

**Report of Data Maintenance Workshop  
In support of Water Data Initiative  
January 7, 2016**

**Summary**

**Initiative Goal:** *Development and widespread use of a comprehensive and consistent surface water database for the Commonwealth:*

- **Objective** – Designed as default and **authoritative surface water basemap** for the Commonwealth.
- **Objective** –Spatially compatible with **current mapping** such as PAMAP imagery and LiDAR.
- **Objective** –Includes man-made **stormwater infrastructure and other ephemeral flow features**.
- **Objective** –Provides **reference basis** for regulatory and analytical studies, modeling and data.
- **Objective** –Serves as the basis for **data catalog** applications and **regulatory simplification**.

The very process of creating a singular water base map would serve as the essential first step to improve management and regulation of the water resource by clarifying just who is collecting which data. Furthermore, the mandate to establish and maintain such a unified database is already indicated specifically in Section 3117 of Act 220 of 2002 – the State Water Planning Act.

An earlier workshop (December 2014) aimed at defining possible approaches and estimating costs strongly suggested that no new map be created without having the maintenance plan and protocols for the new map designed and generally accepted. This workshop addresses that need.

**Report Contents**

- I. **Purpose**
- II. **Workshop Design**
- III. **Discussion and Recommendations**

**Appendix A – Workshop Participants**

**Appendix B – Discussion Breakout Summaries (by Group)**

**I. Purpose** – Develop a preliminary approach to data integration to provide a complete picture of the surface water resource, with a focus on maintenance schemes for continual updates.

**II. Workshop Design** – The group spent the first two hours reviewing a set of information and results from past and current projects upon which to base their later discussions. The presentation generally described and illustrated the current and possible future water base maps as follows:

**Present Base Map Characteristics:**

- High Resolution National Hydrography Dataset (NHD) contains primarily perennial features
- Backward-looking; relates primarily to legacy program data
- Based on 1980's Digital Line Graph data and 50 feet horizontal tolerance
- Limited digital integration
- 2D Flowline

**Desired Base Map Characteristics:**

- Represents perennial, intermittent, ephemeral features; headwaters prominent
- Forward-oriented; encourages planning and scenario testing
- Incorporates wetlands
- Includes stormwater systems
- Greater positional accuracy; compatible with modern base maps and location technologies;  $\leq 5$  feet horizontal tolerance
- 3D segmentation
- Maintains NHD reach codes; supports legacy needs
- Pennsylvania-specific, but reflecting regional and federal needs

**Major recommendations from the December 2014 budget workshop included:**

- Pilot project in Lancaster County recommended to better understand and establish the scope of the project.
- Initial budget must include both technical and marketing components; the group did not recommend any permanent funding stream
- Technical and programmatic recommendations included:
  - Maximize the use of "ancillary data/existing data"
  - establish a *minimum* data production standard to ensure regulatory fairness and maximum use of existing data, not a singular production method
  - Examination of some simplified data model(s) that might support later conflation to NHD
  - No statewide project should be undertaken (pilots and project scoping excepted) without maintenance and update defined.
  - Crowdsourc (volunteer-generated information) should be a component of the maintenance plan.

**Breakouts** – After the presentations attendees broke into **four** teams of **3** individuals and worked independently to:

- Discuss the material presented in terms of four alternatives:
  - Data Model? **Comprehensive or Simplified Database**
  - Data Stewardship? **Multi-agency vs. Single Steward**
  - Data precision? **Schematic vs. Measurable**
  - Map purpose? **Regulatory vs. Planning/Multi-Purpose**

- Decide which combination makes sense looking forward
- Summarize the approach and document the key points of your discussion.

