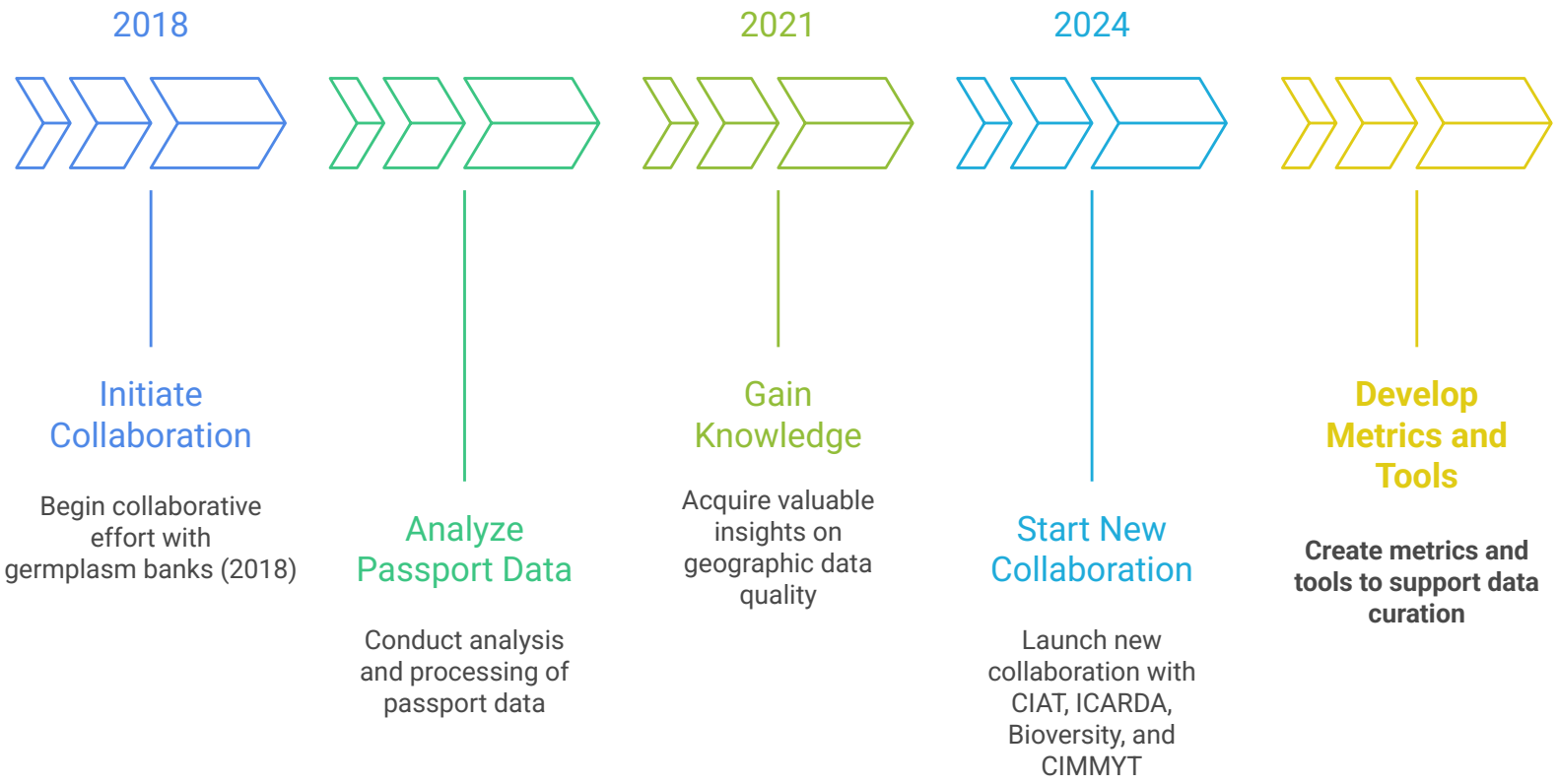
Geographical Quality Score (GQS)

Alliance of Bioversity International &CIAT
November 2025

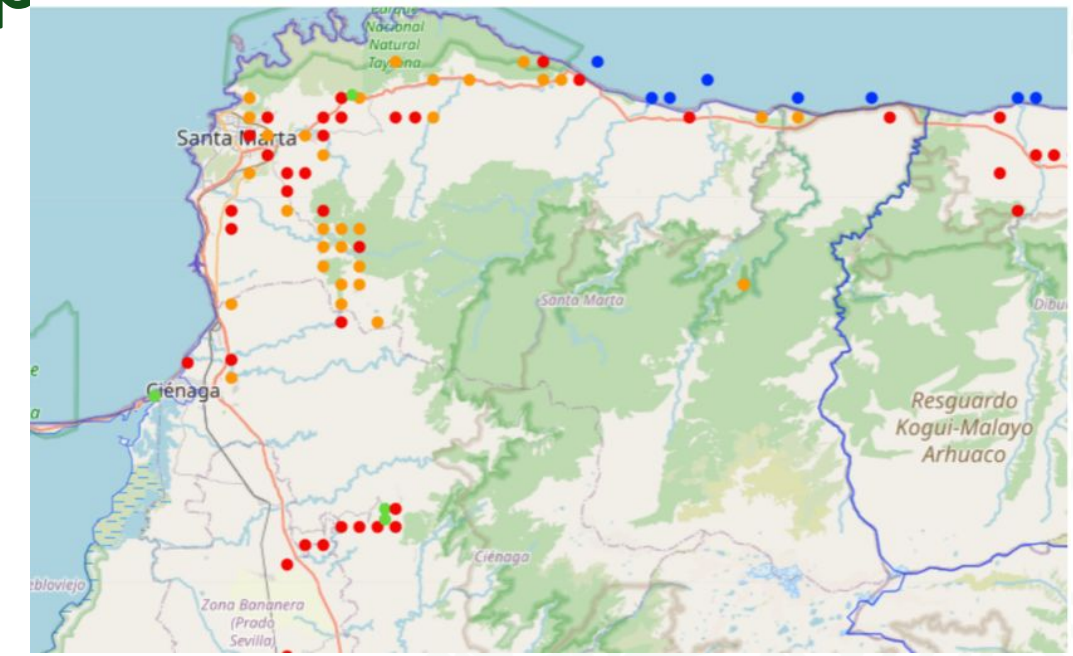
Passport data curation energizer activity

