**Title:** Technical workflow for near real-time reporting of forest census results

**Authors (tentative / incomplete list, not in order):**

David Mitre

Valentine Herrmann1 ( Orcid ID : 0000-0002-4519-481X)

Jess Shue

Krystal Bagnaschi

Rachel Hoffman

Iris Kennedy

Erin MacMonigle

Caroline Troy

Luca Morreale

Albert Kim

Lukas Magee

Amy Wolf

Daniel Zuleta

Stuart Davies?

(others)

William McShea

Kristina J. Anderson-Teixeira1,2\* ( Orcid ID : 0000-0001-8461-9713)

**Author Affiliations:**

1. Conservation Ecology Center; Smithsonian’s National Zoo & Conservation Biology Institute; Front Royal, Virginia 22630, USA
2. Forest Global Earth Observatory, Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Panama City, Panama

\*corresponding author: [teixeirak@si.edu](mailto:teixeirak@si.edu); +1 540 635 6546

**Running head:**

# Abstract

**Keywords**:

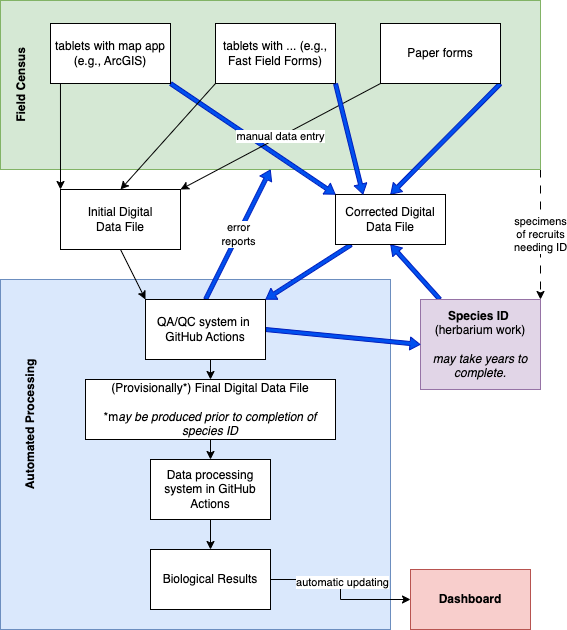
# 1 Introduction

*Par. 1: Forests are important for climate regulation and support of biodiversity (UN goals). Therefore we need quality, up-to-date data on forest C and biodiversity.*

*Par. 2: While remote sensing is vital to mapping forest biomass and diversity globally, ground-based censuses are essential for calibrating/ interpreting satellite data. In short, ground-based censuses are vital to quantifying forest C, diversity, and trends therein.*

*Par. 3: Yet, collection of accurate census data is not trivial. (challenges include developing a good, efficient system for recording data and time lags in checking data)*

*Par 4: To address these challenges, we have developed a workflow for data collection, quality control, and automated production and publication of results. This is developed at ForestGEO sites (Anderson-Teixeira et al., 2015; Davies et al., 2021), but applicable to any forest census.*



**Figure 1. Workflow Diagram.** (Blue arrows indicate iterative process that is completed until the data set is found to be error free.)

# 2 Materials and Methods

We prototyped the system described here– either in full or in part – at **##** ForestGEO sites representative of a range of conditions under which data are collected across forest research sites. *(describe all sites, perhaps include a table*)

## 2.1 Step 1: Production of Digital Data Records from Field Census

Data were collected via a variety of mechanisms, ultimately leading to digital data files. Here, we applied three different methods for censuses at ForestGEO plots.

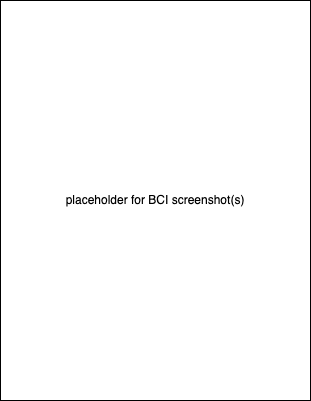
### 2.1.1 ArcGIS App

Examples: BCI, SCBI

ArcGIS Field Maps saves time both with navigation and data entry. After downloading an offline area of the map, all the trees are symbolized as dots on the map falling within labeled quadrats.