**Assumptions:**

- Watershed boundary dataset integration is needed
- Set as a priority through the Geospatial Advisory Board
- Collaboration by multiple agencies and organizations is expected
- Maintenance will be a collaborative effort by state and local governments
- If not an Enhanced NHD version, must relate efficiently

**Discussable Assumptions:**

- Cost is a factor but not a limitation
- Complete by 2019

**III. Discussion and Recommendations –**

**Summary Conclusions**

The workshop participants certainly did not agree on all aspects, however the summary opinion was that we should have a surface water database and base map with the following characteristics:

Begin with a **simplified database** of the best accuracy we can afford and sufficient attributes to link to existing regulatory datasets, and taking advantage of technology advances to **add complexity** and connections in the near future. It was unanimous that there should be a **single steward** managing a system that allows **multiple contributors** to add to the official database; the idea of citizen scientists contributing and of crowd-sourced event data was a common expectation. It was also unanimous to require **measurable** data precision more compatible with modern imagery and data collection tools. Finally, in spite of the fact that regulatory needs drive current data management, that a **multi-purpose** surface water database and base map that supports modeling and engineering is possible and advisable.

**Appendix B** contains the actual notes captured for each Group as they presented the results of their deliberations. This section combines main ideas from the various Groups for each particular decision point.

The **Comprehensive or Simplified Database** decision point was not unanimous, as seen by the summary:

- |         |   |
|---------|---|
| Group 1 | Lines only w/ data keys (perhaps not even polygons at first)              |
| Group 2 | Geometry-oriented with limited attributes                                 |
| Group 3 | <b><i>Start with geometry and basic attributes; evolve to complex</i></b> |
| Group 4 | NHD connection to comprehensive DB  |

Participants were not asked to design the database nor was that their expertise. Discussions instead reflected the fact that we maintain many databases that are only loosely connected and that do not update each other in any meaningful way. There was broad agreement of the need for continual updates and for a single surface water geometry utilized and supported by all who maintain water data. There was also general agreement that evolving technology should allow increasing complexity and

connections among data sets even with a simple geometric network as the root. Crowd sourcing data collection was seen as a way to engage the public in care of the water resource, and a model built to allow and validate inputs from multiple agencies would lend itself to that as well.

Participants were unanimous that a **Single Steward** is superior to Multi-agency control, as seen in the summary comments:

- Group 1 One manager and several gateways to get the data in
- Group 2 One keeper with multiple support/validation
- Group 3 Sole steward (ensures accountability); DEP recommended
- Group 4 Sole steward, funded

The detailed notes support multiple contributors updating data through a portal where changes are validated before incorporation. Reasons for a single steward often centered on continual (stable) funding. One can see that Group 3 recommended DEP (and the State Water Plan requires it) but others suggested that a private or semi-public entity specific to the task be investigated.

It was within this topic that participants identified the need for an inventory of existing data sets and their stewards as a first step. Another common theme was the need for a marketing and business plan for creating better water data; this topic sprang from the need for stable funding and given the diversity of interests already maintaining their own data in their own databases.

Groups were unanimous too on the importance of graduating from Schematic stream networks that show relationships and stream order very well, but are incompatible with the accuracy of current maps and measuring systems. We used the term **Measurable** in our discussions, and the Group comments included some specifics to this unanimous conclusion:

- Group 1 Precision to ensure compatibility with topo
- Group 2 Measurable precision (1:2400)
- Group 3 Based on best available (PAMAP is statewide); inform with catalogue sources
- Group 4 Measurable; but related to funding

Discussions touched on cost and the fact that better accuracy costs more, but also that inaccurate data provokes inefficiencies and resultant costs. Participants noted that we have spatially accurate data from multiple sources (Counties were mentioned most often) which we have not yet taken advantage to improve base mapping. Measurable features allow modeling and consistency of analysis, particularly if we include elevation as a feature.

The concept of measurable data includes compatibility among maps depicting the many networks we rely on in PA. For example, road networks are currently more accurate than surface water, so mapping culverts and pipes where the two intersect is difficult, and we rely more on road data than water data to find them. Less obvious is the fact that the current NHD database does not coincide with the statewide contour data so that assessing steepness of a certain stream reach is not possible from that data (but could be). A different example of incompatibility is that even when we gather field data about disturbances (e.g. -pipe crossings or wetland locations), the spatial deficiencies make it impossible to incorporate the good data into the existing maps.