A little bit of history



Some Findings in geographical coordinates

- Coordinates in the ocean
- Duplicated coordinates
- Low coordinates precision: Coordinates can fall right outside the correct region or country.
- Inverted coordinates sign: Either the latitude, the longitude, or both have the wrong sign.
- Zero coordinates
- Coordinates out of collecting country : Country of collection in the passport data does not match to where the coordinate falls.
- Straight-line pattern in latitude or longitude
- Missing E/W and/or N/S letters and symbols º or ’ or “,
- Lack of Collection Site Information for some accessions.
- Wrong elevation data
- Incomplete Geographical Data



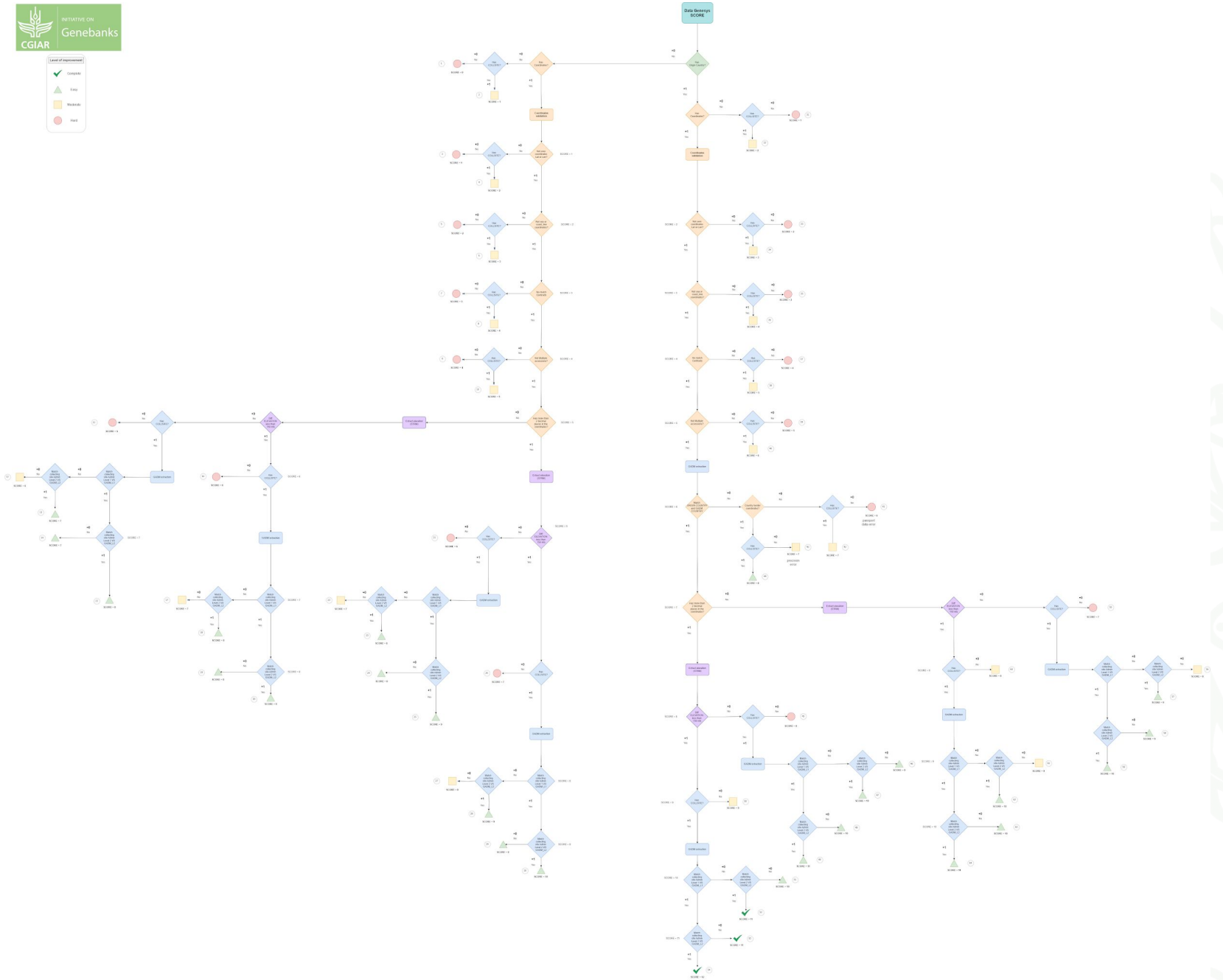
Common_names	Latitude	Longitude
Rochelas	2.9167 º N	75 º W
Revoltura	2.9167N	75W
Revoltura	2.9167N	75

Scope

Design and develop tools to facilitate and strengthen the curation of passport data in germplasm bank collections.

