Quantitative Guidelines for Establishing and Operating Soil Archives

Edward Ayres*
National Ecological Observatory Network (NEON)
1685 38th Street, Suite 100
Boulder, CO 80301

Soil archives are widely recognized as important, but there is limited information available on the basic elements of archive design, such as how much material to archive, which sites and horizons to prioritize, and the most effective way to promote an archive. Here I address these gaps based on 6 yr of operating the National Ecological Observatory Network's (NEON) Megapit Soil Archive, which has provided over 1500 samples to researchers. Active outreach by NEON staff was the most effective way of raising awareness of the archive among the user community. On average, 14.6 ± 4.3 g soil was sent to requesters per sample per year, with soils from A horizons and intensively studied sites consumed fastest. Based on these consumption rates and depending on soil horizon and site type, archiving between 22.3 and 39.9 g soil for each year of a sample's desired lifespan would be sufficient to ensure that it is not exhausted earlier than expected with ≥89% confidence. Similarly, 500 g of archived soil would be sufficient for the vast majority of samples to last at least 12 to 22 yr, depending on horizon and site type. A simple equation is proposed, and parameterized using NEON data, to determine the quantity of soil necessary to archive if the desired sample lifespan is known. Setting a desired lifespan can inform archive management decisions, such as the number of requests to approve and whether to increase or decrease the allocation of resources to raising awareness of the archive.

Abbreviations: NEON, National Ecological Observatory Network; NRCS, Natural Resource Conservation Service; USDA, US Department of Agriculture.

Core Ideas

- **Soil archive samples were consumed at 15 ± 4 g yr^{-1} (mean ± SD).**
- **Soils from A horizons and intensively studied sites were consumed fastest.**
- **Depending on site and horizon, 500 g archived soil would last at least 12 to 22 yr.**
- **Informal interactions were the most effective way to raise archive awareness.**

Collections of physical samples, including archived soils, are widely considered an important and valuable resource for scientific advancement (Boone et al., 1999; Cary and Fierer, 2014; Dolfing and Feng, 2015; Karssies and Wilson, 2015; Manter et al., 2017; Richter et al., 2007). Archived soil samples have been used in countless studies, with notable examples including the identification of declines in environmental lead pollution following regulatory changes (Friedland et al., 1992) and the proliferation of antibiotic-resistant bacteria in the environment following industrial-scale antibiotic production (Knapp et al., 2010). Soil archives take many different forms, from national archives spanning wide regions and often consisting of tens of thousands of samples to smaller archives established by an individual or small group of researchers in relation to a specific project (e.g., Hubbard Brook Sample Archive, 2019; Karssies and Wilson, 2015; Liebig et al. 2008; National Soil Archive, 2019; Rothamsted Research, 2006; Russell Ranch Soil Archive, 2019). Some archives focus on soils from a particular location exposed to various experimental treatments, whereas others focus on natural variation in soils over space, and in both cases archives may span samples collected over years to decades or more. Despite this variety, all soil archives share commonalities, such as recognition that (i) physical samples can be extremely valuable even if we don't know how they will be used in the future; (ii) technological progress,

Soil Sci. Soc. Am. J. 83:973–981
doi:10.2136/sssaj2019.02.0050
Received 20 Feb. 2019.
Accepted 30 Apr. 2019.
*Corresponding author (eayres@battelleEcology.org).

such as new analytical methods, will allow ever more information to be generated from samples; (iii) the world is changing, often making such samples irreplaceable references for conditions at a specific place and time; and (iv) sample collection is often one of the most expensive components of projects, particularly ones encompassing wide geographical areas and deeper soil horizons, and thus archiving samples can be a cost-effective approach to facilitate scientific progress.

The National Ecological Observatory Network (NEON) is a continental-scale network of terrestrial and aquatic research sites throughout the United States where environmental variables are monitored using standardized methods and data are made freely available. In addition to data, large numbers of physical samples, including beetles, mosquitos, ticks, fish, plants, DNA extracts, and soils, are archived and made available for research. The Megapit Soil Archive (hereafter the archive) comprises one component of this collection and consists of soil collected from a single temporary soil pit at each of the 47 NEON terrestrial sites. The soil pit was excavated during site construction to (i) inform soil sensor installation depths, (ii) characterize soil properties (NEON data products: DP1.10066.001, DP1.00096.001, and DP1.00097.001), (iii) collect intact samples to generate soil-specific soil moisture sensor calibrations (Roberti et al., 2018), and (iv) collect soil for archiving. The archive contains fewer samples (n = 335) than most soil archives (median, 620; Nave [2015]), but the quantity of soil stored per sample (1.2–3.6 kg) is greater than most other archives (median, 0.15 kg; Nave [2015]). The first samples were added to the archive in 2012, and the first request for samples was received in 2013. Since then, over 1500 samples have been sent out from the archive (Fig. 1). Samples can be requested at https://www.neonscience.org/request-megapit-soil-samples.

Although some reports have addressed practical aspects of establishing a soil archive, including samples preparation (air-dried or frozen, sieved or ground), storage containers, and labeling (Boone et al., 1999; Nave, 2015), there is limited quantitative information on other basic elements of archive design, such as how much material to archive, which sites and soil horizons to prioritize, which methods of promoting archive use are most successful, and how to balance requests for samples received today with the preservation of samples for future research. Boone et al. (1999) suggested archiving a minimum of a few hundred grams per sample but noted that the quantity depends on unknowns, including the expected number of requests and the quantity needed for the proposed analyses. A summary of responses to a soil archiving questionnaire found that 77% of respondents (n = 39) archived <500 g soil per sample, with 54% archiving <200 g (Nave, 2015), indicating that most archives do not meet Boone et al.'s (1999) suggestion. I am not aware of any recommendations on whether, and by how much, to vary the quantity of soil archived for different sites or horizons, nor the most effective approaches for generating requests for archived soils (e.g., website, conference presentations, list-serv emails).