Discussion groups did not reach a common conclusion regarding whether maps should be **Regulatory or Planning/Multi-Purpose**:

- |         |  |
|---------|--|
| Group 1 | <b><i>Geometry/Planning, since regulations change; same data for all</i></b> |
| Group 2 | Both   |
| Group 3 | -----  |
| Group 4 | Regulatory, because links to funding   |

This perhaps reflects the fact that we currently maintain regulatory maps funded by specific programs and which were made for specific reporting needs; for example. NHD reach identifiers are required for EPA reporting. Other discussion points in the workshop tended to favor a multi-purpose map since the single-purpose program maps are not coordinated and will not support modeling.

In reality, many factors were common to all the decision points:

- Topography, geology, roads, culverts, stormwater systems and other infrastructure all inform and improve our understanding of water data *when they are spatially compatible*.
- NHD will generalize for each needed scale, and we need different levels of complexity and detail depending on the application, but *many applications are unavailable at current map scales*.
- *Data stewardship needs budget*, but so does contribution and maintenance of data.
- The data model design and the stewardship model are entwined. *A forward-looking approach that engages citizens* is needed.
- *Marketing is needed to increase awareness* of the complexities and requirements for modernized water data.
- *Key champions can be found* in the Legislature, Counties, Municipalities, State and Federal agencies, and private enterprise.

Report by:

Eric Jespersen  
President – PaMAGIC 2013-2016  
[ecj@epix.net](mailto:ecj@epix.net)  
717-475-7869

## Appendix A– Workshop Participants

### **Moderators**

Mike Bialousz *	DCNR GIS, Harrisburg
Eric Jespersen *	PaMAGIC, Drums

### **Group 1**

Eric Moore *	West Branch Regional Authority, Muncy
Rich Augenti	Hanover Engineers, Bethlehem
Mike Goeckel	Self-employed, Marysville

### **Group 2**

Tom Denslinger *	Self-employed, Harrisburg
Marcie Dunn *	NRCS, Harrisburg
Patrick McKinney	Cumberland County GIS, Cumberland

### **Group 3**

Karen Feather	Lebanon Valley College, Annville
Glenn Mohler	Lancaster County GIS, Lancaster
Andrew Sechrist	Student – HACC, Lancaster

### **Group 4**

Scott Hoffman *	USGS Water Science Center, New Cumberland
Andy Pickford *	BAE Systems, Mount Holly NJ
Gail Jackson	Pennsylvania DEP, Harrisburg

### **Observer**

Dave Goerman	Pennsylvania DEP, Harrisburg
--------------	------------------------------

*\* denotes participants of both the December 2014 and January 2016 Workshops*

## Appendix B - Breakout Discussion Summaries

### Group 1 –

- If “comprehensive” this is a statewide basemap – not just surface water and topography, but also roads and infrastructure which all impact hydrology (*At least ensure compatibility with all pertinent features*)
- Simpler is better – limited attributes but as good an accuracy as we can get (*hydro portion of larger state basemap*)
  - Different levels of complexity and detail depending on basemap layers
  - (*Many*) Data key(s) to join additional attributes
  - Polygons for some streams of a certain width, with scale dependencies
- Hosting – input from the masses (“Open Streams Map”) with one manager and several gateways to get the data in
  - Gateways – universities, agencies, etc. with qualifications
  - Dynamic and web-based approach (dash boarding)
    - **Updates are continual** – not batch per se
    - Use a web service to post data to users
    - If it is easy to upload data, people will provide data, need QAQC oversight
    - Funding is minimal, only need to pay for hosting and QC
    - Need to create archived hydro layers
  - **Consider private sector involvement (management)** who don’t necessarily have restrictions that state agencies do and have more flexibility
    - Scenario – DEP manager overseeing a consultant doing most of the heavy lifting and leveraging gateway organizations to get the data in
  - Trunk and branch approach
- Accuracy – reliant on cost
  - Line level is priority and the start; polygons eventually (i.e. islands etc.)
- Regulatory vs planning – trunk and branch approach here too where regulatory is a branch as it is very dynamic/changing and very impactful (build it off the trunk)
- Blending County data will be challenging as there are major differences spatially and with schemas → need to **survey Counties** first before deciding on this approach
  - Blank slate approach would be very expensive but could be less work than blending County sources? *Probably need a hybrid approach.*
  - Lancaster Pilot project ties into this as it will hopefully be a model to look at doing LiDAR-based hydro generation statewide
- Private sector will pay for data, if it is a reasonable cost – i.e. - “Official data” has value. Fund it using a subscription based system, annual or monthly.
- Boundary issues between counties affect edge-matching
- MS4 should be a possible funding source for this effort, municipalities need quality data to support their mapping efforts.
- Need to **develop a business and marketing plan for this project.**

