

CDI Stream Summarization Face-to-face Meeting Agenda

August 3-5th

Meeting Location

CSAS Central Conference Room Suite 2500
Denver Federal Center, Building 810
Lakewood, CO 80225

Remote Participation

Webex: <https://usgs.webex.com/meet/dwieferich>
Teleconference: Voice: DOI locations: 703.648.4848, Toll Free Non-DOI locations: 855.547.8255
Code: 98094#

Contact Information

Coordination: Daniel Wieferich, 303-202-4594 or dwieferich@usgs.gov
Travel: Robin Schafer, 303-236-1701 or rschafer@usgs.gov

Objectives of Meeting (From Proposal):

- Get feedback on existing summarization workflows, highlighting both similarities and differences
- Based on feedback and past experiences, set guidelines for summarization workflow
- Determine list of high priority variables for all participating collaborators
- Discuss and determine best options for code and data dissemination to make it most valuable to users and collaborators. For example, discuss appropriate data type of released summary information (e.g. shapefile vs. .hdf vs. .csv)

Note: This will be a two part meeting. Phase 1 will include potential user groups and PIs, and Phase 2 will only include PIs to follow up and develop a study plan.

August 3rd – Phase 1

The morning session will focus on introducing three national efforts of stream summarization and their associated workflows.

9:00 AM – Cheryl Morris (Director of CSAS&L) : Welcome and CDI Overview

Wieferich: Logistics, Introductions, and Agenda

9:15 AM – Wieferich: Overview of CDI Stream Summarization Meeting, document terminologies -

9:25 AM – Give a general overview of three commonly used versions of the national hydrography dataset (i.e. 1:100,000 scale NHDPlusV1, 1:100,000 scale NHDPlusV2, and 1:24,000 scale NHD High Resolution), the differences between these datasets and provide an oversight of future versions.

Cindy McKay (Horizon Systems): NHDPlus Medium Resolution (1:100,000) and CA3T (15 minutes)

Alan Rea: NHD High Resolution and Future Directions (10 minutes)

Group Questions and Discussion: 10 minutes

Three groups will discuss national stream summarization efforts. The general breakdown of these presentations will include context of why stream summarizations workflows were developed (e.g. projects and efforts needing the summaries)(~10minutes), brief description of summarization methods (~10 minutes), and group questions/discussion (~10minutes) AM – Discussions, Resolve Questions, Comments

10:00 AM – Mike Wieczorek: National Water Quality Assessment Program (NAQWA)

10:30 AM - Break

10:45 AM – Scott Leibowitz / Marc Weber: EPA Safe and Sustainable Water Resources National Program

11:15 AM – Dana Infante / Daniel Wieferich: Michigan State University (MSU)

11:45 AM - Wrap up morning session. Participants are encouraged to begin brainstorming about how presented workflows may influence their work. Are current workflows meeting the needs of your efforts and/or your program needs? Think about documented decision points below.

12:15 Lunch – Ordered in Jason’s Deli

Afternoon will start with a recap of the three national workflows. Session will follow with presentations of applications and needs of landscape information summaries for specific user groups. During these presentations, participants should begin thinking of how user group needs compare to existing workflows and decision points. (e.g. As PIs develop a standardized workflow what can be done to make summarized data more useful for users? Are there additional decision points that need to be discussed?)

1:15 PM – Wieferich (and other PIs): Recap and Comparisons of existing workflows. Introduce known decision points and

15 minute Presentations with 5 minutes of questions and discussions

2:00 PM – Abigail Lynch: USGS National Climate Center

2:20 PM – Jeff Houser: USGS UMESC Science Center

2:40 PM – ~~USFWS Landscape Conservation Cooperatives~~

- Trip Hook: EPA Office of Pesticide Programs (OPP)

3:00 PM – Break

3:10 PM – Roland Viger: USGS Water Resources National Research Program

3:30 PM – Arthur Cooper: Michigan Department of Natural Resources

3:50 PM - Group Discussions

Goals of this Discussion: Identify any newly identified decision points within the summarization workflow based on user group presentations and discussions. Discuss Decision Points (e.g. Feedback on decisions points based on use cases)

5:00 PM - Wrap Up

6:00 – Social: Demarra’s
1100 West Alameda Ave.
Lakewood, CO 80226

August 4th – Morning Continue Part 1; Afternoon Part 2

9:00 AM – Regroup

9:10 AM - Goncalo Duarte (University of Lisbon, Portugal): RivTool
Recorded @ <https://usgs.webex.com/usgs/lsr.php?RCID=bee7a3824c494e0392cfed76df966cf0>

9:30 AM - Wieferich: Recap from August 3rd discussions and progress

11:30 Wrap-up Collaborator Discussions

Noon – Lunch (Walk to Federal Center Farmers Market @ Bicentennial Park at corner of 4th and Main)

1:30 PM PIs reflect on discussions and past experiences, work to set guidelines for summarization workflow

PI discussions on how to move forward given the feedback from cooperators and knowledge of our products and previous work

5:00 PM – Wrap Up

August 5th – Resume PI discussions and Programming

Tasks:

9:00 Regroup

9:15 Discussion about NHDPlus Improvements with Tommy Dewald, Dave Wollock

10:30 Wrap up Meeting

Draft Summary of Meeting

Draft Study Plan

Discuss in more detail methodologies and programming (more technical information)

Wieferich / Brad Williams: Discuss HDF file format and multi-core computing findings

Programmers get to work...

