# Alaska as a prototype for a national soil carbon ^living database

Kristofer Johnson
USGS-funded postdoc
University of Alaska Fairbanks

National Soil Carbon Database Development Townhall Meeting

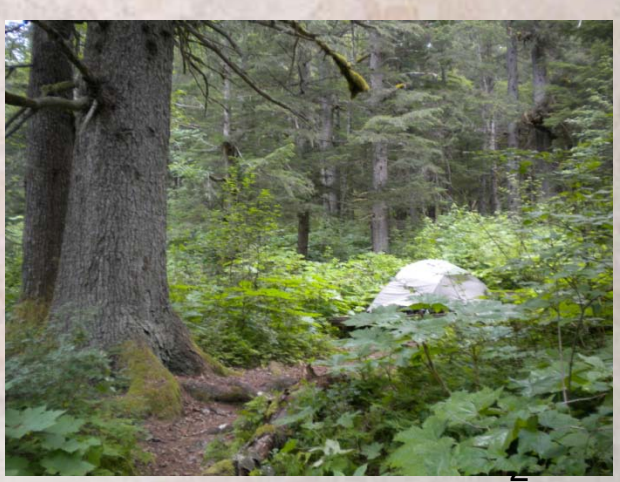
# Overview

1. Introduction – the need for soil carbon databases

2. Building a database – what we learned

3. The data cube

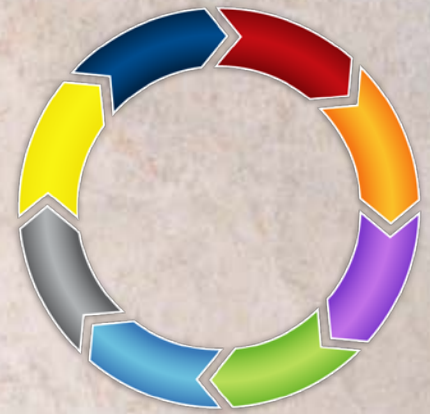
4. How can the database help?



# 1. Introduction

The need for a soil carbon database:

The scientific impetus for understanding the dynamics of soil carbon comes from the important role soils play in the terrestrial carbon cycle.



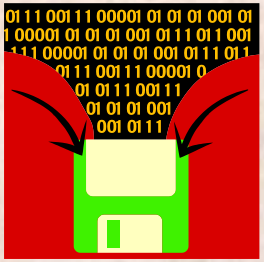
To address this role most effectively, data-sharing activities that strengthen scientific collaboration are needed.

# 1. Introduction

To be an effective tool, a database must be “living”.

This requires:

1. People and their data

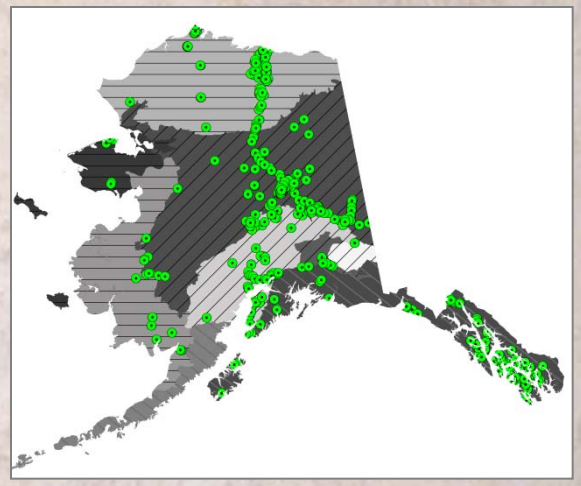


2. Technology

# 2. Building a “living” database

In Alaska, a soil carbon database project funded by the USGS was designed to:

1) promote a stronger soils community in Alaska



2) assess the state of soil carbon in Alaska (i.e. identify gaps)

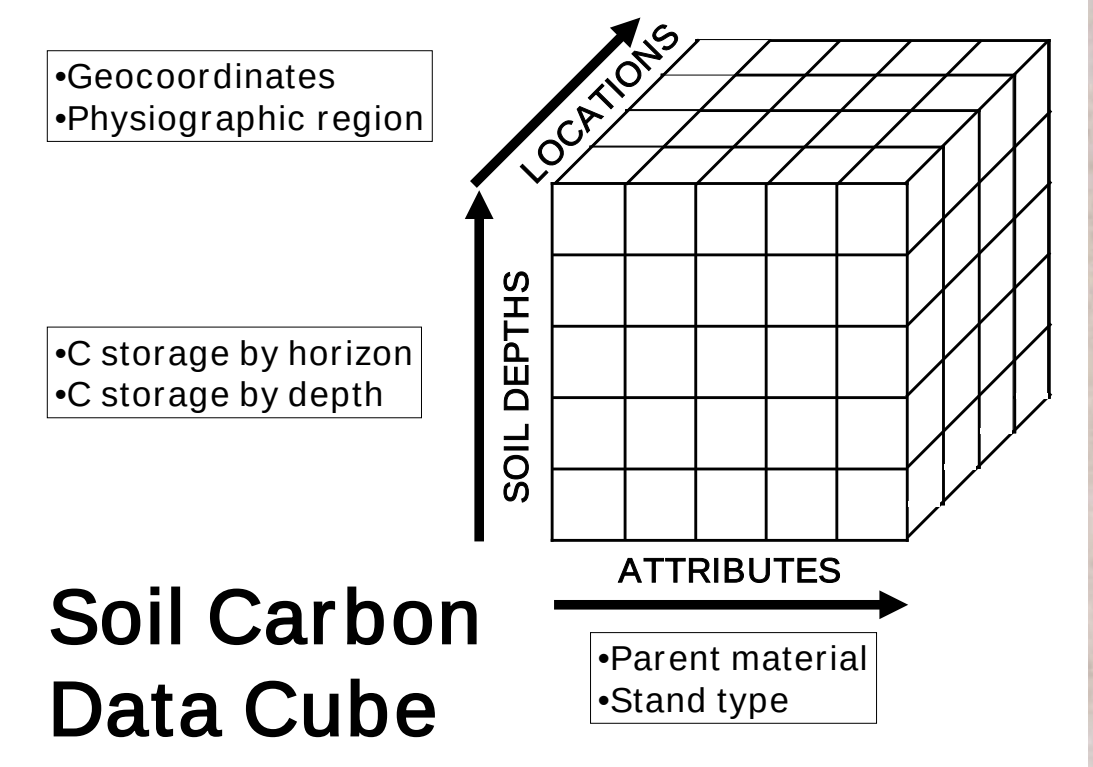
3) develop a “living” database as a tool for soil carbon research

# 3. The Data Cube

Features:

