**Process**:

It seems the process could be:

* Machine-intelligent processes try to discover the values of the fields. Iterative.
* Recognize what is wrong or incomplete (Machine-intelligent processes)
* Correct or complete with other Machine-intelligent processes
* Recognize what is wrong or incomplete (Machine-intelligent processes)
* Correct or complete the remaining with Human-intelligent processes

**Data Model:**

* Fields:
  1. Original image
  2. Image after OCR
  3. Scientific name
  4. Authorship (of the name)
  5. Country
  6. State
  7. County *(I believe this is difficult to collect, the notion of county is North American)*
  8. Location (Geographic Description)
  9. Specimen’s Habitat and Description
  10. Collector(s) name(s)
  11. Collector(s) number(s)
  12. Year (collected)
  13. Month (collected)
  14. Day (collected)
* Other fields could be:
  1. Geographical coordinates
  2. Institution, organization
  3. Study name
* All fields, except Original image, are multivalued: there can be more than one version. Because for example: one process generates one date, other process says is a different date (ambiguous data). Maybe we could work with probabilities.
* Secondary structures:
  1. Dictionary of species (with synonyms)
     + Plants list. <http://plants.usda.gov/dl_all.html>
     + The Plant List. <http://www.theplantlist.org/> Accepted: 350,699; Synonyms: 470,624 terms.
     + Integrated Taxonomic Information System, <http://www.itis.gov/> 703,662 plants, animals, fungi, and microbes of North America and the world.
     + Gbif has a scientific name parser: <http://tools.gbif.org/nameparser/api.do>
  2. Table of countries, states, counties. Problem: Different languages.
     + <http://www.geonames.org/export/>
     + <https://www.maxmind.com/en/free-world-cities-database>
     + <https://developer.yahoo.com/geo/geoplanet/>

O T H E R A N N O T A T I O N S :

Notes from notesfromnature.org:

* Images without country, but they have a city or place, from which we could get the Country
* There is no help for the scientific name: we should avoid errors. We should have a list of all the scientific names, with common names and synonyms. <http://resolver.globalnames.org/api>

Fields for plants:

Species name

* Scientific name

Where:

* Country
* State
* County
* Location (Geographic description)
* Habitat and Description

Date (When)

* Month
* Day
* Year

Who:

* Name of the person (I believe it should be last name, first name)
* Collector number

Fields for insects (calbug expeditions):

* Country
* State/Province
* County
* Locality
* Begin date collected
* End date collected
* Collector
* Elevation (There are no units specified here)
* Other notes

Ornithological:

* Page number
* Registration number (year, batch number, record)
* Scientific name
* Location
* Date Collected (Day, month, year)

Macrofungi:

* Country
* State/Province
* County
* Locality and Habitat
* Collection date (Day, month, year)