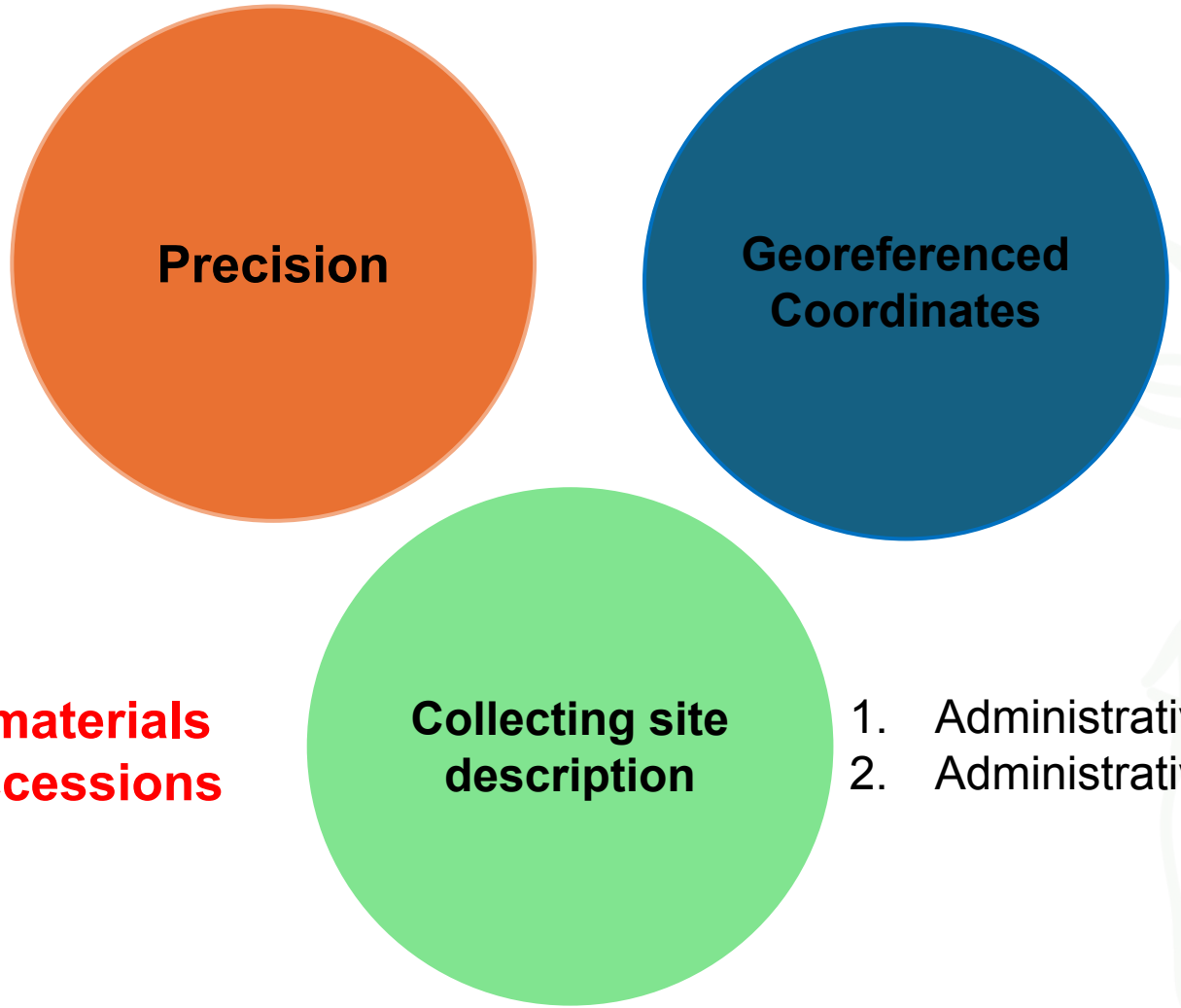
How

Leveraging the expertise of curators from ICARDA, CIAT, Bioversity, and CIMMYT, we developed a strategic plan to assess the quality of reported passport data


INITIATIVE ON
Genebanks
CGIAR
Level of improvement
Complete
Easy
Moderate
Hard
Data Greenys SCORE

Three main categories

1. Low number of decimal Places in decimal Lat/Lon
2. Origin country
3. Zero decimal coordinate
4. Coordinate on SEA



1. Country centroid
2. Province Centroid
3. High pop. density City centroid
4. Institute centroid

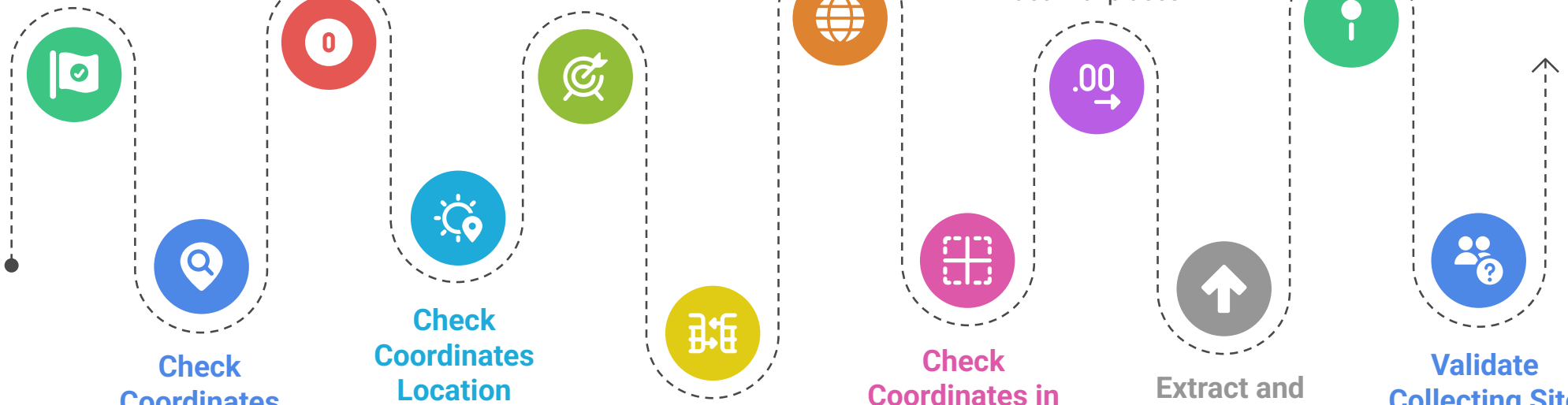
Donated materials
Market accessions

1. Administrative level 1 Miss-match
2. Administrative level 2 Miss-match

Required Fields

1. **ORIGCTY**: country of origin
2. **ELEVATION**: elevation reported in meters
3. **INSTCODE**: Institution code (FAO WIEWS)
4. **DECLATITUDE**: decimal degree latitude
5. **DECLONGITUDE**: decimal degree longitude
6. **ACCENUMB**: accession unique identifier
7. **COLLSITE**: description of the collecting site