Includes relevant measurements

Time-saver
- Gap-filling
- Data filtering



# 3. The Data Cube

## *Build your own pivot table in excel*

H10 | 19.8600006103515

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | |
| 2 | ContributorOrganization | USGS | | | | | | |
| 3 | LayerIndicator | (Multiple Items) | | | | | | |
| 4 | ProfileIndicator | contiguous | | | | | | |
| 5 | Ecoregion | All | | | | | | |
| 6 | | | | | | | | |
| 7 | | | | | | | Column Labels | |
| 8 | | | | | | | c_tot | |
| 9 | Row Labels | Latitude | Longitude | ProfileId | Sample | HznDesgn | Count | Cumulative Lay |
| 10 | BZEC | 64.700821 | -148.31129 | BZEC 1 | BZEC 1.11 | Oe | 1 | 19.86000061 |
| 11 | | | | | | | | 0999985 |
| 12 | | | | | | | | 0000038 |
| 13 | | | | | | | | 8999977 |
| 14 | | | | | | | | 2000046 |
| 15 | | | | | | | | 11 |
| 16 | | | | | BZEC 1.23 | A | 1 | 9.149999619 |
| 17 | | | | | BZEC 1.25 | A | 1 | 7.199999809 |
| 18 | | | | | BZEC 1.27 | A | 1 | 6.480000019 |
| 19 | | | | | BZEC 1.4 | Oi | 1 | 34.99000168 |
| 20 | | | | | BZEC 1.7 | Oe | 1 | 32.31999969 |
| 21 | | | | | BZEC 1.9 | Oe | 1 | 25.09000015 |
| 22 | | | | | | | | |
| 23 | | | | BZEC 2 | BZEC 2.16 | OA | 1 | 35.45999908 |
| 24 | | | | | BZEC 2.28 | A | 1 | 8.039999962 |
| 25 | | | | | BZEC 2.3 | Oi | 1 | 38.83000183 |
| 26 | | | | | BZEC 2.7 | Oe | 1 | 37.93999863 |
| 27 | | | | | | | | |
| 28 | BZGR | 64.700821 | -148.31129 | BZGR 2 | BZGR 2.10 | Oe | 1 | 36.97999954 |
| 29 | | | | | BZGR 2.12 | Oe | 1 | 33.99000168 |

Sheet1 Sheet2 Sheet3

Ready

**Choose Fields**

PivotTable Field List

Show fields related to:

(All)

- Σ Samples
  - SampleCount
- Σ Sites
  - SiteCount
- Ancdatumtype
  - Name
  - Shortname
- Horizon
  - Hzn
  - HznBound
  - HznDesgn

Drag fields between areas below:

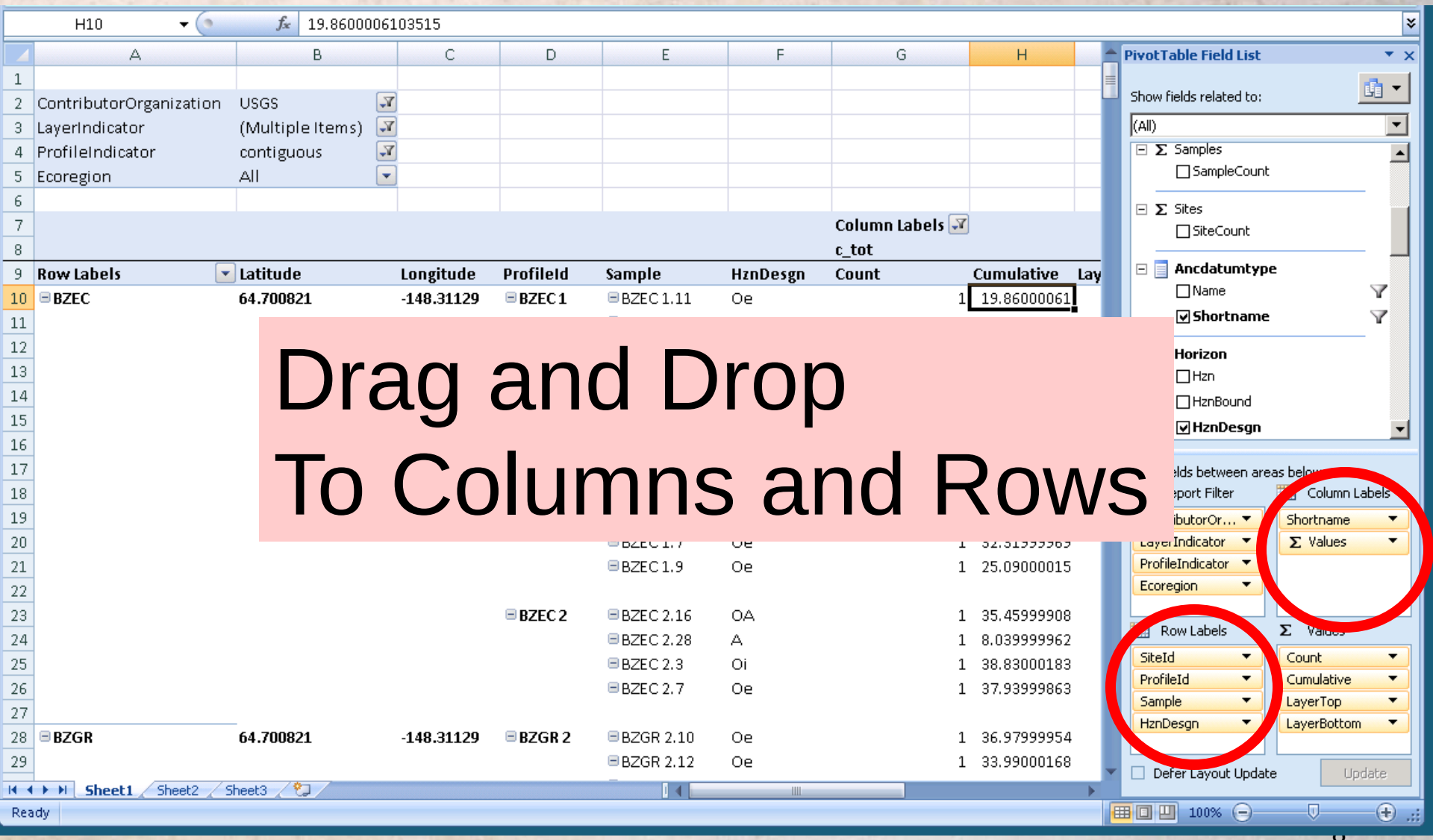
| Report Filter | Column Labels |
|---|---|
| ContributorOr... | Shortname |
| LayerIndicator | Σ Values |
| ProfileIndicator | |
| Ecoregion | |

| Row Labels | Σ Values |
|---|---|
| SiteId | Count |
| ProfileId | Cumulative |
| Sample | LayerTop |
| HznDesgn | LayerBottom |

