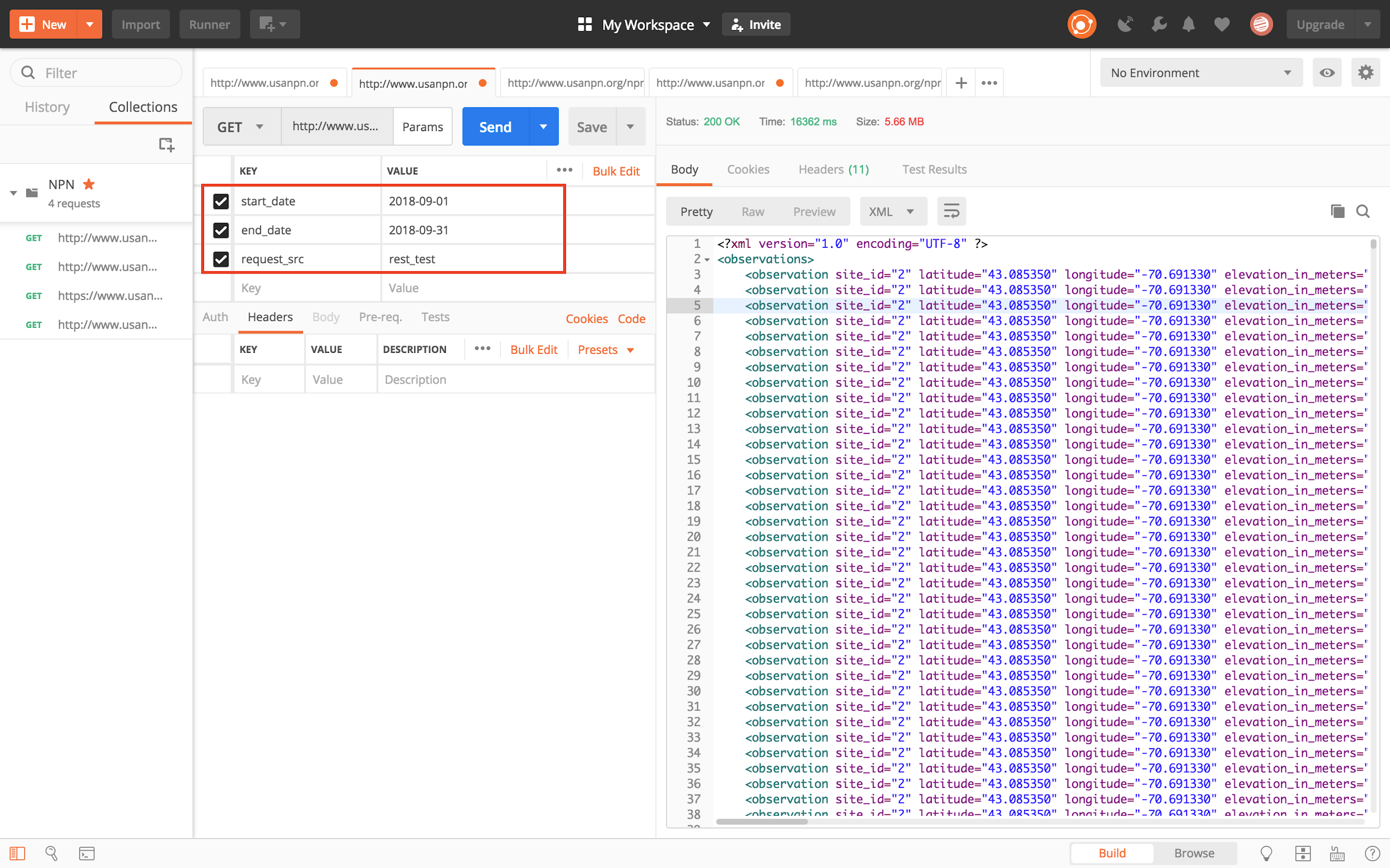
NPN API DATA SOURCING

REST API

**GetSummarizedData**

* Rest api input, xml format
* Parameters: start\_date=yyyy-mm-dd, end\_date=yyyy-mm-dd, request\_src=rest\_test
* Example input:

<http://www.usanpn.org/npn_portal/observations/getSummarizedData.xml?start_date=2018-09-01&end_date=2018-09-31&request_src=rest_test>

Postman testing – available

(data available till 2018-09-12)

Tool used to extract data: R

* Environment: RStudio
* R code required packages and library: “httr”, “jsonlite”, “lubridate”
* Package & functions prep

*#Download the packages and library for format transforming, functions*

install.packages(c("httr", "jsonlite", "lubridate"))

library(httr)

library(jsonlite)

*#For XML format functions*

install.packages("plyr")

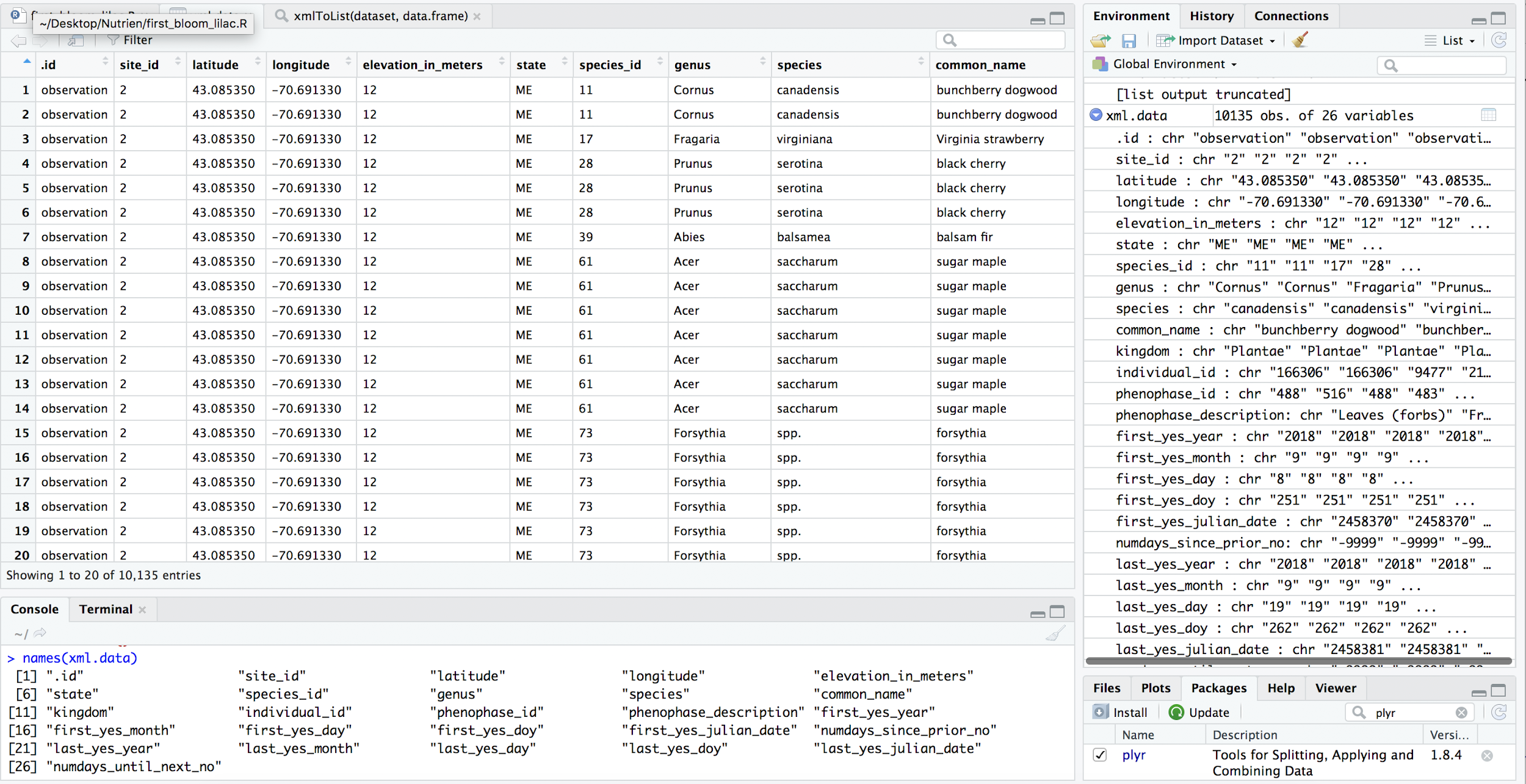
library(plyr)

require("XML")

library(XML)

library(methods)