Here I provide a case study of the Megapit Soil Archive with the hope of providing quantitative guidance on the establishment and operation of a soil archive. Although the results are inevitably specific to the NEON Megapit Archive, the findings are expected to be informative for other archives. Specifically I (i) evaluate the effectiveness of different outreach efforts in generating requests, (ii) quantify the amount of soil consumed per sample per year and evaluate how it varies by site and horizon, and (iii) evaluate the output of projects requesting soil samples. The importance of setting a desired sample lifespan to inform soil archive management decisions is discussed.

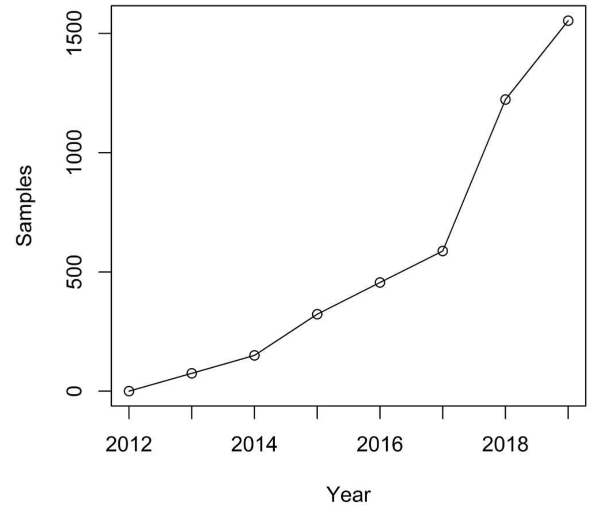

Fig. 1. Cumulative number of samples provided to requesters as of 31 Jan. 2019.

MATERIALS AND METHODS

At each NEON terrestrial site, a temporary soil pit was dug near the meteorological tower in an area representative of the five instrumented soil plots near the tower. A small excavator was used to dig the soil pit (approximately 2 × 2 m wide) at each site, except sites with permafrost, where samples were collected by coring. At most sites, the soil pit depth was 2 m (3 m at Alaskan sites) or bedrock/restrictive feature, whichever was shallower. A detailed pedon description was generated by a US Department of Agriculture (USDA) Natural Resource Conservation Service (NRCS) soil scientist familiar with the local soils, and soil samples were collected by horizon for archiving (Ayres et al., 2016). Additional samples were sent for analysis of various soil physical and chemical properties at the NRCS Kellogg Soil Survey Laboratory (see NEON data products DP1.00096.001 and DP1.00097.001), root biomass distribution was determined (see data product DP1.10066.001), and soil-specific soil moisture sensor calibrations were generated (Roberti et al., 2018). Pedon descriptions and photos are available at https://data.neonscience.org/megapit-info.

Soil for archiving was air-dried in large metal trays (stainless steel or aluminum) and were covered with parchment paper to minimize the risk of contamination. Mineral soils were sieved (2 mm mesh), whereas organic horizons were sorted by hand to remove rocks and undecomposed plant material (Ayres and Durden, 2017). Each sample was given a unique identifier. Sample metadata, including latitude, longitude, horizon, depth increment, and a wide range of soil properties, are available in the megapit soil physical properties (DP1.00096.001) and soil chemical properties (DP1.00097.001) data products available from NEON. Initially 1.2 kg air dried soil was archived from each horizon. However, this was increased to 3.6 kg at later sites due to requests for larger-than-expected quantities of soil. In cases where insufficient soil was collected to meet this quantity, all soil that was collected was archived. The soil was split using a riffle chute splitter and stored in at least four glass jars to minimize the risk of losing all soil from a single horizon (e.g., due to contamination). The jars were stored in fire-resistant lockable cabinets at ambient room temperature (~21°C) in Boulder, CO. A subsample of soil from each archived sample was collected using the subsampling procedure described below and finely ground using a ball mill (8000D Mixer/Mill, Spex SamplePrep) to allow well-mixed representative subsamples to be generated for requests of ≤1 g soil per sample. Soil from each horizon at all 47 NEON terrestrial sites had been added to the archive by 2018.

Requests for Megapit Soil Archive samples can be submitted at https://www.neonscience.org/request-megapit-soil-samples. The request form requires submission of contact information; relevant permits to receive regulated soils; a project description, including justification for the quantity of soil requested; and consent agreement to (i) report publications, presentations, and derived products to NEON, (ii) acknowledge the archive in papers and presentations, and (iii) submit data generated from the samples to NEON; as well as a response to the question "How did you hear about NEON's mega-pit soil archive?" Requests for ≤20 g soil per sample are evaluated internally by NEON staff. Requests for greater quantities of soil are also evaluated by NEON's Soil Sensor Technical Working Group, which includes non-NEON researchers with an interest in soils.

Once a request was approved, the relevant jar containing the sample was identified by its label and weighed. The weight was compared with the weight recorded after the last subsampling event as an additional quality control check to ensure that the correct jar was subsampled. The jar was rotated around its long axis and end-over-end to homogenize the soil, and the subsample was removed using a spatula. Scoop sampling was used to generate subsamples despite lower representativeness during individual subsampling events (Mullins and Hutchison, 1982; Particle Sciences, 2011) because a preliminary test to generate subsample "3_Ordway_Pit1_28–125_TIS Soil Archive_Bw2_1_00001" using a chute splitter indicated that 0.21% of the archived sample mass, primarily comprised of fine particles, was lost during the subsampling event, versus <0.01% mass loss comprised of a range of particle sizes with scoop sampling. The cumulative loss of fine material during repeated subsampling with the chute splitter was expected to result in lower representativeness than scoop sampling over the lifespan of the archived soils. The weight of the jar and the weight of soil removed for the requester were recorded before and after subsampling, and the weights were checked to ensure all soil was accounted for before returning the jar to the archive cabinets (Ayres, 2017). The weights, unique subsample ID, date the subsample was sent to the requester, and requester information are all associated with the unique archive sample ID so that a complete history of each archived sample was maintained. Boone et al. (1999) noted that subsampling archived soils is wasteful because people often request more soil than they require. Instead, they advocate for sending the entire sample to a requester so that only the quantity required is removed. This approach was not taken at the Megapit Soil Archive due to the reduction in quality assurance that can be provided for samples that have changed hands many times and the risk of sample loss or contamination. Returning subsamples to the archive after nondestructive analyses is not permitted for the same reasons. Therefore, every subsampling event results in the permanent loss of soil from the archive.