Defer Layout Update | Update

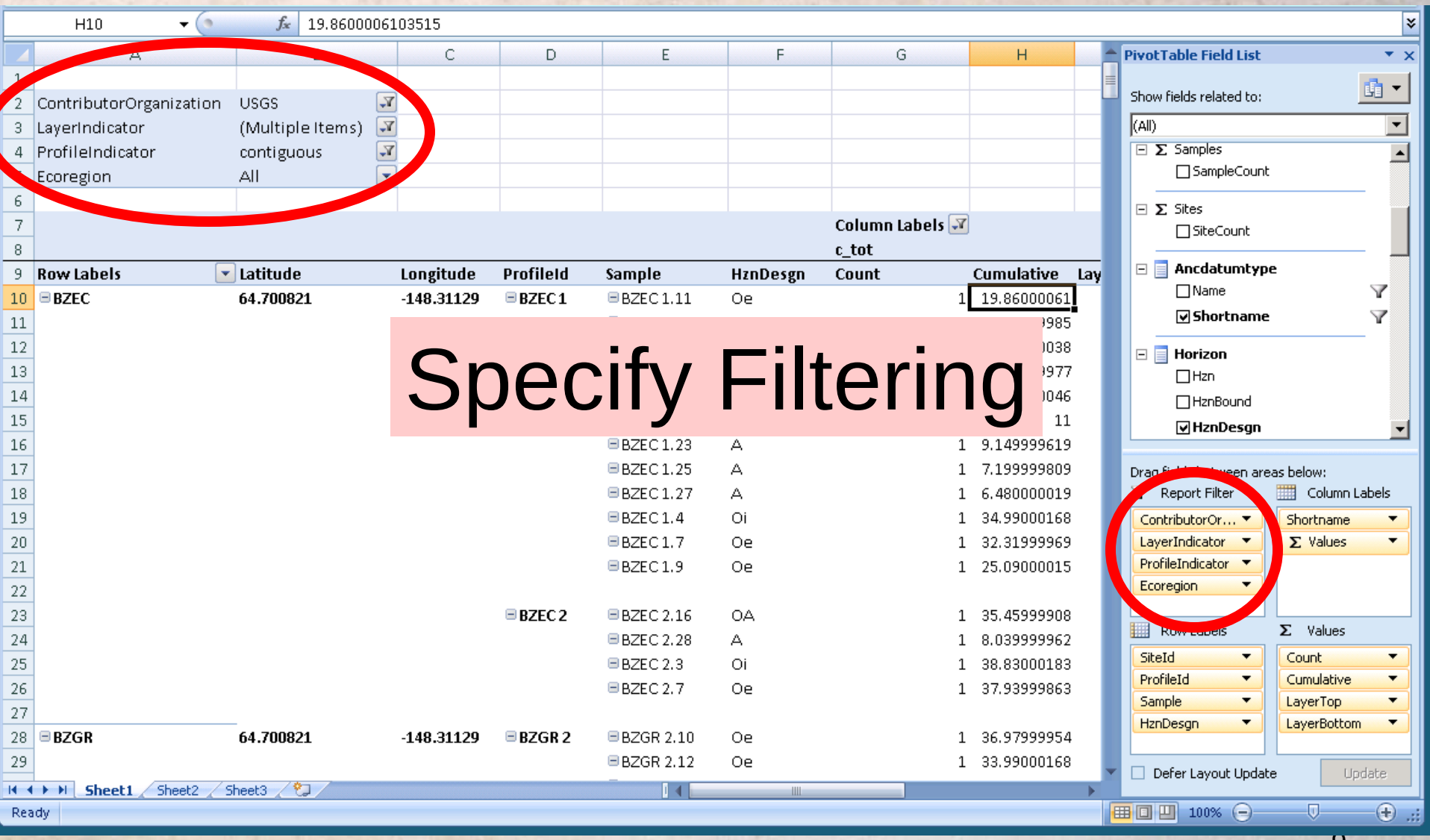
100%

# 3. The Data Cube




Drag and Drop
To Columns and Rows

# 3. The Data Cube



# 3. The Data Cube

G10 | 17

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | ContributorOrganization | USGS | | | | | |
| 2 | LayerIndicator | (Multiple Items) | | | | | |
| 3 | ProfileIndicator | contiguous | | | | | |
| 4 | Ecoregion | | | | | | |
| 5 | HznType | organic | | | | | |
| 6 | | | | | | | |
| 7 | | | | | Column Labels | | |
| 8 | | | | | soc | | |
| 9 | Row Labels | Latitude | Longitude | ProfileId | Cumulative | LayerTop | LayerBottom |
| 10 | BZEC | 64.700821 | -148.31129 | BZEC 1 | 0.936173968 | 0 | 17 |
| 11 | | | | BZEC 2 | [illegible] | 0 | 16 |
| 12 | BZGR | 64.700821 | | | | | |
| 13 | BZMR | 64.700821 | | | | | |
| 14 | D87W | 63.92067 | | | | | |
| 15 | | | | | | | |
| 16 | DF87 | 63.917 | | | | | |
| 17 | DFCB | 63.9217 | -145.74561 | DFCB 11 | 0.513067374 | 0 | 14 |
| 18 | | | | DFCB 12 | 0.590957682 | 0 | 8 |
| 19 | DFTB | 63.9217 | -145.74561 | DFTB 4 | 0.565451991 | 0 | 10 |
| 20 | | | | DFTB DB 1 | 0.266506782 | 0 | 5 |
| 21 | ECHB | 65.664368 | -149.08369 | ECHB 16 | 0.651320979 | 0 | 11 |
| 22 | | | | ECHB 17 | 2.195963943 | 0 | 40 |
| 23 | | | | ECHB 18 | 0.424140483 | 0 | 14 |
| 24 | | | | ECHB 19 | 1.18002902 | 0 | 16.5 |
| 25 | | | | ECHB 20 | 1.820955947 | 0 | 38 |
| 26 | ECHC | 65.664368 | -149.08369 | ECHC 12 | 4.464418948 | 0 | 115 |
| 27 | | | | ECHC 13 | 4.154347435 | 0 | 91 |
| 28 | ECLB | 65.575363 | -148.94881 | ECLB 20 | 0.600886995 | 0 | 17.5 |
| 29 | | | | ECLB 22 | 0.486199994 | 0 | 12 |
| 30 | | | | ECLB 23 | 1.197041959 | 0 | 25 |
| 31 | | | | ECLB 24 | 1.01620798 | 0 | 27 |

Sum by horizon

PivotTable Field List

Show fields related to: (All)