Passport Data Validation Process



Verify Country of Origin

Ensure the reported country is accurate

Check Coordinates Availability

Confirm latitude and longitude are present

Verify Zero Coordinates

Identify if coordinates have zero values

Check Coordinates Location

Determine if coordinates are in the sea or coast

Match Coordinates with Centroids

Compare coordinates with known centroids

Validate Accessions with Coordinates

Ensure accessions match duplicated coordinates

Match Origin Country with GADM

Align the origin country with GADM data

Check Coordinates in Border

Verify if coordinates are in border areas

Check Decimal Places

Ensure coordinates have at least two decimal places

Extract and Compare Elevation

Calculate and compare elevation data

Check Collecting Site Availability

Confirm the collecting site is available

Validate Collecting Site with GADM

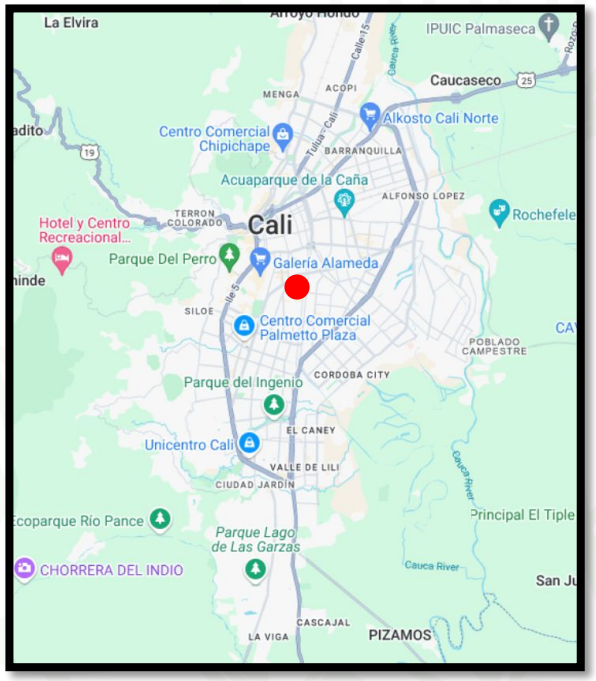
Match the collecting site with GADM administrative levels

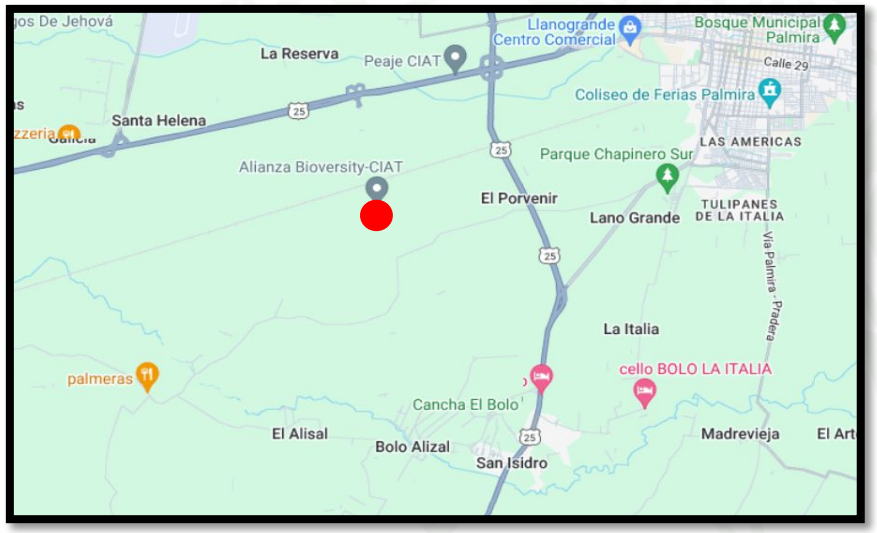
Verification fields

1. Check if accession has origin country.
2. **Check if coordinate has values of Zero.**
3. **Check if coordinate falls on sea.**
4. **Check if coordinate was georeferenced to a country, county, city and institution central coordinate.**
5. Check if more than 25 different accession have the same coordinate.
6. **Check if country of origin matches coordinate country.**
7. Check if coordinate has more than 2 decimal places.
8. Check if difference between ELEVATION and STRM elevation is less than 150 meters.
9. Check if COLLSITE is reported
10. Check if coordinate GADM administrative levels 1 and 2 matches COLLSITE description

Check if coordinate was georeferenced to a country, county, city and institution central coordinate

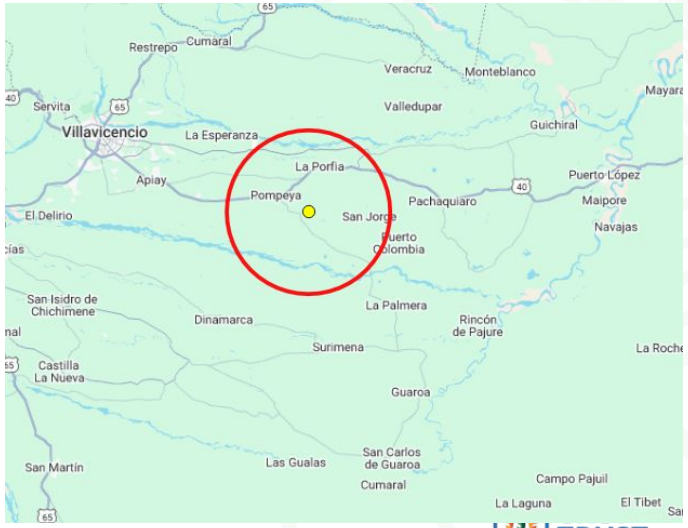
- Accession does not have coordinate, but still have COLLSITE description, nevertheless COLLSITE description lacks precision e.g:
  - Colombia
  - Cali
  - CIAT
- Through a geolocation process, missing coordinates were determined by using either the central coordinate (centroid) of the city or the location of the institution/genebank.



