ERDC/ITL TR-20-4

Information Technology Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Data Enrichment and Enhanced Accessibility of Waterborne Commerce Numerical Data

Spatially Depicting the National Waterway Network

Timothy W. Garton

December 2020

ERDC/ITL TR-20-4
December 2020

Data Enrichment and Enhanced Accessibility of Waterborne Commerce Numerical Data

Spatially Depicting the National Waterway Network

Timothy W. Garton

Information Technology Laboratory
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Final report

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000

Under Coastal Inlets Research Program,
Funding account code U4357241
AMSCO code 060000

Abstract

This report provides methodologies and processes of data enrichment and enhanced accessibility of Waterborne Commerce and Statistics Center (WCSC) maintained databases. These databases house tabular and statistical data that reports on The U.S. Army Corps of Engineers (USACE) Civil Works Division National Waterway Network (NWN), which geospatially represents approximately 1,000 harbors and 25,000 miles of channels and waterways.

WCSC is a division of The Institute for Water Resources (IWR). They have been tasked with the international collection, maintenance, and archival of all records involving commercial movements and commerce that occur on federal waterways. The current records structure is a large, tabular dataset and limited to the systems and processes put in place prior to the computing standards and capabilities available today. Methods have been tested and utilized to bring the tabular datasets into an optimized, modern geospatial network and expanded upon to create a higher resolution than previously maintained by the WCSC.

This report will expand upon the applied methodologies to optimize data queries and the overall enhancement of the data system to allow for linkages to various other sources of information for commerce data enhancement for decision support assistance.

Contents

Figures and Tables

Figures

Tables

Preface

This report is a deliverable product under the Coastal Inlets Research Program (CIRP), which is funded by the Operation and Maintenance (O&M) Navigation business line of the Headquarters, U.S. Army Corps of Engineers (HQUSACE). The CIRP is administered for Headquarters by the U.S. Army Engineer Research and Development Center (ERDC), Coastal and Hydraulics Laboratory (CHL), Vicksburg, MS, under the Navigation Program of HQUSACE.

The work was performed by efforts of the Scientific Software Branch (SSB) of the Computational Science and Engineering Division (CSED), ERDC, Information Technology Laboratory (ITL), Vicksburg, MS.

At the time of publication, Mr. Timothy W. Dunaway was Chief, SSB; and Dr. Jerrell R. Ballard, Jr. was Chief, CSED. The Deputy Director of ITL was Ms. Patti S. Duett and the Director was Dr. David A. Horner.

COL Teresa A. Schlosser was the Commander of ERDC, and Dr. David W. Pittman was the Director.

1 Introduction

Background

Archived statistical data

The U.S. Army Corps of Engineers (USACE) Civil Works mission is to provide safe, reliable, efficient, and environmentally sustainable waterborne transportation systems of channels, harbors, and waterways for movements of commerce, national security needs, and recreation.[1] The systems comprise approximately 1,000 harbors and 25,000 miles of channels and waterways. The USACE Institute for Water Resources (IWR) has multiple technical centers that are tasked with tracking the operations of these systems. The IWR Waterborne Commerce Statistics Center (WCSC) is the technical center that collects, processes, distributes, and archives vessel trip and cargo data.[2,3] The U.S. Army Engineer Research and Development Center (ERDC), Information technology Laboratory (ITL) provides mechanisms for analyzing the complex data structures to compare and contrast statistics in order to assist with annual budget creation (Mitchell 2010, 2012; Khodakarami et al. 2014; Mitchell and Skully 2014).[4,5]

WCSC processing is continuously on the move providing support for both government agencies and the public along with annual regulatory publications. The vigorous environment has limited time and opportunity for altering the current, stable process to utilize more modern, advanced technologies, which has created a dataset that is constrained by systems and methods that have been in place for many years. These constraints have created archives of tabular data that are not easily or dynamically processed into individual areas of interest. Limitations also exist in projecting the statistical analysis of channels onto the National Waterway Network (NWN) that geospatially represents channel definitions as set forth by design specifications and congressional approval. Publicly releasable subsets of the processed data are made available through the

[1] http://www.usace.army.mil/Missions/CivilWorks/Navigation.aspx

[2] http://www.navigationdatacenter.us/index.htm

[3] https://www.iwr.usace.army.mil/

[4] https://www.erdc.usace.army.mil/

[5] https://www.erdc.usace.army.mil/Locations/ITL/

IWR Navigation Data Center (NDC) at http://www.navigationdatacenter.us/index.htm. These datasets are released annually with publications that provide a pre-defined, high-level summary of the network throughput. These publications are essential for everyday public consumption, but are not sufficient when trying to develop low-level methodologies to assist in channel effectiveness, rankings, and budget formation for USACE representatives.