* Data extraction Result:
* Transformed Format: Data Frame



* A glance of data frame
* Columns:

> names(xml.data)

[1] ".id" "site\_id" "latitude" "longitude" "elevation\_in\_meters"

[6] "state" "species\_id" "genus" "species" "common\_name"

[11] "kingdom" "individual\_id" "phenophase\_id" "phenophase\_description" "first\_yes\_year"

[16] "first\_yes\_month" "first\_yes\_day" "first\_yes\_doy" "first\_yes\_julian\_date" "numdays\_since\_prior\_no"

[21] "last\_yes\_year" "last\_yes\_month" "last\_yes\_day" "last\_yes\_doy" "last\_yes\_julian\_date"

[26] "numdays\_until\_next\_no"

* Fields Explanation (\*valuable fields)

.id: observation id

Site.id

\*Latitude (float)\*Longitude (float)

Elevation\_in\_meters (float): size of the plant/animal

\*kingdom: Animalia/Plantae

\*genus

\*species

species\_id

\*common\_name

individual\_id

**\*phenophase\_description: describe the plants/animals’ status/behavior**

> unique(xml.data$phenophase\_description)

[1] "Leaves (forbs)" "Fruits"

[3] "Leaves" "Colored leaves"

[5] "Recent fruit or seed drop" "Young needles (conifers)"

[7] "Falling leaves" "Flowers or flower buds"

[9] "Open flowers" "Ripe fruits"

[11] "Breaking leaf buds" "Young leaves (tree/shrub)"

[13] "Ripe seed cones" "Needles (deciduous)"

[15] "Colored needles" "Increasing leaf size"

[17] "Calls or song (birds)" "Active individuals"

[19] "Initial growth (forbs)" "All leaf buds broken (lilac/honeysuckle)"

[21] "Unripe seed cones" "Leaves (grasses)"

[23] "Flower visitation" "Active adults"

[25] "Pollen cones (conifers)" "Territorial individuals (birds)"

[27] "Insect consumption" "Individuals at a feeding station"

[29] "Feeding (or foraging)" "Singing individuals (birds)"

[31] "Fruit/seed consumption" "Nut gathering"

[33] "Young individuals" "Feeding"

[35] "Pollen release (flowers)" "Fledged young (precocial birds)"

[37] "Active caterpillars" "**Caterpillars feeding"**

[39] "Young needles (pines)" "Emerging needles (pines)"

[41] "Falling needles" "Individuals on land"

[43] "Individuals in water" "Individuals on land (rock/log)"

[45] "Young individuals (hatched from eggs)" "Migrating adults (in uniform direction)"

[47] "Flower heads (grasses/sedges)" "Open flowers (grasses/sedges)"

[49] "Adults on land" "Mating (lepidoptera)"

[51] "End of flowering (lilac/honeysuckle)" "Occupied nest"

[53] "Leaves (sedges)" "Male adults"

[55] "Open pollen cones (conifers)" "Recent cone or seed drop"

[57] "Adults in water" "Individuals in water (rock/log)"

[59] "Vocalizing" "Fledged young (altricial birds)"

[61] "Nestlings (altricial birds)" "Dead caterpillars (lepidoptera)"

[63] "Nest building (birds)" "**Pollen release (conifers)"**

[65] "Young leaves (forbs)" "Initial growth (grasses/sedges)"

[67] "Dead adults" "Adults feeding"

[69] "Dead individuals" "Partially-fledged young (precocial birds)"

[71] "Summer coat" "Breaking leaf buds (lilac/honeysuckle)"

[73] "Mating (male on top)" "Dead nestlings or young (birds)"