- Pipeline Task Force should have a need for quality data. Report identifies many issues that require Geo spatial solutions.
- Group talked with DEP about quality data needed around water sources.
- Recreational and environmental enthusiasts could be interested in crowd sourcing.
- PA Wilds, connect it to that program in some way.
- Report from GeoBoard

#### Group 2 –

- Start with geometry and limited attributes
  - Must be tied back to NHD
  - Cost a big factor in this, but geometry affordable to start
  - Agencies and organizations build onto this dataset over time
- Stewards → **needs to be multi agency input with one data keeper**
  - Too much for one agency; others field verify or add
  - NHD steward needs to work closely with WBD steward (*or same?*)
- 1:2400 scale good
  - **Good for modeling and consistency** and measurable
- This should be able to serve both regulatory and planning uses
- NHD Wet tool in Alabama – citizens inputs data into a database, QA/QC at state and then fed into national NHD. Good citizen engagement. Follow up on this

#### Group 3 –

- Education and **Funding** focus
- Need to start with geometry base with simple attributes
  - Early successes
  - **Evolve** eventually to NHD data model
  - Sole steward is necessary (DEP); ensures accountability
- Precision and accuracy → get best available data (whatever it is) and keep metadata about all the sources specifically (feature level metadata)
  - Complex approach
- Funding
  - Education is key – general public too, legislators and decision makers
    - Need some thought into effective strategies in terms of the approach to these different groups
  - **Crowd sourcing data collection for certain aspects → good way to connect with and engage public**
    - Use **incentive programs** to make this work
    - Regional based topics (i.e. rec uses, flooding, water quality, etc)
      - *Perhaps related to State Water Plan regions*
  - User fees – variety of possible mechanisms (grants (early on), R.E. transfer tax, etc)
- Need a “project champion” potentially in legislature and/or in other organizations
  - Need 3-4 clear/concise **talking points to aid in “selling” the program**



- Need support of Counties (key aspect) and make this as local as possible
  - Challenge – resources (staff and tax bases) are extremely varied by County or within Counties – could result in additional costs to outsource the effort in some cases
    - Lancaster – has similar issues with municipalities. To help, County provides parcel data to municipalities in exchange for data back (gives them a starting point for this GIS too)
    - Need to consider the approach to various Counties especially in terms of making this valuable to Counties of all economic levels and also feel they are part of the process
      - Need **good marketing** – consider bringing in a marketing person/group to the initiative
  - Challenge - Currently a **wide variety of hydro data sets by County**
  - Challenge - Political support is varied

#### Group 4 –

- Single line geometry with at least an attribute link to NHD
- **Counties are likely best government entities to provide best local level/on the ground data**
- Database host/resources in the cloud
  - Influences who the steward or gate keeper is
  - We cannot continue with a single *person* steward
    - i.e. PAMAP was Penn State with other organizations providing funding and support
  - **Funding stream needs to be stable** to whoever it is
- Tighter accuracy needed but tied to funding as cost goes up exponentially with increased accuracy; base accuracy not selected – very tied to cost
- **Timing is critical variable especially with regard to funding sources** (fiscal year – whose?)
- **Need an inventory of existing data sets and their stewards**
- Press releases for **awareness**; also list serve type distribution
  - Need more “modern” approaches to **awareness** too (i.e. social media) and we need an expert in leveraging these (“content strategist”) – **also requires funding**
    - Look at **university support for marketing** with statewide support
  - Need a path forward first and the marketing comes
- Assess each Counties data *and* willingness to share
  - Identify County data stewards with this
  - Data pitfalls and voids identified
  - Requires funding partners and sources too
- Steps:
  - Inventory
  - Compile and check
  - Identify pitfalls and voids
  - Identify funding partners