- Cumulative
- LayerBottom
- LayerTop
- Σ Samples
  - SampleCount
- Σ Sites
  - SiteCount
- Ancdatumtype
  - Name

Drag fields between areas below:

Report Filter: Contributo... / LayerIndic... / ProfileIndi... / Ecoregion

Column Labels: Shortname / Σ Values

Row Labels: SiteId / ProfileId

Σ Values: Cumulative / LayerTop / LayerBottom

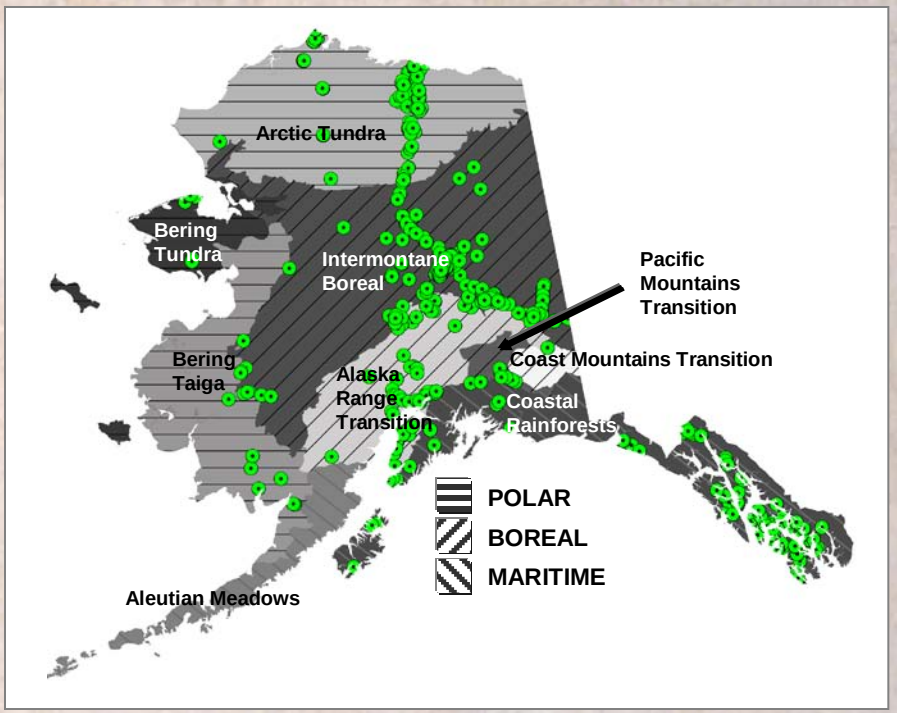
Defer Layout Update | Update

Sheet1 | Sheet2 | Sheet3

Ready | 85%

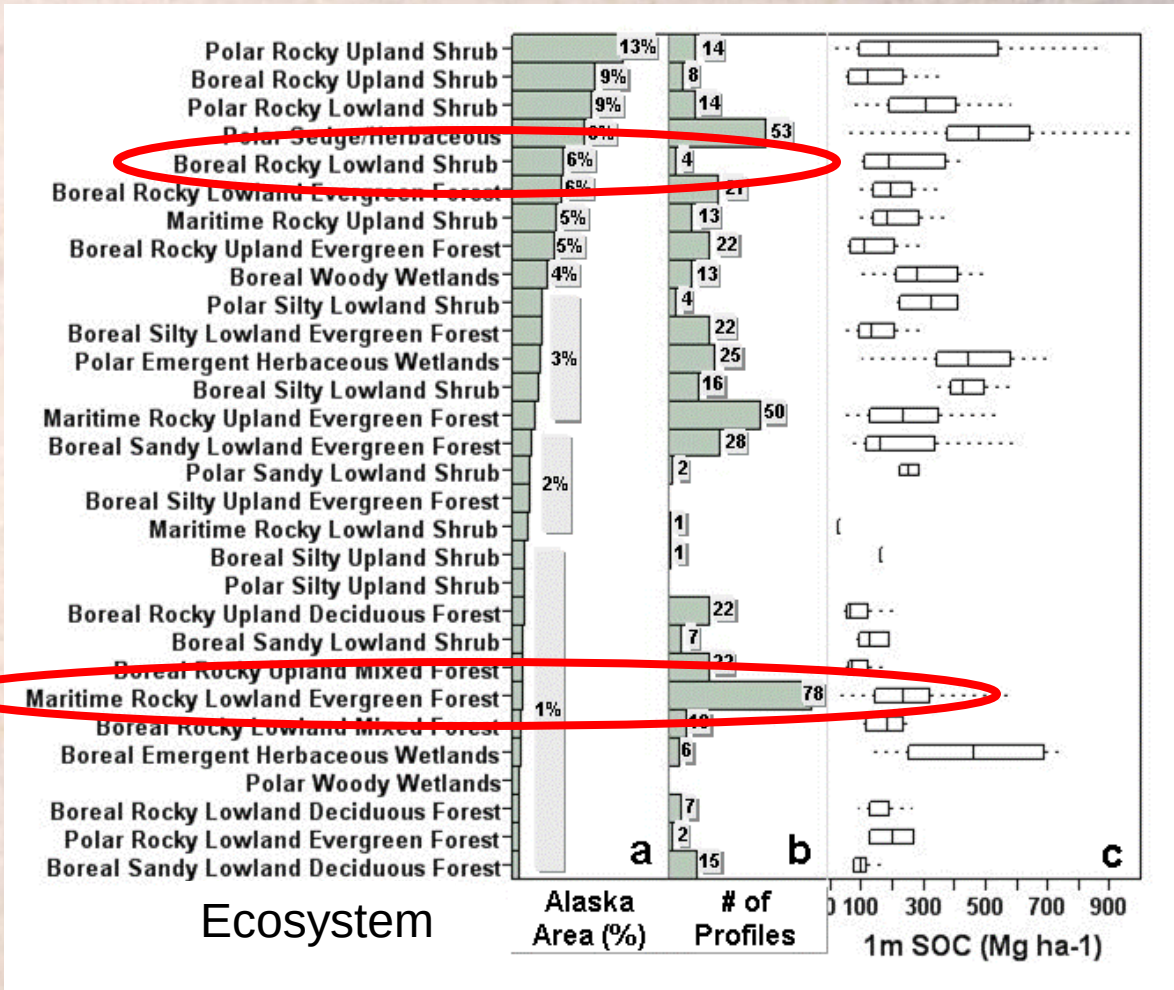
# 4. How can the database help?

Data cube saves time with data preparation.