[75] "Courtship " "Migrating adults (insect swarm)"

[77] "Inactive nymphs (dormant)" "Open flowers (lilac)"

[79] "Full flowering (lilac)" "**Nest building (ground-nesting insect)**"

\*First\_yes\_year :Year first appear

\*First\_yes\_month: Month

\*First\_yes\_day: Day

\*First\_yes\_julian\_date: Convertible, to yyyy-mm-dd

\*Last\_yes\_year :Year

\*Last\_yes\_month: Month

\*Last\_yes\_day: Day

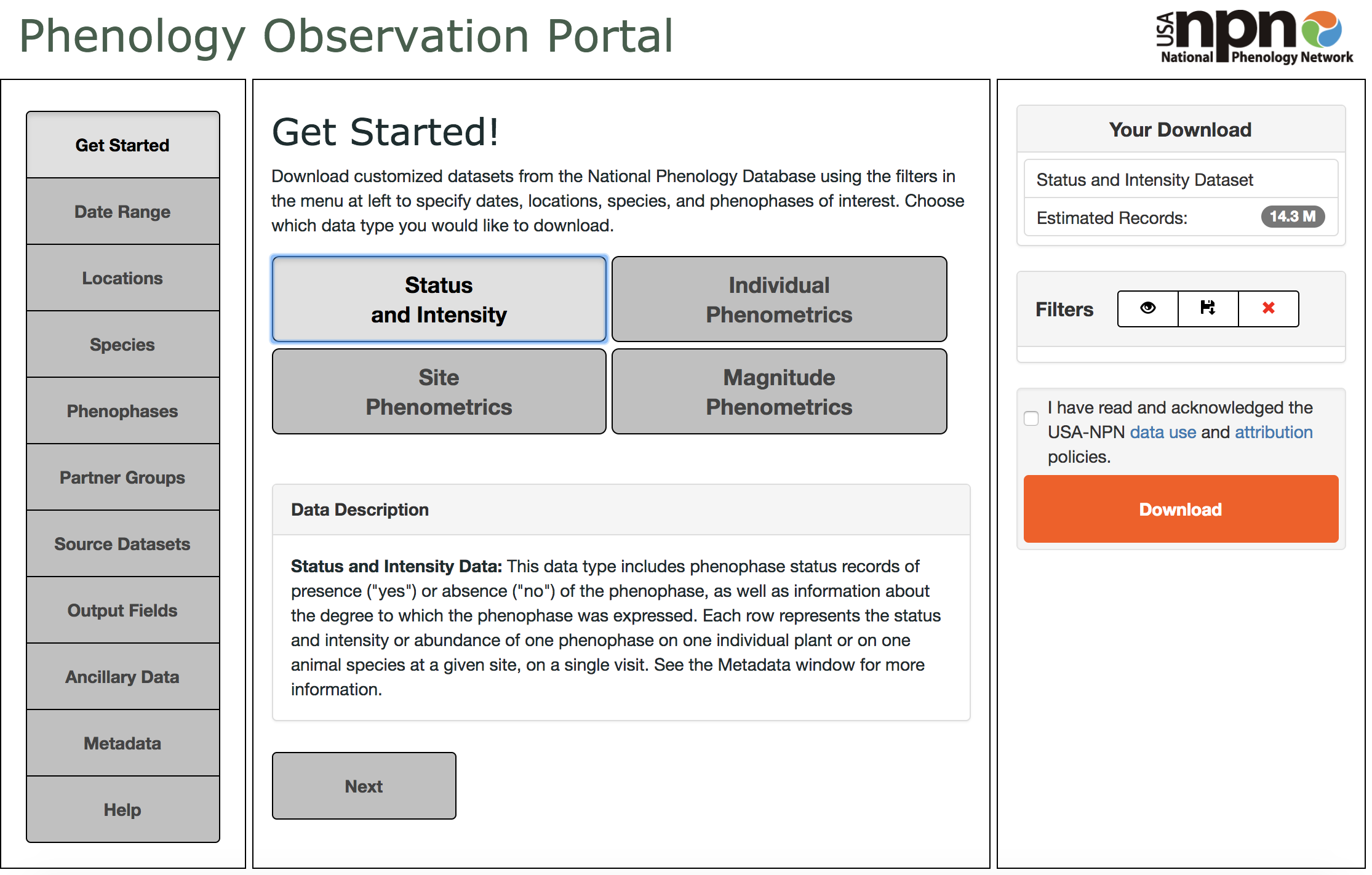
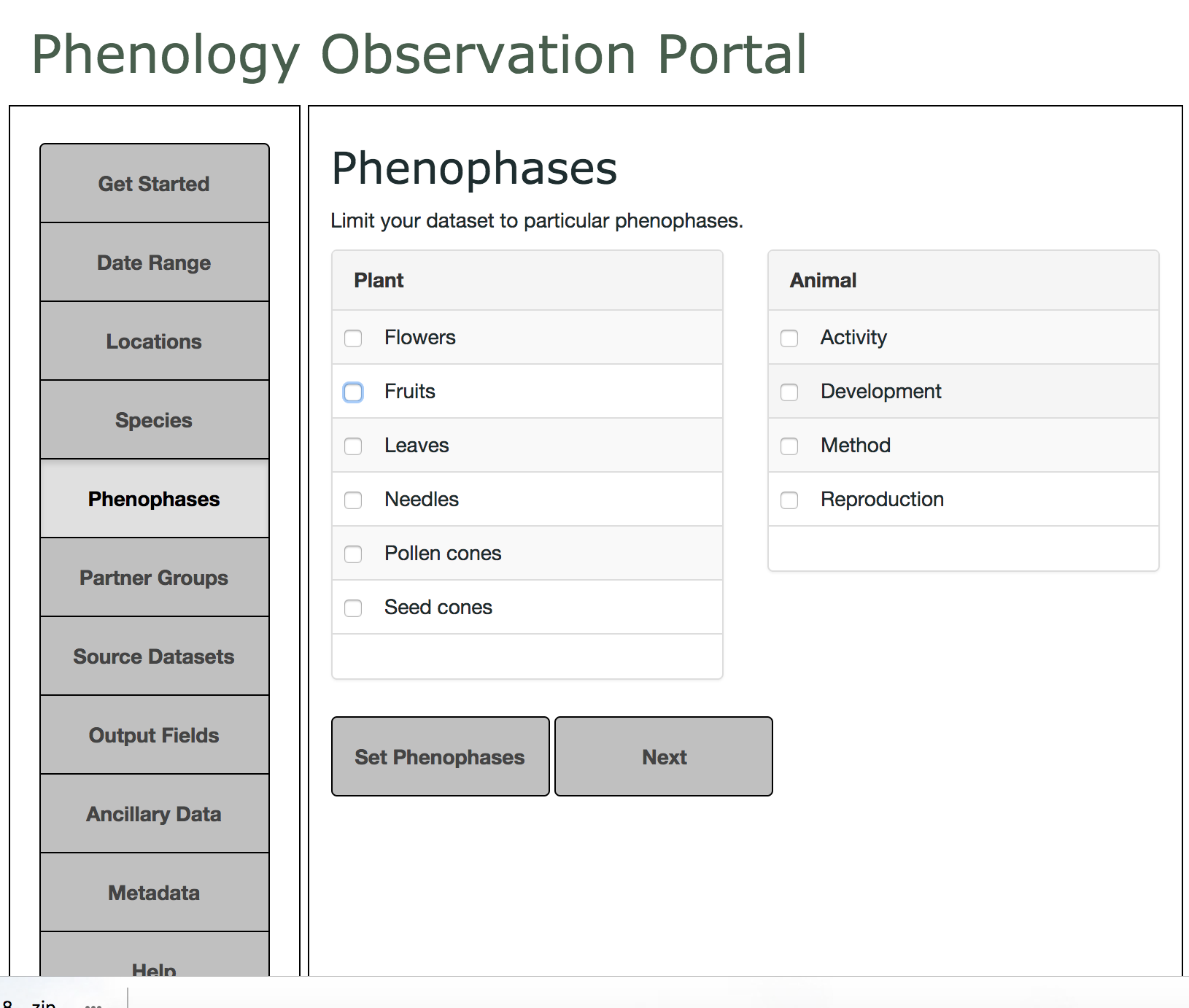
numdays\_since\_prior\_no: -9999 is NULL, only positive integers are valid