Various approaches have been taken to increase awareness of the archive among the potential user community with the aim of generating requests for samples. This has included presentations at targeted conferences (Ayres, 2013; Ayres et al., 2014, 2015; SanClements, 2019), flyers, periodic emails to relevant communities (e.g., Soil Science Society of America Divisions and the International Soil Carbon Network), and an archive website. The effectiveness of these approaches was evaluated by investigating the proportion of users that responded to the question "How did you hear about NEON's mega-pit soil archive?", where one of the following responses can be selected: NEON Presentation, Colleague, NEON Staff, List-serv email, Soil archive flyer or postcard, Web search, or Other. Response to this question is required to submit a request for samples. People who have not yet submitted a request for samples but have informally inquired about the archive were also asked this question, and in situations where they did not respond it was recorded as "Unknown." It is likely that some expressions of interest in the archive that did not result in formal requests went undocumented.

For the purpose of this study, documented expressions of interest and formal requests received over a 6-yr period between 2013 and 2018 were analyzed, as were the 1554 subsamples that were prepared for the approved requests received during this period. Parameters investigated include the sites and soil horizons requested, the number of requests per sample per year, the quantity of soil requested, and output of projects requesting samples. All analyses were performed in the R statistical language, version 3.4.3 (R Foundation for Statistical Computing, 2017). The lm() function (linear regression) was used to test for a significant relationship between year and number of expressions of interest and formal requests received to see if the interest in the archive changed over time. The requests appeared to show a bias toward soils from A horizons and NEON Core sites (see below); as a result, the 1554 subsamples were partitioned into A versus non-A horizons and into Core versus Relocatable sites for analysis. The primary difference between NEON Core and Relocatable sites is the expected duration of monitoring: Core sites will be monitored continuously over NEON's expected 30-yr lifespan, whereas Relocatable sites may move to a different location after several years (Schimel et al., 2009). The data collected from these two

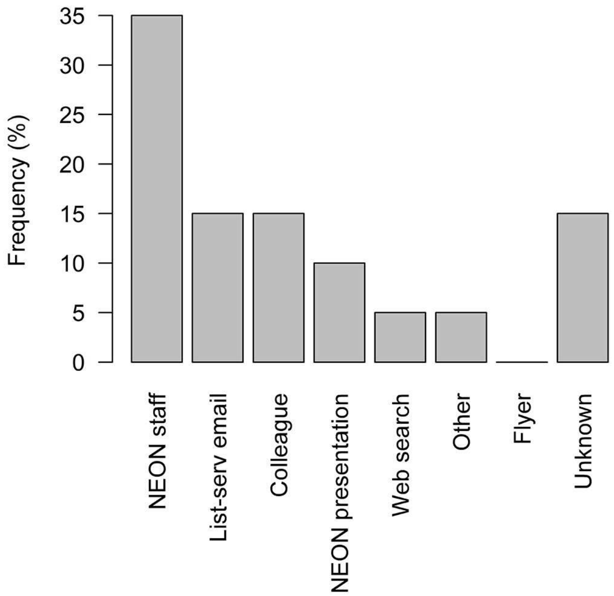

Fig. 2. Responses (n = 20) to the question: How did you hear about the archive?

types of sites are almost identical, and the data generated from the soil pits where the archived soils were collected were identical. In addition, the Core sites tend to have a long history of ecological research, whereas this is less common, albeit not unusual, at Relocatable sites. Shapiro–Wilk tests showed that the quantity of soil consumed (*i.e.*, sent to requesters) per sample, the number of requests per sample per year, and the quantity of soil requested per sample per request were all non-normally distributed ($P < 0.05$). As a result, nonparametric Kruskal–Wallis tests [kruskal.test() function] were used to test for significant differences in these variables for all combinations of Core and Relocatable sites and A and non-A soil horizons for the 1554 subsamples prepared for requesters. In cases where the test was significant ($P < 0.05$), multiple pairwise-comparisons were performed using the pairwise.wilcox.test() function with the Benjamini and Hochberg (1995) correction for multiple comparisons. Anonymized data used in this study are available at https://data.neonscience.org/prototype-search (National Ecological Observatory Network, 2019).

RESULTS

Of the 20 people expressing an interest in the archive between 2013 and 2018, 60% heard about the archive as a result of active promotion of the archive by NEON staff (e.g., informal interactions at conferences, list-serv emails, and conference presentations), with an additional 15% hearing via word of mouth from a colleague (Fig. 2). In contrast, passive promotion of the archive (i.e., flyers and a website that can be found via a web search) was only cited in one instance, highlighting the importance of directly interacting with the target user community.

Of the 20 expressions of interest, 14 (70%) resulted in formal requests for samples, which equates to a mean ± SD of 2.3 ± 1.0 formal requests per year for the period 2013 to 2018. There was a tendency toward larger numbers of expressions of interest

Table 1. Expressions of interest and formal request for samples received between 2013 and 2018.

Year	Expressions of interest	Requests for samples	Approved requests
2013	2	1	1
2014	3	3	2
2015	5	3	2
2016	2	1	1
2017	5	3	3
2018	3	3	3

and requests in later years, but the relationship was not significant ($P > 0.1$) (Table 1).