**Figure 2. Screenshots from the ArcGIS app on iPads taken during the 2023 census of the Smithsonian Conservation Biology Institute ForestGEO plot:** (a) map zoomed into four quadrats (20 x 20 m) showing all trees in the previous (2018) census and their 2023 census status (green = complete, yellow = in progress, red = not started); (b) data form showing information about the previously censused tree selected on the map (turquoise outline); (c) data form for adding a new recruit (grey field editable) at a location selected on the map (+ symbol); (d) warning (“DBH check”, bottom field) generated after a suspicious measurement of diameter at breast height (DBH) was entered in the editable “dbh 2023” field for the tree shown in panel (b).



**Figure 3. Screenshot showing of web application of ArcGIS.** Shown are **##** ha of the 50-ha Barro Colorado Island ForestGEO plot near the completion of the 2022-2023 census..

### 2.1.2 Fast Field Forms App

Examples (potential): Wabikon, Tyson, and Lilly Dickey (2022)

*Figure?: screenshots of app*

### 2.1.3 Paper Records

Example: Amacayacu

At Amacayacu, paper records were collected in the [YEAR] census. Data were digitized (entered into spreadsheets) every weekend. Here, to provide an example of how this system can work with paper data collection, we analyzed digital records as collected in the field (no corrections applied).

## 2.2 Step 2: Automated Data Processing via GitHub Actions

*(GitHub is excellent for scientific data management, including automated processing)* (Kim et al., 2022; **braga\_not\_2023?**)

### 2.2.1 Data QA/QC

### 2.2.2 Biological Results

Generate maps & .csv files for each of the following (by quadrat, as soon as each quadrat is completed and found to be error-free):

* n stems
* n gained (recruitment)
* n lost (mortality)
* ∆ n stems
* biomass
* woody productivity
* woody mortality
* ∆ biomass
* species richness
* n species gained
* n species lost
* ∆ species richness

### 2.2.3 Results Dashboard

### 2.2.4 Step 3: Finalizing Data with Species ID of Recruits

Tropical sites typically have a large lag between data collection and finalization because all new recruits must be identified to species. Given high species diversity, it is not possible for even expert botanists to identify all species in teh field; rather, specimens must be collected and identified later with reference to botanical collections. This process can take years, meaning that census data are often not finalized until years after the census has been completed. While this process must be completed in order to quantify species richness, stem abundance can be calculated, and biomass estimated, before this process is complete.

# 3 Results

## 3.1 Error Rates and Detection

*Table #?. For each site: census pace, initial data error and warning rates (perhaps broken into categories),…*

## 3.2 Biological Results

# 4 Discussion

# 5 Conclusions (optional)

# Acknowledgements

* Funding: ForestGEO

# Conflict of Interest statement

The authors declare no conflict of interest.

# Authors’ contributions

*[Name of author 1] and [Name of author 2] conceived the ideas and designed methodology; [Name of author 1] and [Name of author 3] collected the data; [Name of author 2] and [Name of author 4] analysed the data; [Name of author 1] and [Name of author 4] led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.*

# Data availability

# References

Anderson-Teixeira, K. J., Davies, S. J., Bennett, A. C., Gonzalez-Akre, E. B., Muller-Landau, H. C., Joseph Wright, S., Abu Salim, K., Almeyda Zambrano, A. M., Alonso, A., Baltzer, J. L., Basset, Y., Bourg, N. A., Broadbent, E. N., Brockelman, W. Y., Bunyavejchewin, S., Burslem, D. F. R. P., Butt, N., Cao, M., Cardenas, D., … Zimmerman, J. (2015). CTFS-ForestGEO : A worldwide network monitoring forests in an era of global change. *Global Change Biology*, *21*(2), 528–549. <https://doi.org/10.1111/gcb.12712>

Davies, S. J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., Alonso, A., Anderson-Teixeira, K., Andrade, A., Arellano, G., Ashton, P. S., Baker, P. J., Baker, M. E., Baltzer, J. L., Basset, Y., Bissiengou, P., Bohlman, S., Bourg, N. A., Brockelman, W. Y., Bunyavejchewin, S., … Zuleta, D. (2021). ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biological Conservation*, *253*, 108907. <https://doi.org/10.1016/j.biocon.2020.108907>

Kim, A. Y., Herrmann, V., Bareto, R., Calkins, B., Gonzalez-Akre, E., Johnson, D. J., Jordan, J. A., Magee, L., McGregor, I. R., Montero, N., Novak, K., Rogers, T., Shue, J., & Anderson-Teixeira, K. J. (2022). Implementing GitHub Actions continuous integration to reduce error rates in ecological data collection. *Methods in Ecology and Evolution*, *13*(11), 2572–2585. <https://doi.org/10.1111/2041-210X.13982>