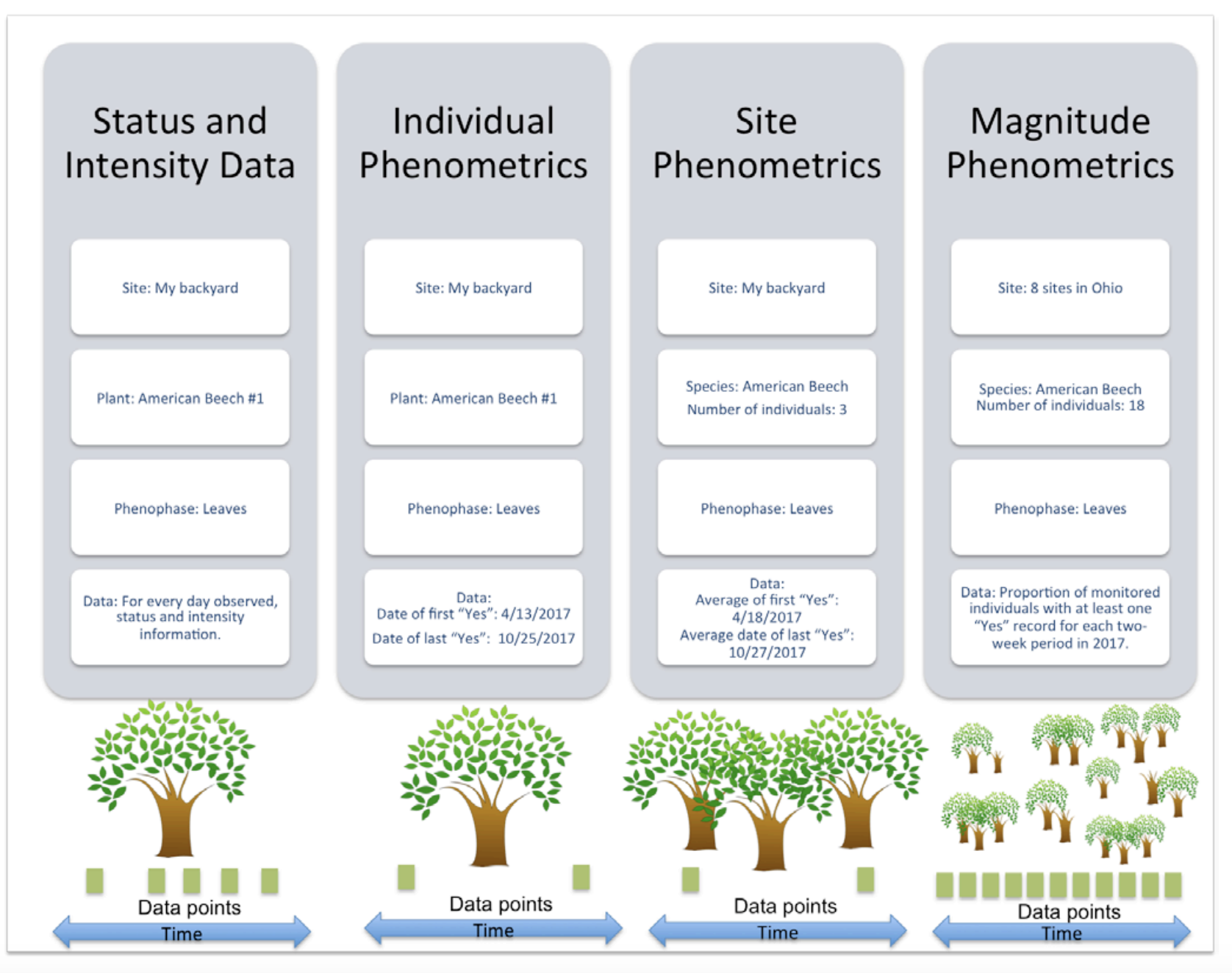
numdays\_until\_next\_no: -9999 is NULL, only positive integers are valid