The mean ± SD quantity of soil requested was 58 ± 129 g per requested sample (minimum, 1 g; maximum, 500 g), with half of the requests being for 20 g or less (Table 2). One request was declined because the quantity of soil requested was considered too large (500 g per requested sample) relative to the quantity archived (1200–3600 g per sample) and the scientific merit of the project, and another request has been awaiting a response from the requester for >3 mo and is considered to have been withdrawn. The remaining 12 requests were approved. Considering only the 12 requests that were approved (mean ± SD, 2.0 ± 0.9 requests per year), 23 ± 19 g soil was requested per requested sample (minimum, 1 g; maximum, 50 g).

Six of the 12 approved requests asked for soils from all 47 NEON terrestrial sites, with the remaining six requests asking for soils from between 1 and 20 sites (Table 2). The mean ± SD number of sites per request was 30 ± 19, representing 64 ± 40% of NEON terrestrial sites. Core sites comprised 60% of sites in requests for a subset of sites despite accounting for only 43% of NEON terrestrial sites. Four of 12 requests were for soils from all soil horizons, with one additional request for all mineral soil horizons. Of the other requests, A horizon soils were by far the most requested (5 of 12 with an additional request for O and A horizon soils).

The mean amount of soil consumed per archived sample (i.e., sent to requesters) was lower than the mean amount requested per requested sample because some requests were only for a subset of soil samples. Across all samples in the archive, the mean ± SD quantity of soil consumed was 14.6 ± 4.3 g per sample yr^{-1}. However, consistent with the requests (Table 2), significantly more soil was consumed from A horizon samples ($P < 0.001$), and soils from A horizons at Core sites were consumed most rapidly (18.2 ± 7.2 g yr^{-1}); (Fig. 3a), which approached rates significantly greater than A horizon soils from Relocatable sites ($P = 0.08$). Of the 10% of samples with the fastest consumption rates (i.e., > 17.9 g yr^{-1}), 74% were A horizon soils despite comprising only 20% of archived samples, and 62% were from NEON Core sites despite comprising only 41% of archived samples. The higher consumption rate of A horizon soils resulted from 52% more requests than for non-A horizon soils ($P < 0.001$) (Fig. 3b), which outweighed a 19% reduction in the weight requested per request ($P < 0.001$) (Fig. 3c). Site type did not significantly affect the number of requests (Fig. 3b), but significantly more soil

Table 2. Formal requests (n = 14) for samples from the archive received between 2013 and 2018.

Sites†	Horizons	Quantity requested	Request year	Request decision
		g		
All	all	30	2013	approved
All	all	1	2014	approved
All	all	50	2014	approved
All	all	35	2014	awaiting information (>3 mo since last response)
BLAN	Oi	50	2015	approved
HARV, **OSBS**, **ORNL**, **TALL**, **CPER**, JORN	A (shallowest subhorizon)	50	2015	approved
All	0–15 cm	500	2015	declined
HARV, BLAN, **SCBI**, **OSBS**, JERC, DSNY, DELA, **UNDE**, **KONZ**, UKFS, DCFS, **WOOD**, NOGP, **CPER**, STER, RMNP, OAES, **NIWO**, JORN, **ONAQ**	all mineral	31	2016	approved
All	O and A (all subhorizons)	2	2017	approved
All	A (shallowest subhorizon)	3	2017	approved
HARV, **SCBI**, **OSBS**, **GUAN**, **UNDE**, **KONZ**, **TALL**, **WOOD**, **CLBJ**, **NIWO**, **SRER**, **ONAQ**	A (shallowest subhorizon)	15	2017	approved
All	All	20	2018	approved
HARV, **SCBI**, DSNY, **OSBS**, **GUAN**, **UNDE**, **KONZ**, GRSM, LENO, **WOOD**, **CPER**, **NIWO**, **SRER**, **ONAQ**, BART, JERC, DCFS, **CLBJ**, MOAB, JORN	A (shallowest subhorizon)	10	2018	approved
RMNP, LAJA, UKFS, MLBS, **ORNL**, **TALL**, ABBY, **WREF**, **SJER**, SOAP, TEAK, **TOOL**, **BONA**, HEAL	A (shallowest subhorizon; B at TOOL & BONA as A horizon not available)	10	2018	approved

† NEON site code. Core sites are in bold text.

per request was requested from non-A horizons at Relocatable sites than from Core sites (Fig. 3c).

Given that the first request for samples was received in 2013, there has been relatively little time for the requesters to complete their measurements, analyze data, generate conclusions, and disseminate findings, especially for more recent requests. Nonetheless, two peer-reviewed papers using samples from the archive have been published (Ghabbour et al., 2015; Kramer and Chadwick, 2018), and two presentations have been given at conferences (Abney et al., 2018; Davies and Ghabbour, 2015). These publications and presentations come from three separate requests, representing 25% of approved requests.

DISCUSSION

Because the data presented here come from an archive established for a continental-scale research network, the results are likely most relevant to archives established by other large networks and organizations. NEON's large size likely increases awareness of components of the Observatory, such as the archive, which may result in a larger number of requests per sample than experienced by other soil archives. Nonetheless, there are aspects of the results that are likely relevant to most soil archives, regardless of size and focus.

For instance, the results clearly show that active outreach efforts by NEON staff (e.g., community interactions, presentations, and list-serv emails) were essential to raise awareness of the archive and to generate requests. Although presentations and list-serv emails played a role in generating awareness, contact with NEON staff (presumably at conferences, workshops, field sites, and other venues) was the most cited approach for raising awareness, highlighting the importance of informal community engagement. Word of mouth from colleagues also played a significant role in increasing awareness, but, based on these results, it seems reasonable to assume that the colleague heard about the archive via active outreach by NEON. In contrast, only one person (5%) heard about the archive via a web search, suggesting that, although an archive website is probably essential, it is unlikely to generate much awareness among the potential user community.