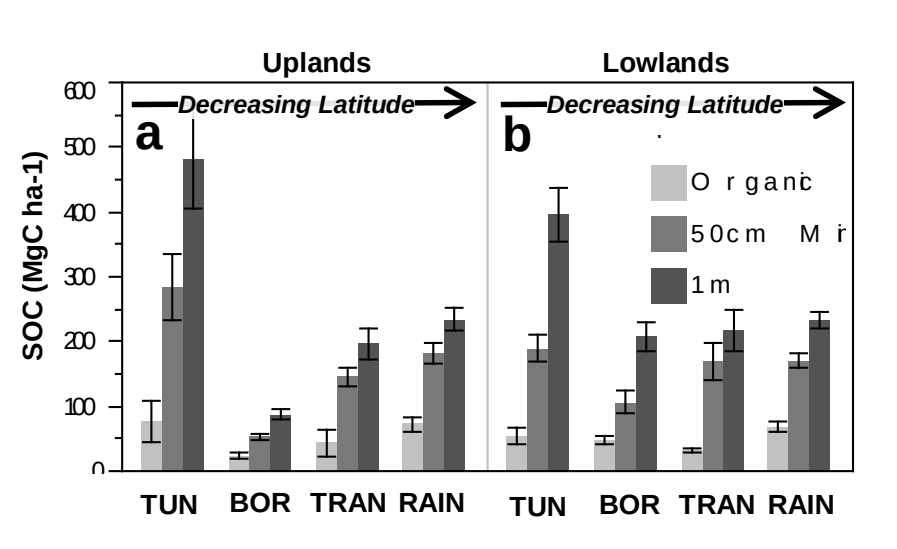
# 4. How can the database help?

Identify data gaps.



# 4. How can the database help?

Meta-analysis using most complete data set



# 4. How can the database help?

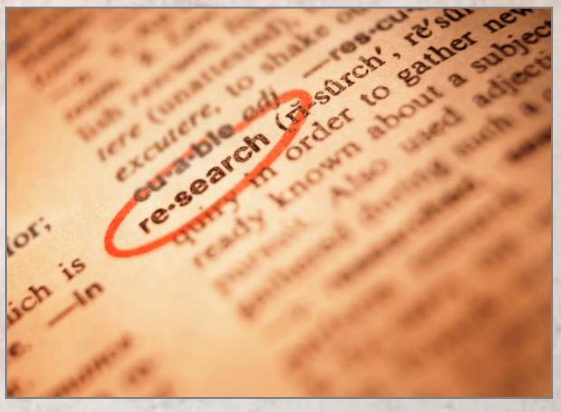
MANY APPLICATIONS

# 5. Summary

It is possible, *and desirable*, to build a community-based soil carbon database

Success depends on a critical mass of participants who will use the database -> “living”

Database technology, such as the “data cube”, provide the means to make the database useful for research.



# Acknowledgments

## Collaborators

- Jennifer Harden
- A. David McGuire
- Norman B. Bliss
- James G. Bockheim
- Mark Clark
- Teresa Nettleton-Hollingsworth
- M. Torre Jorgenson
- Evan S. Kane
- Michelle Mack
- Jonathan O'Donnell
- Chien-Lu Ping
- Edward A. G. Schuur
- Merritt R. Turetsky
- David W. Valentine

Deb Agarwal and Catharine Van Ingen
from Berkeley Labs and Microsoft Research

Funding: USGS, Global Change Office, Deep Soil Carbon Project

# National Soil Carbon Network: Continuing Database Development

Lucas Nave
NSCN Network Coordinator

National Soil Carbon Database Development Town Hall Meeting

# Data Sources

**<u>Follow AK model:</u>**

1. **Engage personal contacts**
2. **Organize data sharing workshop**
3. **Incorporate extensive NRCS & USGS data**

# Database Access: Fluxnet Synthesis Sharepoint Site

**http://www.fluxdata.org/nscn**

Principal developers: Deb Agarwal, Marty Humphrey, Catharine van Ingen

Funding agencies: USDA-Forest Service (NRS and GCRP), USDA-NIFA (REE-NET), and USGS (GCRP)

# Public-accessible Main Page

National Soil Carbon Network - Home - Windows Internet Explorer

Favorites | Suggested Sites | Web Slice Gallery

National Soil Carbon Network - Home

Page | Safety | Tools

Site Actions | Browse | Page | Publish | Sign In

National Soil Carbon Network › Home

Home | Dataset Information | Search this site...

## National Soil Carbon Network

**Welcome!** The National Soil Carbon Network (NSCN) is an organization of scientists from academia, government, and the private sector working towards a large-scale synthesis of soil C research in the United States. The principal goals of the Network are to produce databases, models, and maps that enable users to understand: 1) how much C is stored in U.S. soils, 2) how long this C remains in the soil, and 3) the factors that make this soil C vulnerable to being lost (*i.e.*, emitted to the atmosphere). Overarching these goals is the need for a spatially explicit approach, since measurements of soil C storage, turnover, and vulnerability all vary at spatial scales from several meters to thousands of kilometers, as well as across different depths within an individual soil profile. In addition to database development, modeling, and mapping, the Network also facilitates topical workshops, peer-reviewed publication, and presentations at professional conferences.

### The Network is:

- compiling soil sampling, archiving and analysis protocols
- sharing scientific, analytical, and logistical infrastructure
- synthesizing products beneficial to stakeholders and scientists
- developing community-accessible data resources