APPENDIX

You can download the dataset data directly from NPN site, but dataset is not comprehensive

[Phenology Observation Portal](https://data.usanpn.org/observations/get-started)

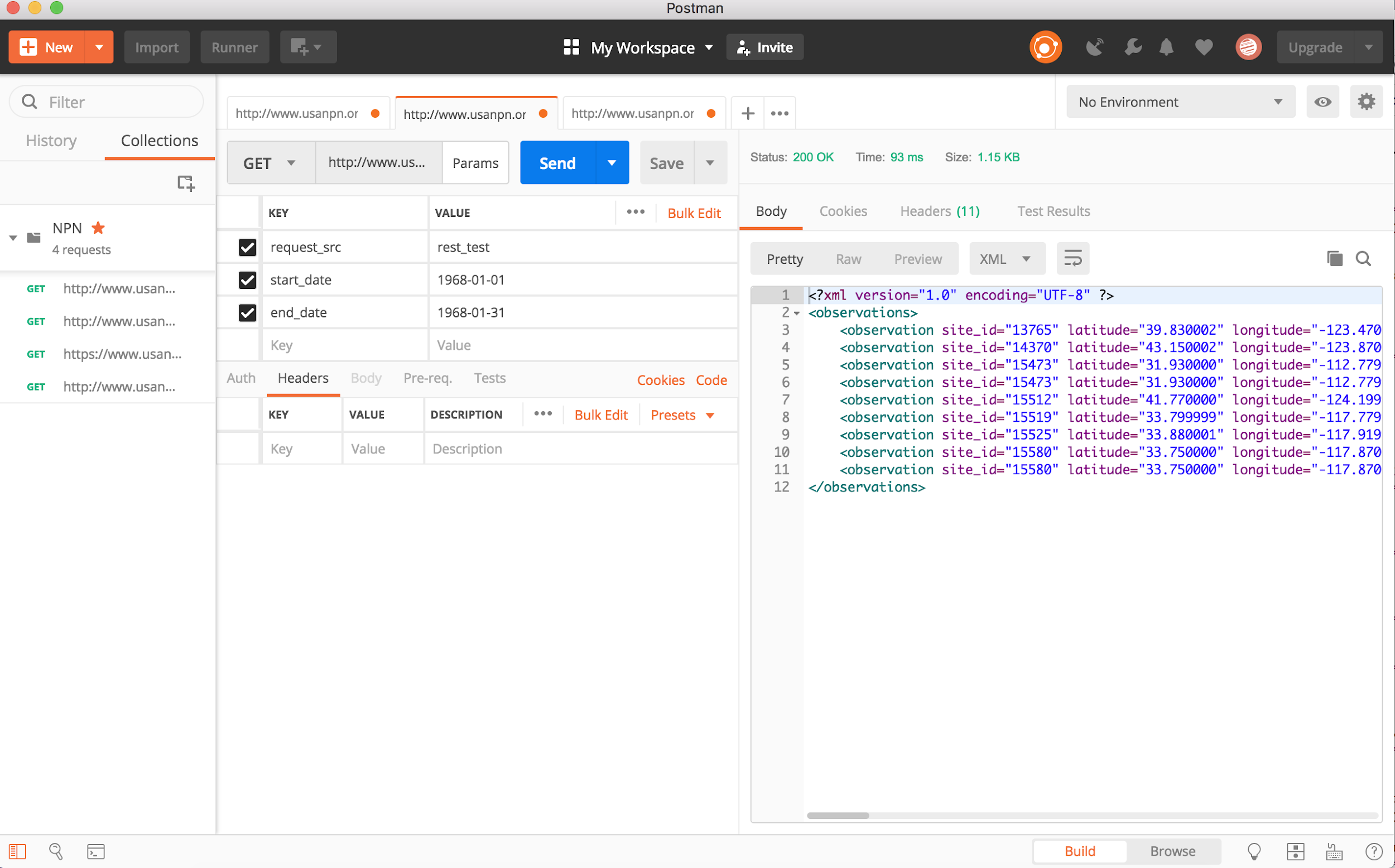




http://www.usanpn.org/npn\_portal/observations/getSummarizedData.xml?start\_date=2015-01-01&end\_date=2015-01-31&start\_date=2015-02-01&end\_date=2015-02-31&start\_date=2015-03-01&end\_date=2015-03-31&start\_date=2015-04-01&end\_date=2015-04-31&start\_date=2015-05-01&end\_date=2015-05-31&start\_date=2015-06-01&end\_date=2015-06-31&start\_date=2015-07-01&end\_date=2015-07-31&start\_date=2015-08-01&end\_date=2015-08-31&start\_date=2015-09-01&end\_date=2015-09-31&start\_date=2015-10-01&end\_date=2015-10-31&start\_date=2015-11-01&end\_date=2015-11-31&start\_date=2015-12-01&end\_date=2015-12-31&start\_date=2016-01-01&end\_date=2016-01-31&start\_date=2016-02-01&end\_date=2016-02-31&start\_date=2016-03-01&end\_date=2016-03-31&start\_date=2016-04-01&end\_date=2016-04-31&start\_date=2016-05-01&end\_date=2016-05-31&start\_date=2016-06-01&end\_date=2016-06-31&start\_date=2016-07-01&end\_date=2016-07-31&start\_date=2016-08-01&end\_date=2016-08-31&start\_date=2016-09-01&end\_date=2016-09-31&start\_date=2016-10-01&end\_date=2016-10-31&start\_date=2016-11-01&end\_date=2016-11-31&start\_date=2016-12-01&end\_date=2016-12-31&start\_date=2017-01-01&end\_date=2017-01-31&start\_date=2017-02-01&end\_date=2017-02-31&start\_date=2017-03-01&end\_date=2017-03-31&start\_date=2017-04-01&end\_date=2017-04-31&start\_date=2017-05-01&end\_date=2017-05-31&start\_date=2017-06-01&end\_date=2017-06-31&start\_date=2017-07-01&end\_date=2017-07-31&start\_date=2017-08-01&end\_date=2017-08-31&start\_date=2017-09-01&end\_date=2017-09-31&start\_date=2017-10-01&end\_date=2017-10-31