Across all samples in the archive, an average of 14.6 ± 4.3 g soil was sent to requesters per sample yr^{-1}, which could inform the quantity of soil that new archives need to store if the expected number of requests and request size is unknown. For example, to ensure that almost no samples were exhausted earlier than planned, one could conservatively choose to archive 3 SD above this weight (i.e., 27.5 g soil per sample) for each year of the archived sample's desired lifespan, assuming a similar consumption rate to NEON's Megapit Soil Archive. In this example, 3 SD was used because it gives a high level of confidence that the archived soil will not be exhausted earlier than expected. Based on this consumption rate, 500 g archived soil would be sufficient to last at least 18 yr (i.e., 500/27.5) for the vast majority of samples and ~50% of samples would have been exhausted after 34 yr (500/14.6). Responses to a survey of soil archives indicated that the median mass archived was 150 g per sample (Nave, 2015), which implies that samples would start to be exhausted after ~5 yr (i.e., 150/27.5) and around 50% of samples would be exhausted within ~10 yr (i.e., 150/14.6) if they experienced the same consumption rate as the Megapit Soil

Archive. In practice, some of these archives may experience slower consumption rates, increasing their lifespan.

It is notable that the consumption rate described above excludes the request for 500 g per sample that was declined due to the proportion of archived material that would have been consumed, but had this been approved the consumption rate would have been substantially greater. Given that only 8% of archives surveyed stored ≥1000 g soil and none stored >2400 g (Nave, 2015), such large requests are unlikely to be fulfilled by most archives, including the Megapit Soil Archive.

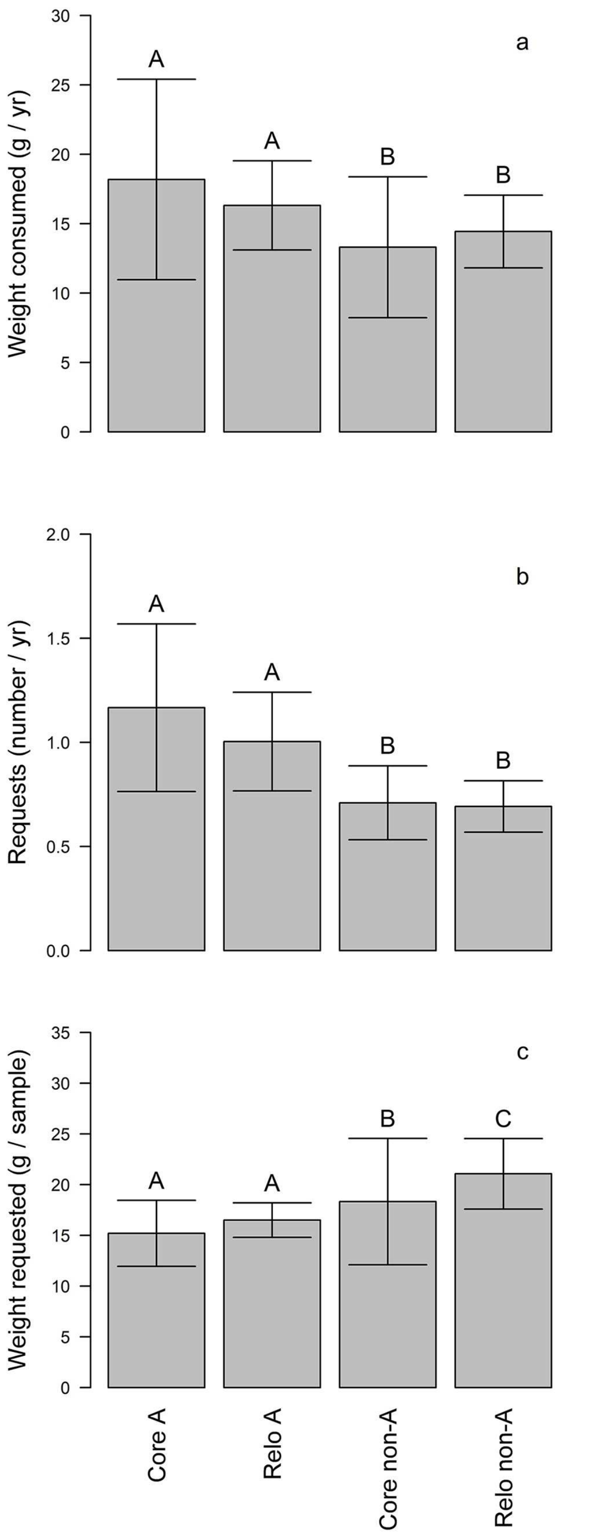

Fig. 3. Mean ± 1 SD quantity of soil consumed (a), number of requests per year (b), and soil weight per sample per request (c) for A horizon soils at Core sites (Core A; *n* = 24), A horizons at Relocatable sites (Relo A; *n* = 44), non-A horizons at Core sites (Core non-A; *n* = 112), and non-A horizons at Relocatable sites (Relo non-A; *n* = 155). Different uppercase letters denote significant differences (*P* < 0.05).

The single most common approved request (33%) was for soil from all sites and horizons. The remaining requests were for soils from a subset of sites and/or horizons, which resulted in different consumption rates among the archived samples. In particular, soils from A horizons were consumed more rapidly, and there was a tendency for soils from Core sites to be consumed faster as well. This suggests that archiving larger quantities of soil from A horizons, and perhaps more intensively studied sites, may be warranted. Using this information, a simple equation can be created to estimate the quantity of soil required for archiving:

$$Q = [C_{\mu} + (C_{\sigma} \times B)]L \quad [1]$$

where Q is the quantity of soil necessary to archive per sample (g), C_{μ} is the mean consumption rate (18.2, 16.3, 13.3, and 14.4 g yr^{-1} for A horizons at Core sites, A horizons at Relocatable sites, non-A horizons at Core sites, and non-A horizons at Relocatable sites, respectively, in this study), C_{σ} is the consumption rate SD (7.2, 3.2, 5.1, and 2.6 g yr^{-1} for A horizons at Core sites, A horizons at Relocatable sites, non-A horizons at Core sites, and non-A horizons at Relocatable sites, respectively, in this study), B is the desired level of confidence that the sample will last L years (unitless)—3 times SD (i.e., ≥89% confidence for non-normal distributions based on Chebyshev's inequality [Markov, 1884]) in this study, and L is the desired lifespan of the archived sample (yr) (36 yr in this study [2013–2049] based on the archive's establishment and NEON's expected end date).