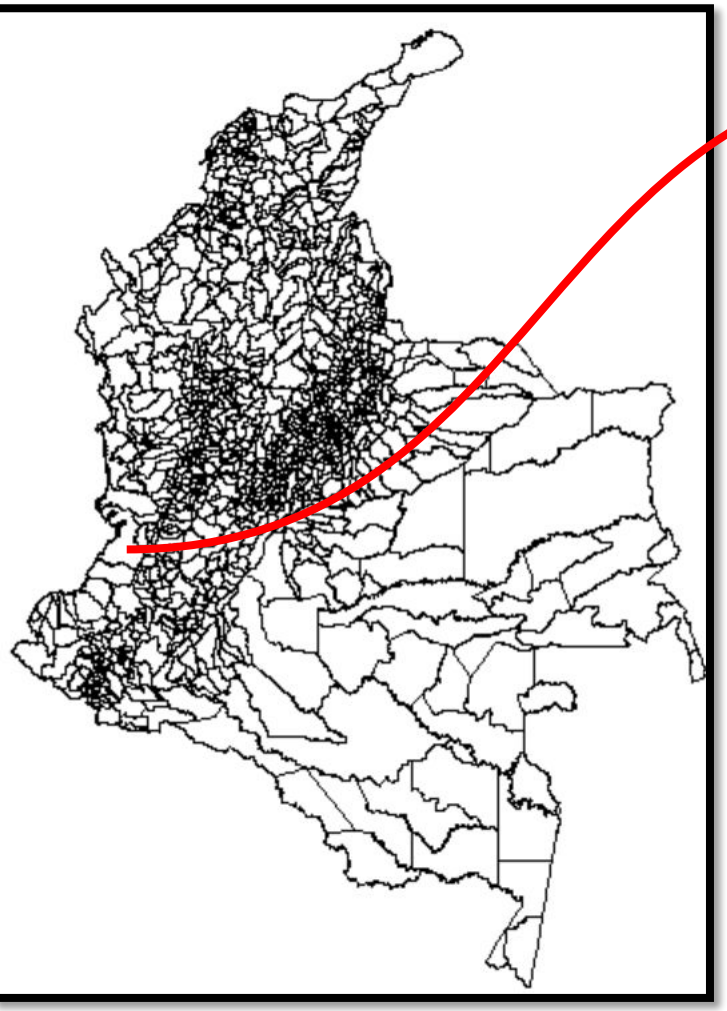


Check if coordinate has more than 2 decimal places.

- No decimals in coordinate could represent an uncertainty of 111.1 kilometers e.g., lon = 76; lat = 2
- One single decimal place in coordinate could represent an uncertainty of 11.1 kilometers

