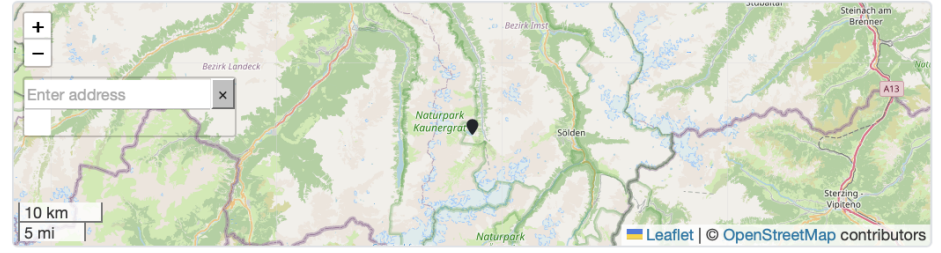
LOCALISATION GÉOGRAPHIQUE

IA ★ COMMUNE IA ★ LIEU-DIT IA ★ DATE DE COLLECTE
PAYS Autriche In monte Brandkogel Stubalpe

COORDONNÉES GPS

RÉFÉRENCE DE LA CARTE IA ★ _LONGITUDE IA ★ _LATITUDE 10.8513784 1500
46.9755144
Decimal degrees (e.g., 50.5833)

Get Coordinates Vérifier les coordonnées Obtenir les coordonnées: récupérer lat/long depuis l'adresse | Vérifier: afficher sur la carte



Stockage

Contexte écologique & Données complémentaires



Encodage Enregistrer

Ré-analyser

Information à propos du collecteur IA ★★

SÉLECTIONNER UN OU DES COLLECTEURS DANS LA LISTE

× Eggers Baron

ou encoder de nouveaux collecteurs

[+]

Information à propos du déterminateur

Données taxonomiques

CLASSIFICATION

ESPÈCE IA ★★★ GENRE IA ★★★ FAMILLE IA ★★★ ORDRE IA ★★★

Indigofera tinctoria L.

Indigofera

Fabaceae

Fabales

AUTEUR

L.

IDENTIFIANTS & STATUT

TAXID	TYPE DE SPÉCIMEN	ÉCHANGE AVEC	N° CATALOGUE	RÉFÉRENCE
	Herbariu...		YEI30A	

_istype _iscultivated vérifié ?

SUGGESTIONS TAXONOMIQUES

RANG	NOM CANONIQUE	NOM SCIENTIFIQUE	GENRE	FAMILLE
------	---------------	------------------	-------	---------

Localisation

LOCALISATION GÉOGRAPHIQUE

IA ★	COMMUNE IA ★	LIEU-DIT IA ★	IA ★
PAYS			DATE DE COLLECTE
Îles Vierges des	St. Thomas	St. Thomas, partis orientales insulae	1880-01-01

COORDONNÉES GPS

RÉFÉRENCE DE LA CARTE	_LATITUDE	_LONGITUDE	ALTITUDE	INFORMATIONS COMPLÉMENTAIRES

Decimal degrees (e.g., 50.5833)