Based on Eq. [1] and the parameters specified above, between 22.3 and 39.9 g soil, depending on site type and horizon type, would be needed for each year of archiving to have ≥89% confidence (i.e., 3 SD for non-normally distributed data based on Chebyshev's inequality [Markov, 1884]) that the sample would not be exhausted prematurely (Fig. 4). With this consumption rate, 500 g archived soil would last at least 12, 19, 17, and 22 yr for A horizons at Core sites, A horizons at Relocatable sites, non-A horizons at Core sites, and non-A horizons at Relocatable sites, respectively, for the vast majority of samples. Conversely, to ensure the same proportion of samples reached their desired lifespan regardless of horizon and site type, the amount of soil archived from A horizons at Core sites, A horizons at Relocatable sites, and non-A horizons at Core sites would need to be increased by 79, 16, and 28%, respectively, relative to non-A horizons at Relocatable sites.

Using these values in relation to the Megapit Soil Archive, NEON should have archived 1435, 934, 1026, and 803 g soil per horizon for A horizons at Core sites, A horizons at Relocatable sites, non-A horizons at Core sites, and non-A horizons at Relocatable sites, respectively. These values are less than the quantity that was archived for most horizons (1200 g for early sites and 3600 g for later sites), which suggests that, assuming the current consumption rate remains unchanged, it is likely that soil from most horizons will not be exhausted prior to NEON's planned end date (2049), with the possible exception of a few A horizon soils from the early Core sites. It is unknown if the current consumption rate will change over time. Although it is possible that the quantity of soil requested per sample will remain relatively unchanged over the lifespan of the archive given that most soil properties can be measured on 1 to 50 g soil (i.e., the range requested to date), it is conceivable that that the number of requests per year may change. For example, awareness of the archive among the potential user community may increase as more studies are published based on data from the samples, and the samples may become more attractive to researchers as the amount of information known about each sample increases.

Because soil from all NEON terrestrial sites have already been added to the archive, the estimated quantities of soil required for archiving based on Eq. [1] will not be implemented at this time. However, if new NEON sites are established in the future, these values will be recalculated based on the latest consumption rates to maximize resource use efficiency. Such a change would have little impact on the field collection and laboratory processing of the samples prior to archiving, with the exception that the volume of field moist soil collected per horizon could be adjusted. The current field collection protocol specifies that ~12 L of soil be collected from mineral horizons and ~36 L from organic horizons for archiving (Ayres et al., 2016); however, based on the values from Eq. [1], this could be reduced to ~5 L from mineral horizons and ~11 L from organic horizons. These volumes of field moist soil are expected to result in the specified weight of air-dried soil for archiving, even for soils with relatively low bulk density or high rock content.

Although the archive is still relatively young, 25% of approved requests have already produced at least one known publication or presentation. It is possible that this value is underestimated given that not all products resulting from a request may be reported to the archive (although requesters agree to report all resulting publications and presentations when submitting a request). Notably, the two publications come from the first and third requests that were approved, suggesting that the other requesters may not have yet had sufficient time to complete their analyses and have a publication accepted. Although it is impossible to say whether the findings from these studies would have been made without the archive, it is difficult to imagine that they could have been achieved in such a cost-effective manner given the expense involved in characterizing and collecting soils to 2 m deep at 47 (often remote) sites throughout the United States, including Alaska, Hawaii, and Puerto Rico.

There is an inherent tension between promoting the use of archived samples, especially when many analyses are destructive, and retaining samples for future use. Archived samples that are never used are a waste of resources, but samples that are exhausted too quickly have an opportunity cost due to the loss of information that may come from future analytical advances or

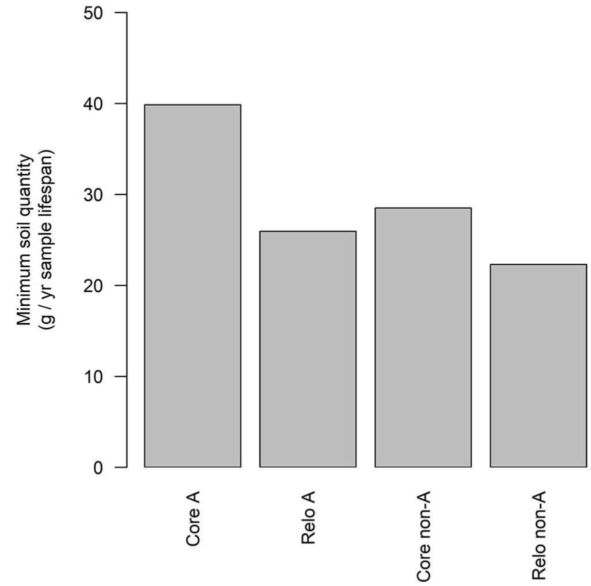

Fig. 4. Minimum quantity of soil required for each year of archiving to ensure with ≥89% confidence (i.e., mean consumption rate plus 3 SD for non-normal distributions) that samples are not prematurely exhausted for A horizon soils at Core sites (Core A), A horizons at Relocatable sites (Relo A), non-A horizons at Core sites (Core non-A), and non-A horizons at Relocatable sites (Relo non-A).

novel research ideas. Without a clear framework, archive management decisions, such as the level of effort devoted to raising awareness of the archive or how many requests to fulfill, may be prone to arbitrariness and be difficult to justify. However, this can be mitigated by setting a desired lifespan for each archived sample, as demonstrated below.