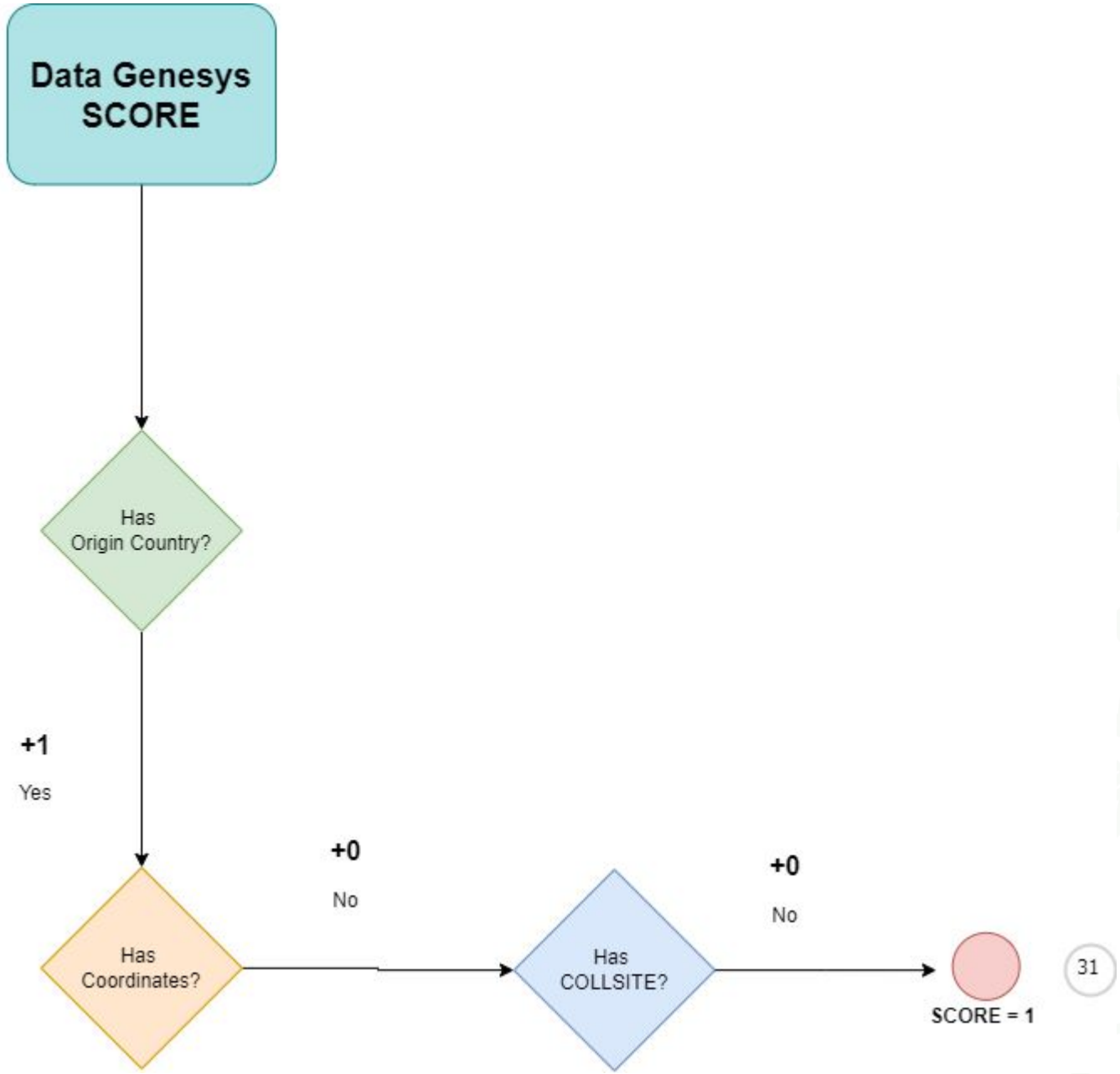


Check if coordinate GADM administrative levels 1 and 2 matches COLLSITE description

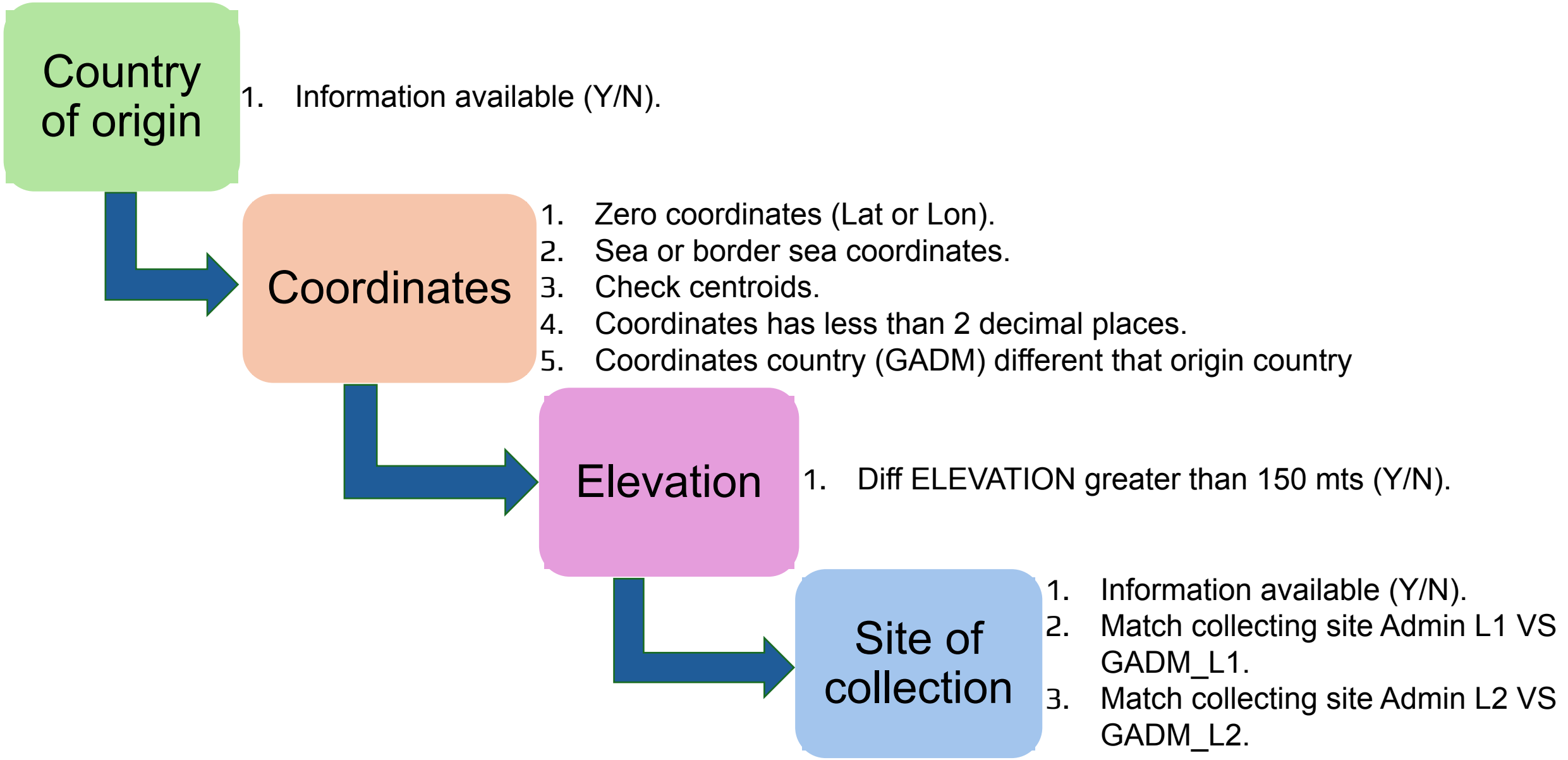


- Coordinate: Lat:3.488 Lon: -76.45
- GADM:
  - Level 0: Colombia
  - Level 1: Valle del cauca
  - Level 2: Cali

- **COLLSITE:** 2 Km east of **Cali**, **valle del cauca**, colombia**.** **(level 1 and level 2 matched)**
- **COLLSITE: Colombia (No match)**


Data Genesys
SCORE
Has
Origin Country?
+1
Yes
Has
Coordinates?
+0
No
Has
COLLSITE?
+0
No
SCORE = 1
31

Hierarchy of Variables



Why does this matter for conducting a gap analysis ?