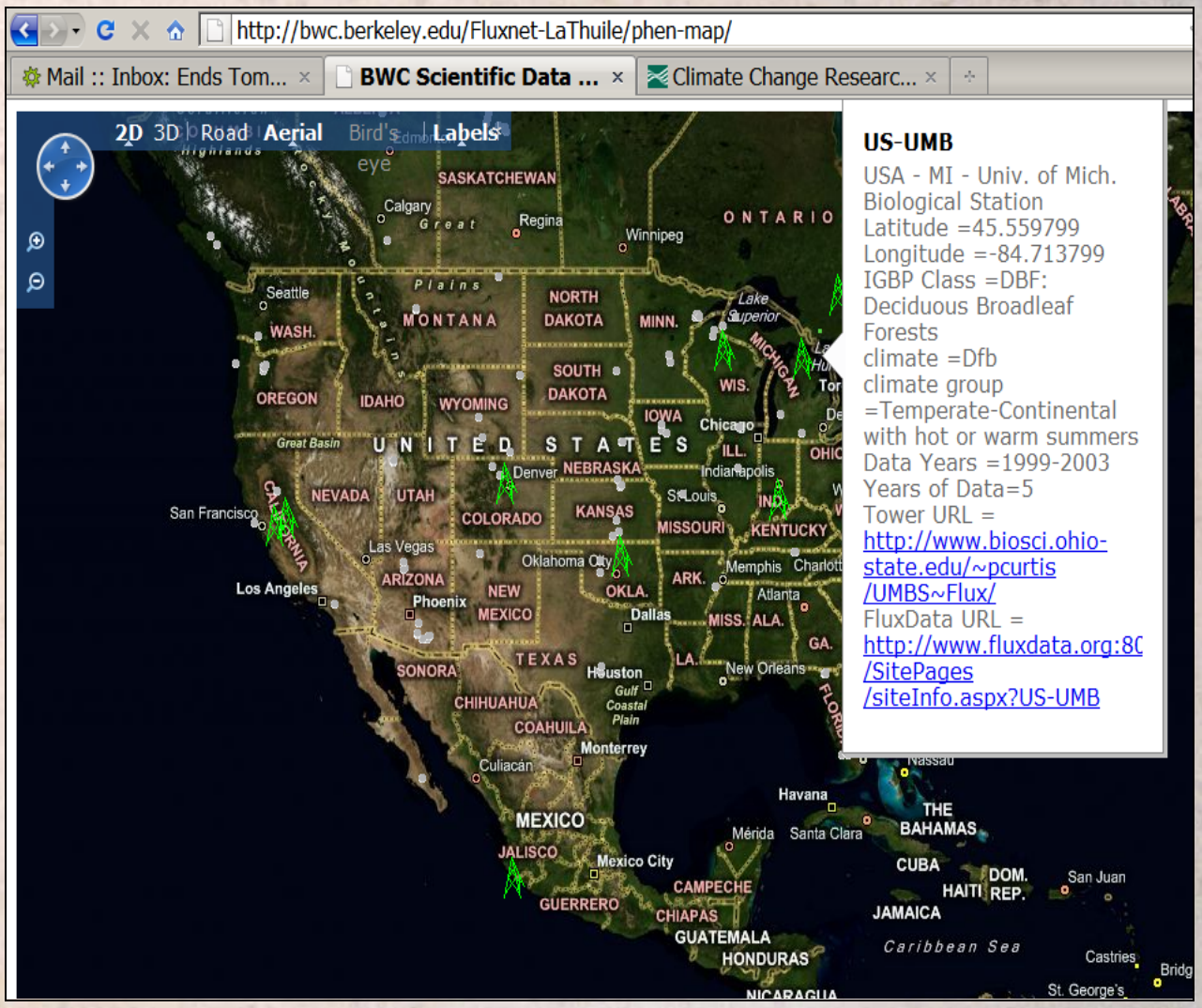
### Why soil carbon?

Soil is a vital national resource and soil carbon is an integral component of soil structure and function. Soil is the largest terrestrial reservoir of carbon, containing an estimated 1550 Pg of organic carbon in the top meter alone. Soil contains twice as much carbon as the atmosphere (800 Pg in 2007), and three times that in global vegetation (~500 Pg). Although the global stock of soil carbon is immense, it is not static: about 120 Pg of carbon moves annually between soil and the carbon reservoirs in the atmosphere and vegetation. Soil carbon may thus play a singular but uncertain role in climate forcing during the coming decades, with significant net losses contributing to positive feedbacks, or significant sequestration helping to mitigate climate forcing.

Internet | Protected Mode: On | 100%

# Online Data Source Map

- Site-level metadata and contact information
- Self-updating in parallel with living database
- MS-Virtual Earth



# Online Data Availability Table

- Organized by contributor or location (ecoregion)
- Definitions of columns in the database, number of sites, profiles, samples

Dataset Information - Datumtype

Site Actions | Browse | Page | Publish

Dataset Information ▸ Datumtype

Home | Dataset Information

Excel Web Access - Datumtype

File | Data | Find

## Database and Datacube Datumtypes

Download latest spreadsheet

Generated:12/9/2010 6:59:59 PM

| Variable Type | Datumtype | Name | Type | Units | Description | Para |
|---|---|---|---|---|---|---|
| Profile | | | | | | |
| | ccon_prof_proc | Processed Profile Organic Carbon Content | REAL | g cm-2 | These are organic carbon content calculations for individual profiles based on special procedures for handling missing data (e.g. %C v. bulk density relations). Most ideal are estimates to a specific depth or horizon in the profile. Please document the method in the associated ccon_proc_depth, ccon_proc_type (mineral or organic), and ccon_proc_ref. Units: g cm-2 | ccon_ |
| | cryo_flag | Cryoturbated Flag | TEXT | | The profile shows a significant degree of cryoturbation. Enter "yes" if true, otherwise leave blank. | ;;;; |
| | desc_profile | Profile Description | TEXT | | Enter information that helps describe the profile. | ;;;; |
| | sampler_names | Sampler Names | TEXT | | The names of the persons who described and sampled the profile | ;;;; |
| | soiltaxon | Soil Taxonomy | TEXT | | Enter the taxonomic classification of the profile | ;;;; |
| | soil_series | Soil Series | TEXT | | The names of the NRCS soil series that the profile belongs to. | ;;;; |
| | add_taxon_flag | Complete Taxonomy Flag | TEXT | | Indicate that whether there is more taxonomic information for this profile by entering "yes", otherwise leave the field blank. | ;;;; |
| | thaw_depth_profile | Thaw Depth of the Profile | TEXT | | The depth to the frozen surface of the profile IF SAMPLED AFTER AUGUST 15. | ;;;; |
| | surface_veg | Surface Vegetation | TEXT | | Add details for the vegetation at the exact location of the profile | ;;;; |

Done | Internet | Protected Mode: On | 100%

# Database Access

**Direct access to data cube using PivotTables, for advanced users and customizable data viewing**

| | |
|---|---|
| ContributorOrganization | USGS |
| LayerIndicator | (Multiple Items) |
| ProfileIndicator | contiguous |
| Ecoregion | All |
| HznType | organic |