NEON is expected to operate until 2049, which provides a clear longevity goal for the archive and indicates that ideally all samples should be exhausted by, but no sooner than, 2049 to maximize the benefits generated by the archive. Based on current consumption rates, 334 of 335 samples in the archive are expected to contain some remaining material by 2049 (Fig. 5), indicating that the archive is being underutilized and should aim to increase consumption rates (i.e., send out more soil to requesters). Given that all feasible formal requests for samples have been approved to date (Table 2), there is currently no justification for loosening the request evaluation criteria to increase sample consumption rates. As a result, the only option to increase consumption rates is to increase the amount of soil requested each year, suggesting that any available resources should be directed to raising awareness of the archive among the potential user community, which can be best achieved by active outreach by NEON (Fig. 2). As demonstrated by this example, specifying a desired lifespan for archived samples, in conjunction with other archive data, simplifies the decision-making process and results in a clear justification that can be articulated to archive funders, project managers, community stakeholders, and others. If the archived samples were being depleted too rapidly, this decision-making process would have suggested tightening the request evaluation criteria and/or reducing outreach activities aimed at generating requests. This decision-making process is dynamic and can be changed at any time based

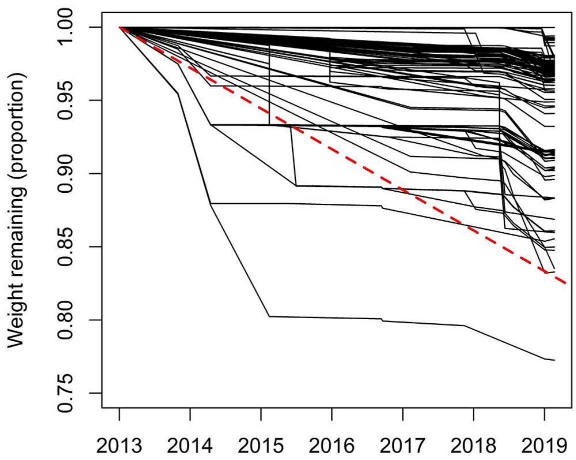

Fig. 5. Proportion of soil remaining for each sample in the archive. Each line represents soil from an individual archived sample (*n* = 335). The dashed line represents a constant consumption rate that results in sample exhaustion in 2049 (NEON's expected end date).

on new data (i.e., changes in consumption rates or even a new desired lifespan) and can be tailored to individual samples (e.g., tightening the evaluation criteria for the one sample that is currently being depleted too fast to survive until 2049). Last, in this example sample consumption rates were compared with a modeled constant consumption rate that results in sample exhaustion by its desired lifespan. However, if near-term research findings were more highly valued than findings generated in the more distant future (i.e., temporal discounting), the modeled consumption rate could be relatively high initially and gradually decrease over time. Conversely, if there was a desire to save archived soil for future researchers, the modeled consumption rate could be relatively low initially and gradually increase over time.

Although most soil archives already track sample consumption rates (at least informally), I believe that most do not specify a desired lifespan for their samples; as a result, it may be difficult for archive managers to maximize the benefits of these archives and provide a justification for their decisions. Although everybody would agree that specifying a lifespan for an archived soil sample is neither trivial nor even has one "correct" value, it allows for a quantitative and consistent approach to inform management decisions and optimize resource use (e.g., money, time, effort, and space), which is preferable to making these decisions based on judgement alone.

The results from this study provide a quantitative basis to inform the establishment and operation of soil archives. Although these finding are expected to be broadly applicable, they are to some extent specific to the NEON Megapit Soil Archive. As a result, other soil archives are encouraged to publish similar data based on their usage and operation to provide a diverse suite of quantitative information that can be used to maximize the benefits provided by valuable, often irreplaceable, archived soil samples.

ACKNOWLEDGMENTS

Numerous researchers at the field sites provided valuable assistance with selecting the soil pit location and other aspects of the NEON site design. In addition, I'm very grateful for support from USDA–NRCS soil scientists, especially Larry West, Jon Hempel, Doug Wysocki, Ellis Benham, Scarlett Bailey, Rich Ferguson, Steve Monteith, and the many others who assisted with the soil pit location selection, performed soil descriptions, and analyzed soil physical and chemical properties and to Rommel Zulueta, David Durden, Josh Roberti, Robert Lee, Derek Smith, Holly Abercrombie, Jennifer Everhart, Cody Flagg, Maheteme Gebremedhin, Alison Hogeboom, Rachel Krauss, Courtney Meier, Mike Patterson, Aaron Rothchild, Michael SanClements, Oliver Smith, Julia Spencer, and Margaret Treese for field and lab support. Three anonymous reviewers provided valuable comments that improved this paper. NEON is sponsored by the US National Science Foundation (NSF) and managed under cooperative agreement by Battelle. This material is based on work supported by the NSF (grant DBI-1638696; 2016-2020). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

REFERENCES

Abney, R., D.L. Bish, and J.D. Raff. 2018. Soil organic matter and mineralogy as controls on NO_2 deposition and HONO emissions from soil surfaces. Paper presented at: American Geophysical Union Fall Meeting, Washington, DC. 10–14 Dec. 2018.

Ayres, E. 2013. The NEON soil archive: A community resource. Paper presented at: Soil Ecology Society 13th Biennial Meeting, Camden, NJ. 11–14 June 2013.

Ayres, E. 2017. NEON.DOC.001306 NEON procedure and protocol: Producing TIS soil archive subsamples for users, Version C. National Ecological Observatory Network, Boulder, CO. https://data.neonscience.org/documents (accessed 20 Feb. 2019).

Ayres, E., M. Denslow, R. Zulueta, D. Durden, J. Roberti, and D. Smith. 2014. The NEON soil archive: A community resource. Paper presented at: Ecological Society of America annual meeting, Sacramento, CA. 10–15 Aug. 2014.

Ayres, E., and D. Durden. 2017. NEON.DOC.000325 field and lab procedure and protocol: TIS soil archiving, version F. National Ecological Observatory Network, Boulder, CO. https://data.neonscience.org/documents (accessed 20 Feb. 2019).