&start\_date=2017-11-01&end\_date=2017-11-31&start\_date=2017-12-01&end\_date=2017-12-31&start\_date=2018-01-01&end\_date=2018-01-31&start\_date=2018-02-01&end\_date=2018-02-31&start\_date=2018-03-01&end\_date=2018-03-31&start\_date=2018-04-01&end\_date=2018-04-31&start\_date=2018-05-01&end\_date=2018-05-31&start\_date=2018-06-01&end\_date=2018-06-31&start\_date=2018-07-01&end\_date=2018-07-31&start\_date=2018-08-01&end\_date=2018-08-31&start\_date=2018-09-01&end\_date=2018-09-31&request\_src=rest\_test

**Q&A**

- Do we know how far back the data goes (at least 3 years)?

The data can be tracked back to **1968-01-01**  
 1968 Year records are relatively limited to State = CA, OR, AZ.

- Do we know how recent the data is? Up to the day, week?

I extracted 2018 Steptember data, the most recent record is on 2018-09-09. I attached the file on the email.

- The phenophase id is great, is there something bit more generic like ‘ start of spring’?

There’s no generic fields like “start of spring”, but we can find keywords from: phenophase: falling leaf/blooming flowers/calling birds

Since the exact date is available in each observation, we can match the phase to the date/seasons to determine “start of the spring” indicators.

- What are the options for extraction by city or by county? Ideally we need a single data point averaged for an area (like city or county)

There’s only “state” information, but both latitude and longitude are available.

o Is there an input parameter for lat/long range or city? If so, does it product a single metric (averaged)?

I guess we can pull out a field featured “city”, we can create a range averaging the lat.long.

Documentation is available <https://docs.google.com/document/d/1_2SBAUpgaU6mcPUlmS8wRSBY2agDPbc_-AT3fTB1wl4/edit?usp=sharing>