Noon: Meeting Dismissal

Discussion Topics for the CDI Stream Summarization Meeting August

Landscape data has been summarized by a number of groups in a variety of ways and to a variety of spatial units. The USGS Community of Data Integration funded a proposal to have several groups work together on developing a standardized workflow for summarization of landscape data to stream networks, focusing current efforts on the National Hydrography Dataset Plus Version 2 dataset. We see this as an opportunity to involve stakeholders and to gather feedback on decision points of summarization methods and variables of interest to help ensure, when possible, our work is valuable to our partners. This document is intended to gather informal feedback on a few discussion topics to help spark conversation at a face-to-face meeting in August.

Context to Help Understand Discussion Topics

Stream spatial network terminology used in discussion points

Blue Line = Stream reach, typically defined as a stream segment connecting 2 of the following; stream origins, stream confluences, lake or reservoir inflows or outflows, and/or stream outflows

Yellow = 90 meter buffer of stream

Blue Polygon = Local Catchment, catchment draining to an individual stream reach

Green Polygon = Upstream Drainage Catchment, upstream drainage of a given reach NOT including the local catchment

Green Polygon + Blue Polygon = Network Catchment, entire upstream drainage of a given reach including the local catchment

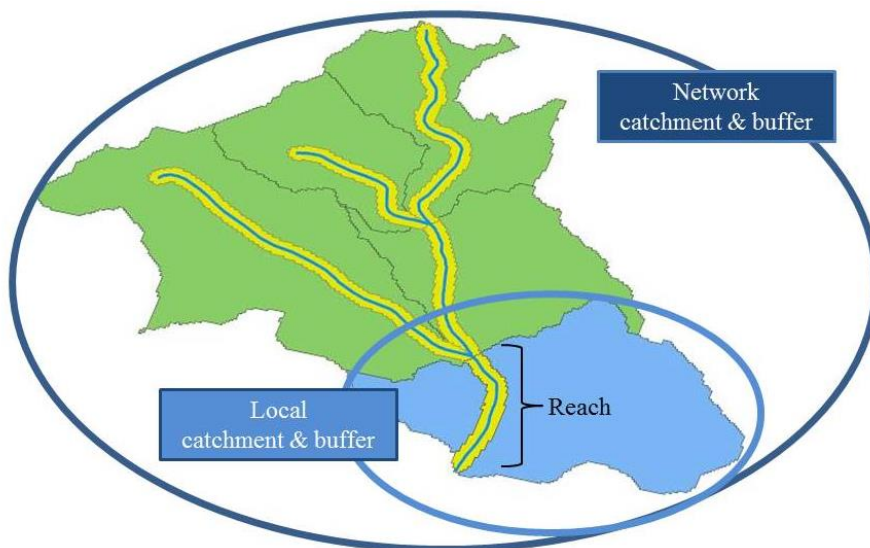


Figure borrowed from NFHP's 2015 Through A Fish's Eye Report

Examples of landscape data and how they are summarized

Data	Example Summary Type
Land Cover	Percent of spatial unit per land use class (e.g.

	percent of 90 meter buffer that is forested)
Elevation	Min, Max, Average per Catchment (e.g. highest elevation within a network catchment)
Area	Area of spatial unit (e.g. area of drainage for a network catchment)
Canals, Pipelines, Ditches	Density (e.g. Density of ditches per local catchment)
Imperviousness	Percent (e.g. Percent of upstream drainage catchment that has impervious surface)
Water Table Depth	Mean
Depth To Bedrock	
Soil / Lithology	Mean % content per class per catchment
Population	Mean Density
Pollutant Measures (NPDES, TRI, Superfund Sites)	Density , Count (e.g. Count of superfund sites within a network catchment)
Climate (e.g. PRISM, Future Projections...)	e.g. Max temperature per network catchment
Roads	# road/stream intersections, density of roads per catchment
Conservation Status	Percent of spatial unit per conservation status designation (e.g. % of catchment protected designated as national park land)
Mines (coals, mineral, uranium)	Density
Large Dams	Density, Count

Discussion Topics:

- When summarizing landscape information which of the following spatial scales are important to your work? (See above figure for more description) When it is possible, give project context and when it applies identify stream network used (e.g. NHD High Resolution, NHDPlusV2 Medium Resolution, NHDPlusV1, other).

Spatial Scale	Yes, No, Explanation/Context
Stream reach (e.g. count of dams per reach)	
90 meter buffer – Specify Stream Network	
Other sized buffers, please explain	
Local Catchment – Specify Stream Network	
Upstream Drainage Catchment – Specify Stream Network	
Network Catchment – Specify Stream Network	
Network Catchment – Main Stem Only (No Tributaries included in summaries)	
Downstream Network Catchment	
Hydrologic Unit 12 (HUC12)	
Hydrologic Unit 8 (HUC8)	
Level 4 Ecoregion	
Other?	

- When summarizing landscape information to local and upstream network catchments or watersheds, current efforts have included the following summary options: Sum, Mean, Max, Min. Are there other summarization statistics that would be helpful for your science products (e.g. Variance, Mode, Mean)? When possible please provide context by using an example dataset and an example of how the summarization might be used.
- What landscape data are important to your science projects? Examples of data are listed in the above table. For your science projects please list the most commonly used or mostly likely valuable landscape data and list any landscape data that would be valuable to your science that are not included in the table. When possible use examples and potential data sources of interest. It is important for us to understand variable needs to ensure our workflow will accommodate the summarization of needed variables.
- What data formats are you most comfortable working with for scientific analyses (e.g. excel, csv, shapefile, raster, netcdf / hdf, web service...)?
- For your research, is it important to account for the percentage of the summarization spatial unit that does not have available data?