| Row Labels | Latitude | Longitude | ProfileId | Cumulative | LayerTop | LayerBottom |
|---|---|---|---|---|---|---|
| BZEC | 64.700821 | -148.31129 | BZEC 1 | 0.936173968 | 0 | 17 |
| | | | BZEC 2 | 0.859842978 | 0 | 16 |
| BZGR | 64.700821 | -148.31129 | BZGR 2 | 1.080353983 | 0 | 18 |
| BZMR | 64.700821 | -148.31129 | BZMR 2 | 4.772603452 | 0 | 79 |
| D87W | 63.92067 | -145.3777 | D87W 7 | 0.427561976 | 0 | 15.5 |
| | | | D87W 8 | 0.443791486 | 0 | 8.5 |
| DF87 | 63.917 | -145.37199 | DF87 1 | 0.218627989 | 0 | 5 |
| DFCB | 63.9217 | -145.74561 | DFCB 11 | 0.513067374 | 0 | 14 |
| | | | DFCB 12 | 0.590957682 | 0 | 8 |
| DFTB | 63.9217 | -145.74561 | DFTB 4 | 0.565451991 | 0 | 10 |
| | | | DFTB DB 1 | 0.266506782 | 0 | 5 |
| ECHB | 65.664368 | -149.08369 | ECHB 16 | 0.651320979 | 0 | 11 |
| | | | ECHB 17 | 2.195963943 | 0 | 40 |
| | | | ECHB 18 | 0.424140483 | 0 | 14 |
| | | | ECHB 19 | 1.18002902 | 0 | 16.5 |
| | | | ECHB 20 | 1.820955947 | 0 | 38 |
| ECHC | 65.664368 | -149.08369 | ECHC 12 | 4.464418948 | 0 | 115 |
| | | | ECHC 13 | 4.154347435 | 0 | 91 |
| ECLB | 65.575363 | -148.94881 | ECLB 20 | 0.600886995 | 0 | 17.5 |
| | | | ECLB 22 | 0.486199994 | 0 | 12 |
| | | | ECLB 23 | 1.197041959 | 0 | 25 |
| | | | ECLB 24 | 1.01620798 | 0 | 27 |

PivotTable Field List

Σ Samples
SampleCount
Σ Sites
SiteCount
Ancdatumtype
Name

Drag fields between areas below:

Report Filter: Contributo... / LayerIndic... / ProfileIndi... / Ecoregion

Column Labels: Shortname / Σ Values

Row Labels: SiteId / ProfileId

Σ Values: Cumulative / LayerTop / LayerBottom

Defer Layout Update — Update

# Database Access

- Automatically self-updating reports generated from the data cube for typical use cases
- Explore in browser or download .xls

Data Browsing - CarbonByHorizon - Windows Internet Explorer

http://www.fluxdata.org/nscn/Data%20

Favorites | Suggested Sites | Web Slice Gallery

Data Browsing - CarbonByHorizon

Site Actions | Browse | Page | Marty Humphrey

Data Browsing › CarbonByHorizon

Home | Dataset Information | Data Browsing

Excel Web Access - CarbonByHorizon

File | Open in Excel | Data | Find

Profile Carbon by Horizon

*Download latest spreadsheet*

Generated:12/10/2010 9:00:43 PM

**Data as of 02/25/2010; Processed 04/11/2010; Profiles with contiguous samples only**

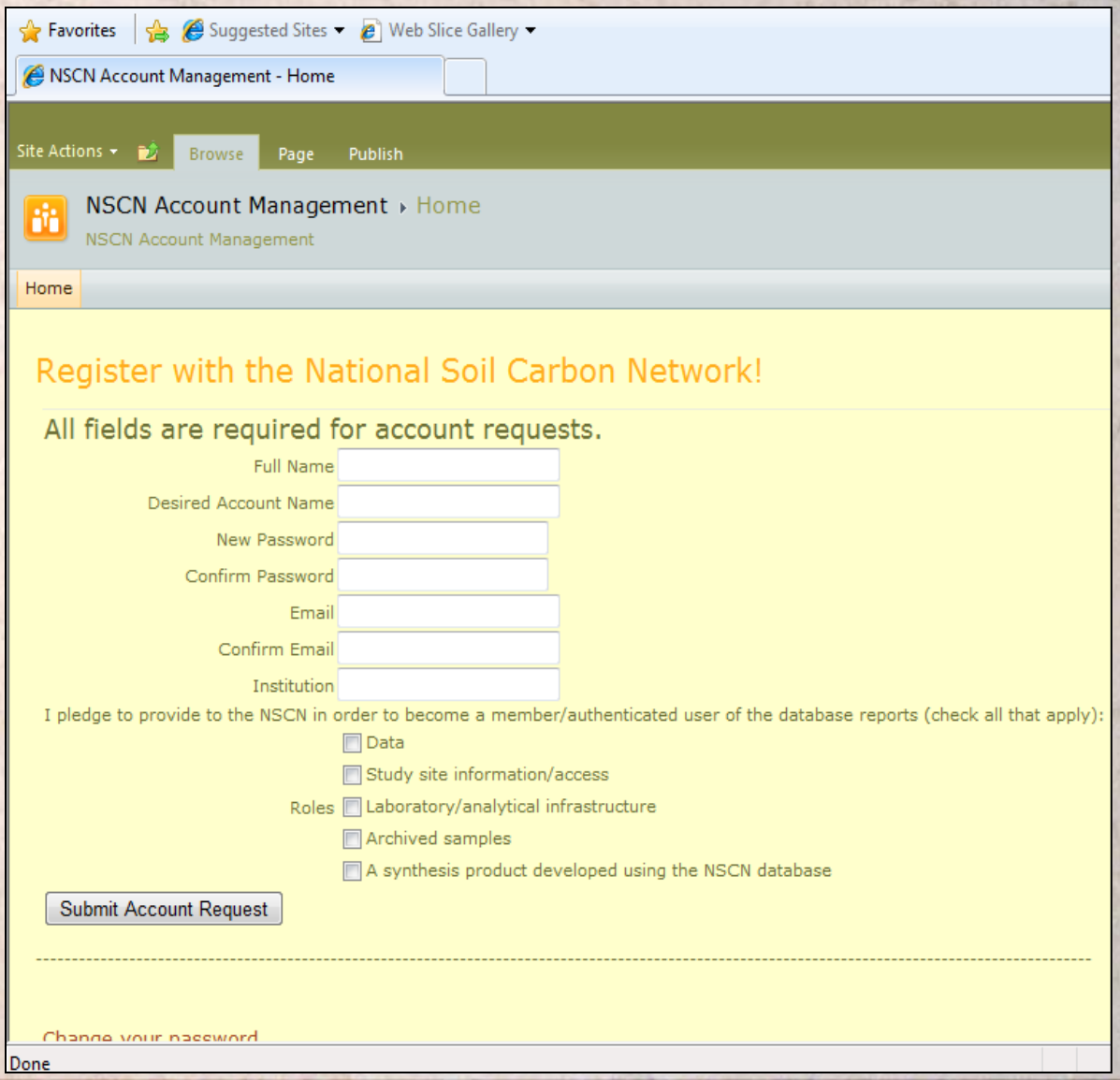
| Site Name | Latitude | Longitude | Eco-Region | Profile Name | Measure Date | Horizon | SOC | Layer Thickness | Sample Count |
|---|---|---|---|---|---|---|---|---|---|
| 00AK068001 | 63.48355484 | -150.1499023 | Alaska Range | 00AK068001 | 8/2/2000 12:00:00 AM | A | 0.2444 | 3 | 1 |
| 00AK068001 | 63.48355484 | -150.1499023 | Alaska Range | 00AK068001 | 8/2/2000 12:00:00 AM | C | 2.2949 | 68 | 4 |
| 00AK068001 | 63.48355484 | -150.1499023 | Alaska Range | 00AK068001 | 8/2/2000 12:00:00 AM | O | 0.9579 | 29 | 4 |
| 00AK068001 | 63.48355484 | -150.1499023 | Alaska Range | 00AK068001 | 8/2/2000 12:00:00 AM | Total | 3.4972 | 100 | 9 |
| 00AK068002 | 63.43358231 | -150.7666931 | Alaska Range | 00AK068002 | 8/2/2000 12:00:00 AM | B | 1.9133 | 68 | 3 |
| 00AK068002 | 63.43358231 | -150.7666931 | Alaska Range | 00AK068002 | 8/2/2000 12:00:00 AM | C | 0.1047 | 16 | 1 |
| 00AK068002 | 63.43358231 | -150.7666931 | Alaska Range | 00AK068002 | 8/2/2000 12:00:00 AM | E | 0.4308 | 3 | 1 |
| 00AK068002 | 63.43358231 | -150.7666931 | Alaska Range | 00AK068002 | 8/2/2000 12:00:00 AM | O | 0.3012 | 9 | 4 |
| 00AK068002 | 63.43358231 | -150.7666931 | Alaska Range | 00AK068002 | 8/2/2000 12:00:00 AM | Total | 2.75 | 96 | 9 |