Get Coordinates

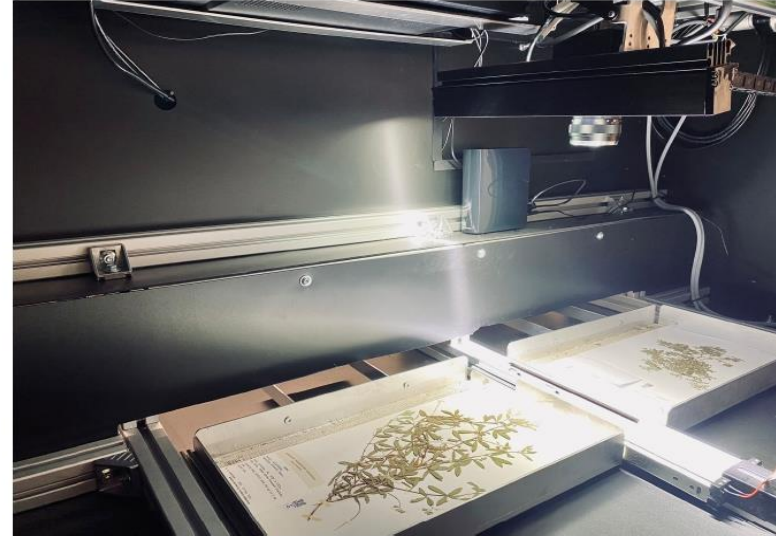
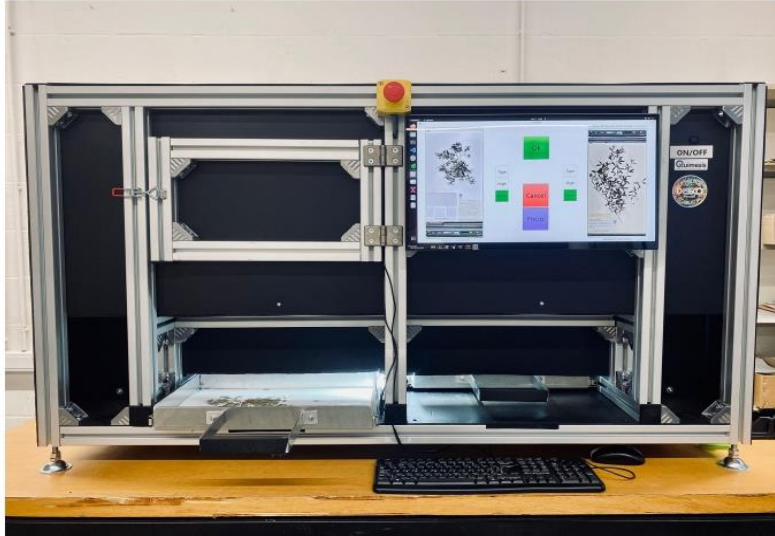
Vérifier les coordonnées

Obtenir les coordonnées: récupérer lat/long depuis l'adresse | Vérifier: afficher sur la carte

				10-17		Date	Loc.					
Lycoperdon perlatum Pers.	EFI30A	Demoulin, Vincent	BE	1969-10-02	2026-06-04	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Lycoperdon perlatum Pers.	DFI30A	Demoulin, Vincent	BE	1969-09-27	2026-06-04	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Cyathus olla	ZBF30A	Lausberg, P.	BE	1971-08-30	2026-06-02	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Cyathus olla	YBF30A	Demoulin, Vincent	BE	1974-07-18	2026-05-29	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Matthiola odoratissima (M.Bieb.) R.Br.	BFI30A	Goghina, E.E.; Mazurenko, M.	AZ	1971-04-22	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Matthiola taurica (Conti) Grossh.	CFI30A		RU	1938-05-12	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Onosma taurica var. brevifolia DC.	AFI30A	Nydegger, Max	TR	1993-05-20	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Carlina oligocephala subsp. oligocephala	ZEI30A	Nydegger, Max	TR	1980-07-29	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Indigofera tinctoria L.	YEI30A	Eggers, Baron	VI	1880-01-01	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Reseda lutea var. vulgaris Müll.Arg.	XEI30A	Henin, Henri	BE	1907-07-26	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Primula auricula L.	WEI30A		AT		2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Viola cazorlensis Gand.	KFI30A	Segura Zubizarreta, A.	ES	1977-05-07	2026-05-26	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Cladonia verticillata	YZH30A		FR	1921-08-01	2026-05-22	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Cladonia cervicornis subsp. verticillata (Hoffm.) Ahti	ZZH30A	Kozioł, Edward	PL	1997-07-28	2026-05-22	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					
Cladonia cervicornis subsp. verticillata (Hoffm.) Ahti	BAI30A	Serusiaux, Emmanuel	BE	1988-07-01	2026-05-22	Tax.	Col.	Det.	Coord.	Alt.		
						Date	Loc.					

Showing 26 to 50 of 38,234 entries

Scan with a digital camera high resolution.



One picture every 18 secondes, perfect at a depth of field of 5cm



Join Us



SCAN ME

THANK YOU