Quality score Dashboard

https://bit.ly/BOLDWP1

(Ecuador, Morocco and Pakistan)

https://bit.ly/PDFFC

(Colombia, Kenya and Zambia)

Examples

ACC_ID: INIAP-ECU-15376

Quiality: **Low**
Level of improvement: **Moderate hard to curate**

Country of origin

1. Information available □ Yes
   ECU (Ecuador)

Coordinates
Lat 1.672
Lon-80.2989

1. Zero coordinates (Lat or Lon) □ No
2. Sea or border sea coordinates □ **Yes**
3. Check centroids □ No
4. Coordinates less than 2 decimal places No
5. Coordinates country (GADM) different that origin country □ NO

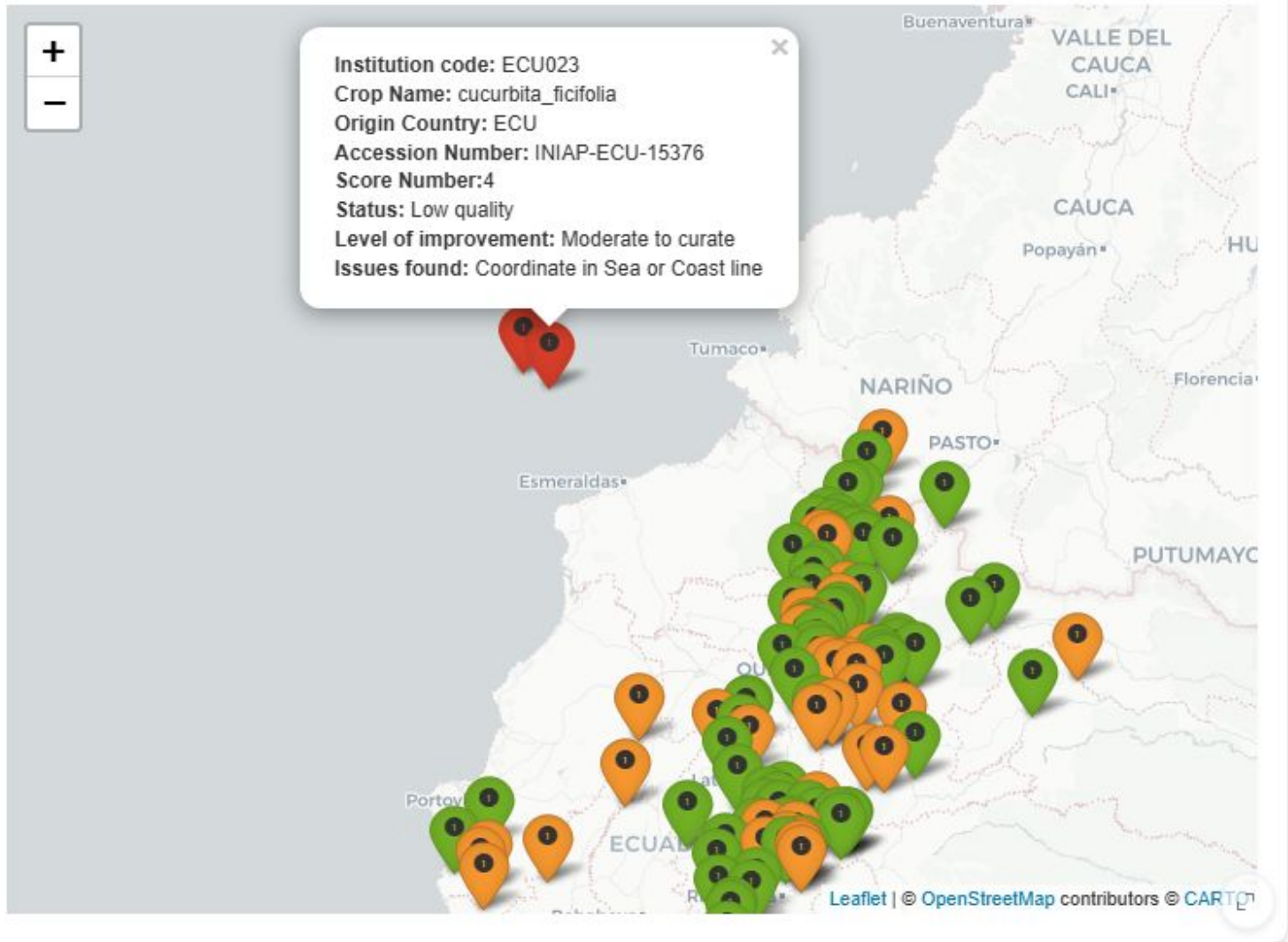
Elevation

1. Diff ELEVATION greater than 150 mts No

Collection site

1. Information available □ Yes

Imbabura, Cotacachi, Imantag


Institution code: ECU023
Crop Name: cucurbita_ficifolia
Origin Country: ECU
Accession Number: INIAP-ECU-15376
Score Number:4
Status: Low quality
Level of improvement: Moderate to curate
Issues found: Coordinate in Sea or Coast line
Buenaventura
VALLE DEL CAUCA
CALI
CAUCA
Popayán
Tumaco
NARIÑO
PASTO
Florencia
Esmeraldas
PUTUMAYO
Portoviejo
ECUADOR
Leaflet | © OpenStreetMap contributors © CARTO