CarbonByHorizon

Done | Internet | Protected Mode: On | 100%

# Join NSCN for Access

- Database browsing and access requires membership in the NSCN
- Online member signup makes joining easy
- Indicate intended contributions to the NSCN at the time of registration (data upload)



# Acquiring Data from New Sources

Microsoft Excel - forest soil C&N database

F119 C/N

| | Line # | Acc# | M-d pgs | R-d pg | R-d source | Parameter | Units | MAT | MAP | horizon | quality flag | method | LAT/LONG | CTRL n | CTRL X | CTRL SE | TRT n | TRT X | TRT SE | Effect size |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 109 | 108 | 2005.61 | 2 | 4 | T4 | N | Mg ha-1 | 11.7 | 872 | BC | D | Kjeldahl | 43S146E | 2 | 7.46 | * | 2 | 5.69 | * | 0.76 |
| 110 | 109 | 2005.61 | 2 | 4 | T4 | N | Mg ha-1 | 11.7 | 872 | BC | D | Kjeldahl | 43S146E | 2 | 7.46 | * | 2 | 5.20 | * | 0.70 |
| 111 | 110 | 2005.61 | 2 | 4 | T4 | N | Mg ha-1 | 11.7 | 872 | BC | D | Kjeldahl | 43S146E | 2 | 7.46 | * | 2 | 3.38 | * | 0.45 |
| 120 | 119 | 2005.8 | 2 | 4 | T2 | C | % | 3 | 625 | O | * | LOI | 61N25E | 1 | 41 | 6.6 | 1 | 64 | 16 | 1.56 |
| 121 | 120 | 2005.8 | 2 | 4 | T2 | C | % | 3 | 625 | O | * | LOI | 61N25E | 1 | 35 | 3 | 1 | 55 | 3.5 | 1.57 |
| 122 | 121 | 2005.8 | 2 | 4 | T2 | N | g kg-1 | 3 | 625 | O | * | CHN | 61N25E | 1 | 24 | 1.7 | 1 | 24 | 1.7 | 1.00 |
| 123 | 122 | 2005.8 | 2 | 4 | T2 | N | g kg-1 | 3 | 625 | O | * | CHN | 61N25E | 1 | 27 | 1.3 | 1 | 24 | 1.7 | 0.89 |
| 124 | 123 | 2005.8 | 2 | 7 | F3 | Nmin | mg kg-1 | 3 | 625 | O | * | lab inc | 61N25E | 1 | 108 | 12 | 1 | 89 | 49 | 0.82 |
| 125 | 124 | 2005.8 | 2 | 7 | F3 | Nmin | mg kg-1 | 3 | 625 | O | * | lab inc | 61N25E | 1 | 193 | 24 | 1 | 128 | 62 | 0.66 |
| 126 | 125 | 2004.22 | 2,3 | 4 | T1 | C | Mg ha-1 | 6.4 | 1032 | A | D | CHN | 51N9E | 4 | 227 | * | 4 | 183 | * | 0.81 |
| 127 | 126 | 2004.22 | 2,3 | 4 | T1 | C | Mg ha-1 | 6.4 | 1032 | BC | D | CHN | 51N9E | 4 | 138 | * | 4 | 147 | * | 1.07 |

- **Incorporate published data from forest management / soil carbon meta-analysis databases**
- **Ask ad hoc groups & formal networks to contribute large datasets (*e.g.*, LTSE, LTSP, LTER, NERC, SMC)**

# New Database Functionalities

- Incorporate control / treatment structure for meta-analysis
- Populate database with $^{14}C$ data for modeling mean residence times
- Develop map-based data access

OUTREACH NOTICE: POSTDOC- SOIL BIOGEOCHEMISTRY
Northern Research Station, USDA Forest Service
Center for Accelerator Mass Spectrometry, LLNL

**Post-doctoral Researcher: Soil Biogeochemistry**

The USDA Forest Service, in collaboration with Lawrence Livermore National Laboratory, is advertising a Post-doctoral Researcher position with the Duty Station located in Livermore, California. The position will focus on terrestrial carbon cycle studies, with an initial appointment for 1 year, renewable annually for up to 3 years

***Questions about the position should be addressed to Chris Swanston at cswanston@fs.fed.us.***
***Application period closes February 15, 2011, or upon the selection of a qualified candidate.***

*THE POSITION*
The incumbent will work as a postdoctoral researcher within the Climate, Fire, and Carbon Cycle Science Group in the Forest Service Northern Research Station, and as a postdoctoral staff member in the Lawrence Livermore National Lab Center for Accelerator Mass Spectrometry (CAMS). The postdoc will interact with numerous projects, with a common

# Encouraging Participation & Continued Growth

Advertise benefits of data sharing:

1. Access to useful synthesis products
2. Access to analytical infrastructure, archives, and study sites
3. Attainment of proposal requirements for data management

Enhance database utility:

1. Develop databases for other soil C measurements (DSC, FTIR, black carbon)
2. Encourage specific individuals to use databases (climate/C cycle modelers, grad students doing lit reviews)
3. Offer synergy with Fluxnet as a terrestrial C data clearinghouse

# Questions, Advice, Comments?

<u>Panel:</u>

Deb Agarwal: database mechanics

Jen Harden: AK deep soil C project leader

Kris Johnson: AK database coordinator

Luke Nave: NSCN coordinator

Chris Swanston: NSCN chair

Catharine van Ingen: database programming