

Meeting Agenda
NAACCR GIS Committee
Wednesday, March 12, 2008
2:00-3:00 PM Eastern Time
1:00-2:00 PM Central Time
12:00-1:00 PM Mountain Time
11:00-12:00 noon Pacific Time
10:00-11:00 AM Alaskan Time

Toll free number: 1-866-784-0143

Passcode: 9425222

A. Roll Call

Members Present: Frank Boscoe – Chair, Kevin Henry –Vice Chair, Pam Agovino, Stephanie Foster, Chris Newton, Carolyn Halbert, Dave Stinchcomb, Vijay Raghavan, Annette Gardner (CDC), Christian Klaus, Jennifer Kachajian, Cathy Riddell, Yongping Hao, Bob Borchers, Jim Files

NAACCR Staff Present: Joshua Whitley

Members Absent: Recinda Sherman, Jim Martin, Myles Cockburn, Susan Brown, Heidi Gianella, Holly Hodges, David O'Brien, Mary Mroszczyk, Chris Johnson,

B. Minutes

1. Approval of January and February minutes (note: January minutes still not available as of 3/12, 9:30 am) – February Minutes: Chris Johnson motioned to approve the minutes as written, Carolyn seconded. The motion was passed.

C. Old Business

1. Geocoding Best-Practices Document – Dan
Copy editing is in progress. NAACCR will retain the copyright to the document. A limited number of physical copies will be printed and the document will be mainly published online. The document is still being copy-edited by Dan's supervisor (head of geography at USC). He will have it back to the committee in a few weeks.
2. Rules for address consolidation – Bob
See attached excel file. Bob asked Frank to flesh out the statistics on the Wisconsin situation for one-to-many and variability of elements used for criteria about one month ago. There were multiple addresses among multiple abstracts that get consolidated. The committee wanted to know how many cases we were talking about. Bob tried to lay this out in the spreadsheet. He noted that he would be happy to do

further queries off of the database, if requested. Someone asked whether the abstracts are being geocoded or if just the consolidated records are being geocoded. Bob encouraged everyone to take a similar look at their own data to flesh out the ranged variation of these factors. The finding that the county attribute can really change according to which abstract you're taking is consistent with what Christian and Bob came up with. The variation can be up to 20%-30%. Dave Stinchcomb asked if there was an analysis that shows how many abstracts you'll have. This document doesn't describe how you went about consolidation. Bob said that his records only have address information after geocoding. This document is that output.

Stephanie Foster from the CDC said that she hoped that more than a few were aware that NPCR is attempting to create a Geocoding system that will be available through some sort of web system to everyone. Carolyn asked when the software would be available. Stephanie said that the CDC is hoping by the end of 2009. Christian and Bob have a program working that is designed to be used pre-geocoding for address validation for abstractors. Dan Goldberg is working on making it a web application. It is designed to be free and open-source. In NC, no less than 5 abstractors have asked for this application. This should help save CCRs money by cleaning up data on the front end. Frank asked if it would be possible for those from the CDC working on the Geocoding application to address the GIS committee. Stephanie said they are at a team development point and more should be fleshed out in a few weeks. Stephanie will give an update on the next call.

Frank told Bob that he thought there was sufficient evidence on the Excel spreadsheet to show that a lot of cases come up as having to make a choice. Bob would like to give one more month for people to look at this. Bob was asked how he is defining "different addresses." The information he was using to generate the spreadsheet was all post-geocoding, so there shouldn't be any very close but different addresses because they should be already standardized.

3. Geocoding NPI –Kevin

Kevin has finished geocoding the file. His concern is about how often we do it since it is very time-consuming. The geocoding takes between 6-8 hours. Post-processing takes total of about 4-5 hours on his PC. It was decided that we really can't do enhancements, however he did code it so we can have coordinate quality in the file too. Distribution of the information was discussed. The group wondered if registries would just get their own state's data or download the complete file.

Semantic Interoperability is going through the NPI file too. Kevin is going to sit on their next call. Bob thinks there might be a difference in individual address quality versus large provider address quality. He also thought that more recent GDT data might have created better matches. Bob will run through their geocoder to see how well the data meshes. Understanding the reasons for non-matches might be useful for release to the NAACCR website.

The question was raised as to which fields to give back. Bob thinks he should just give back the 30 fields and NPI number. Josh asked about the output format. Kevin thought it would be best to use a fixed-width format, in case there are commas.

D. New Business

No new business. Meeting adjourned at 1:51 PM.

Next meeting on April 9th.