ACC_ID: INIAP-ECU-12363

Quiality: **Moderate**
Level of improvement: **Moderate hard to curate**

Country of origin

1. Information available □ Yes
   ECU (Ecuador)

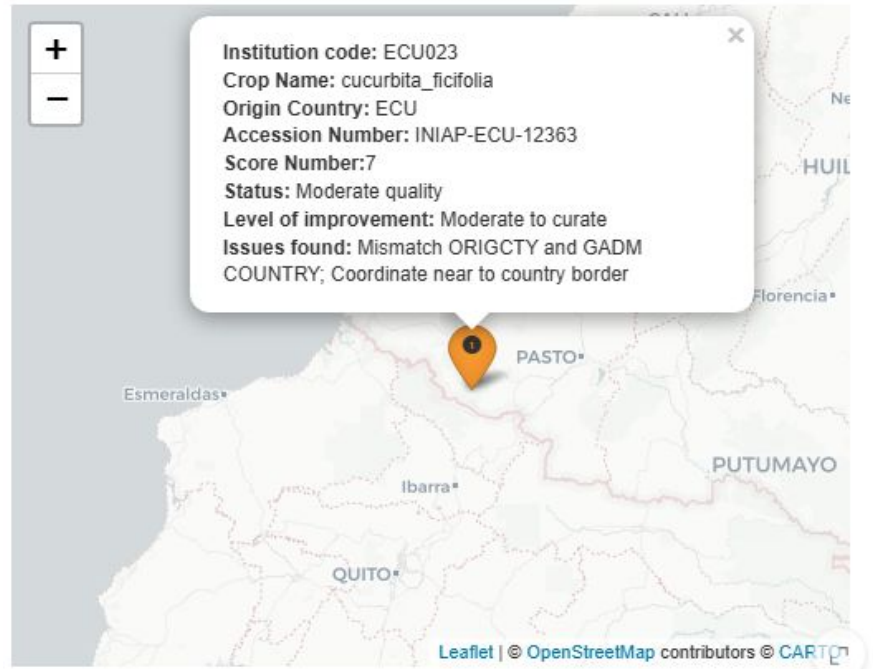
Coordinates

Lat 1.0063
Lon -78.01666

1. Zero coordinates (Lat or Lon) □ No
2. Sea or border sea coordinates □ No
3. Check centroids □ No
4. **Coordinates country (GADM) different that origin country**
   □ **Yes**
   COL
5. **Country border coordinates**

Site of collection

1. Information available □ Yes
   Ambuquí,Ibarra,Imbabura



Filter by issue

Mismatch ORIGCTY and GADM COUNTRY

	ACCENUMB	COLLSITE	CROPNAME	DECLONGITUDE	DECLATITUDE	ORIGCTY	GADM_ISO3	QUALITY_SCORE	LEVEL_OF_IMPROVEMENT
	Filter		Filter					Filter	Filter
○	INIAP-ECU-12363	Ambuquí,Ibarra,Imbabura	cucurbita_ficifolia	-78.0166666666667	1.00638888888889	ECU	COL	Moderate	Moderate

ACC_ID: INIAP-ECU-5582

Quiality: **Moderate quality**
Level of improvement: **Moderate hard to curate**

Country of origin

1. Information available □ Yes
   ECU (Ecuador)

Coordinates
Lat 10.3
Lon –64.4

1. Zero coordinates (Lat or Lon) □ No
2. Sea or border sea coordinates □ No
3. Check centroids □ No
4. **Georeferenced to a centroid**

Collection site

1. Information available □ Yes
   San Lucas,Loja,Loja

ChatGPT 5

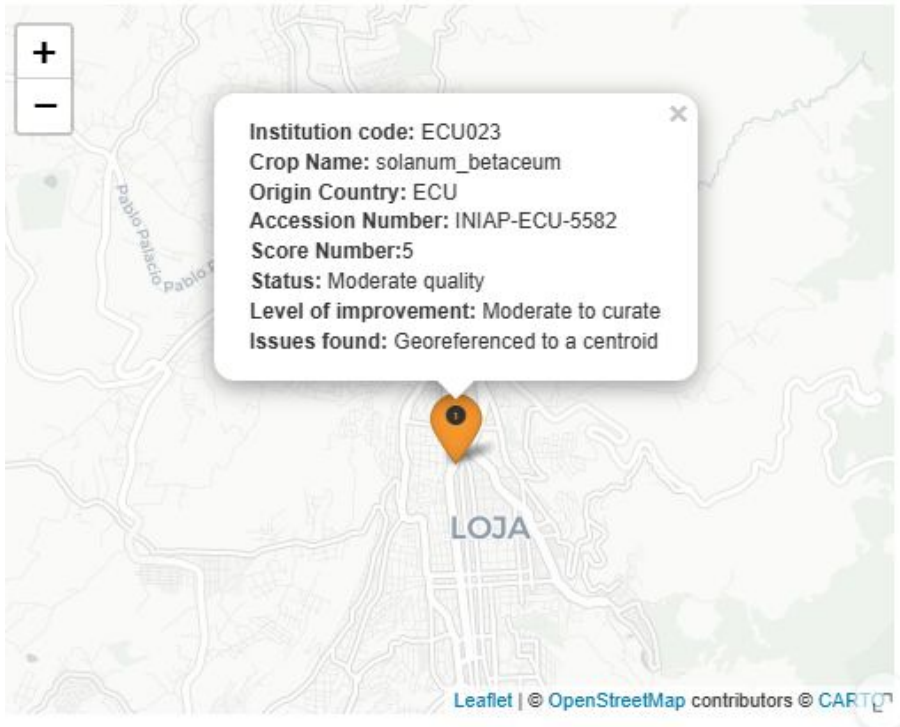
cual es el centroide de la ciudad de loja ecuador

El centroide aproximado de la ciudad de Loja, Ecuador (coordenada decimal) es:

Latitud: -3.993130°

Longitud: -79.204220° Date and Time +2

Sources



	ACCENUMB	COLLSITE	CROPNAME	DECLONGITUDE	DECLATITUDE	ORIGCTY	GADM_ISO3	QUALITY_...
	Filter		Filter					Filter
○	INIAP-ECU-5582	San Lucas,Loja,Loja	solanum_...	-79.205	-3.99055555555...	ECU	ECU	Moderate

Questions?