Ayres, E., R. Zulueta, D. Durden, J. Roberti, and D. Smith. 2015. The NEON soil archive: A community resource. Paper presented at: National Cooperative Soil Survey National Conference, Duluth, MN. 7–11 June 2015.

Ayres, E., R. Zulueta, D. Smith, and J. Roberti. 2016. NEON.DOC.001307 TIS soil pit sampling protocol, version C. National Ecological Observatory Network, Boulder, CO. https://data.neonscience.org/documents (accessed 20 Feb. 2019).

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57:289–300.

Boone, R.D., D.F. Grigal, P. Sollins, R.J. Ahrens, and D.E. Armstrong. 1999. Soil sampling, preparation, archiving, and quality control. In: G.P. Robertson, D.C. Coleman, C.S. Bledsoe, and P. Sollins, editors, Standard soil methods for long-term ecological research. Oxford Univ. Press, Oxford, UK. p. 3–28.

Cary, S.C., and N. Fierer. 2014. The importance of sample archiving in microbial ecology. Nat. Rev. Microbiol. 12:789–790. doi:10.1038/nrmicro3382

Davies, G., and E.A. Ghabbour. 2015. Measuring sequestered carbon contents of grassland and forest soil profiles in the NEON initiative. Paper presented at: National Cooperative Soil Survey National Conference, Duluth, MN. 7–11 June 2015.

Dolfing, J., and Y. Feng. 2015. The importance of soil archive for microbial ecology. Nat. Rev. Microbiol. doi:10.1038/nrmicro3382-c1

Friedland, A.J., B.W. Craig, E.K. Miller, G.T. Herrick, T.G. Siccama, and A.H. Johnson. 1992. Decreasing lead levels in the forest floor of the Northeastern USA. Ambio 21:400–403.

Ghabbour, E.A., G. Davies, A.A. Sayeed, M.T. Croman, B.A. Hoehing, and E. Ayres. 2015. Measuring the total and sequestered organic matter contents of grassland and forest soil profiles in the National Ecological Observatory Network Initiative. Soil Horiz. 56. doi:10.2136/sh15-07-0014

Hubbard Brook Sample Archive. 2019. https://hubbardbrook.org/d/sample-archive-overview (accessed 14 Jan. 2019).

Karssies, L., and P. Wilson. 2015. Soil archive: Supporting research into soil changes. IOP Conf. Ser.: Earth Environ. Sci. 25:012021. doi:10.1088/1755-1315/25/1/012021

Knapp, C.W., J. Dolfing, P.A.I. Ehlert, and D.W. Graham. 2010. Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. Environ. Sci. Technol. 44:580–587. doi:10.1021/es901221x

Kramer, M.G., and O.A. Chadwick. 2018. Climate-driven thresholds in reactive mineral retention of soil carbon at the global scale. Nat. Clim. Chang. 8:1104–1108. doi:10.1038/s41558-018-0341-4

Liebig, M.A., D.J. Wikenheiser, and K.A. Nichols. 2008. Opportunities to utilize the USDA-ARS Northern Great Plains Research Laboratory soil sample archive. Soil Sci. Soc. Am. J. 72:975–977. doi:10.2136/sssaj2007.0324N

Manter, D.K., J.A. Delgado, H.D. Blackburn, D. Harmel, A.A. Pérez de León, and C.W. Honeycutt. 2017. Why we need a national living soils repository. Proc. Natl. Acad. Sci. USA 114:13587–13590. doi:10.1073/pnas.1720262115

Markov, A. 1884. On certain applications of algebraic continued fractions. Ph.D. thesis, St. Petersburg.

Mullins, C.E., and B.J. Hutchison. 1982. The variability introduced by various subsampling techniques. Eur. J. Soil Sci. 33:547–561. doi:10.1111/j.1365-2389.1982.tb01788.x

National Ecological Observatory Network. 2019. Megapit Soil Archive use data, 2013-2018. https://data.neonscience.org/prototype-search (accessed 20 Feb. 2019).

National Soil Archive. 2019. https://www.hutton.ac.uk/about/facilities/national-soils-archive (accessed 14 Jan. 2019).

Nave, L. 2015. Summary of archiving practices. International Soil Carbon Network. http://iscn.fluxdata.org/wp-content/uploads/sites/15/ISCN-archiving-summary.pdf (accessed 20 Feb. 2019).

Particle Sciences. 2011. Sampling of powders, Technical Brief 2011, Volume 5. Particle Sciences Inc., Bethlehem, PA. https://www.particlesciences.com/docs/technical_briefs/TB_2011_5.pdf (accessed 20 Feb. 2019).

R Foundation for Statistical Computing. 2017. R version 3.4.3 (2017-11-30)–"Kite-Eating Tree." R Foundation for Statistical Computing, Vienna.

Richter, D.deB., M. Hofmockel, M.A. Callaham, D.S. Powlson, and P. Smith. 2007. Long-term soil experiments: Keys to managing Earth's rapidly changing ecosystems. Soil Sci. Soc. Am. J. 71:266–279. doi:10.2136/sssaj2006.0181

Roberti, J.A., E. Ayres, H.W. Loescher, J. Tang, G. Starr, D.J. Durden, et al. 2018. A robust calibration method for continental-scale soil water content measurements. Vadose Zone J. 17:170–177. doi:10.2136/vzj2017.10.0177

Rothamsted Research. 2006. Guide to the classical and other long-term experiments, datasets and sample archive. Rothamsted Research, Harpenden, UK.

Russell Ranch Soil Archive. 2019. https://asi.ucdavis.edu/programs/rr/research/data/soil-sample-archives (accessed 10 Apr. 2019).

SanClements, M.D. 2019. Accessing NEON's thousands of archived soil samples. Paper presented at: SSSA International Soils Meeting, San Diego, CA. 6–9 Jan. 2019.

Schimel, D., M. Keller, P. Duffy, L. Alves, S. Aulenbach, W. Gram, et al. 2009. The NEON strategy: Enabling continental scale ecological forecasting. National Ecological Observatory Network, Boulder, CO.