Custom reports and exports are available by request through the WCSC. These requests must be made in advance to allow time to compile the data for day-to-day intra-agency business needs of USACE. These requests are also limited to manually defined areas of interest, which may also require that a dock level search be done instead of a reach level search since many projects are defined by a single reach.

This hard-coded methodology of disseminating the data creates higher costs for maintenance when project definitions are added or updated and exposes the imminent risk of human error.

Statistical data limitations

The current design structure consists of a shipping and receiving dock, commodity code, tonnage, and route string. The only method for querying custom reports is writing Structured Query Language (SQL) queries that consist of the link number that represents a channel, and can be expanded upon by including the set of docks for the area of interest. These queries utilize full-text indexing of the route string in order to find where the route contains the link numbers of interest. This is inefficient as an individual channel could be represented by a link, SD Link, or Super Link resulting in the need to query the route string for multiple string contain conditions. Also, the docks are not always publicly accessible due to privacy concerns, so a specific location may not be defined for a customer. This high-level manner of database searching creates problems when an individual wants to know conditions at an exact location in the network. It also limits the reports to only identify movements that dock at a location and do not include transiting commerce that utilizes a channel for throughput, but does not stop at that specific location.

Objective

Network spatial representation

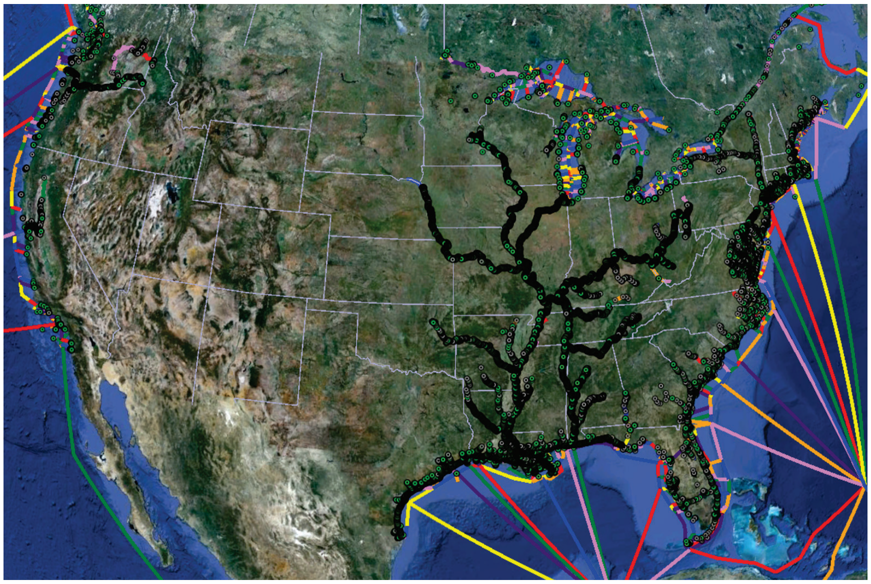
Many routine requests by USACE representatives are needed to easily represent the characteristics of a channel in order to support or defend decisions of the organization. This can easily be done with geospatial representation of the data. Static printouts and pamphlets are available through the NDC annual publications. The ability of any authorized person being able to have access to a dataset that can be utilized to create their own subset of data projected onto a realistic geospatial representation is a powerful tool.

The National Waterway Network (NWN) is maintained and utilized by WCSC and provides a spatial window into the link and project definitions of the entire system. However, the resolution is too low for accurate representation of commerce. The existing network, however, does allow for augmenting to provide a resolution of data that strengthens the USACE justifications.

The final part of the system that all movements depend on is the geospatial representation of docks, which is collected and maintained in Master Docks Plus (MD+). This is posted and maintained by the NDC.

With a high-resolution network and accurate representation of the location of all docks involved (Figure 1), all movements can be broken down into any resolution desired.

Figure 1. NWN and valid docks overlay.



Approach

Augmenting reach definitions

The NWN that is maintained and utilized by WCSC is defined on a one-mile level of resolution.

This is sufficient for long sections of inland rivers where the volume of traffic does not change (Figure 2).

However, this does not reflect a refined enough resolution within harbors or projects that may have many docks creating a significant change in the volume of traffic entering or exiting the network. Also, dissemination of movements to side and sub-reaches of interest that a USACE study may require is often assigned to a single reach in the NWN that may or may not be geospatially correct (Figure 3). This necessitates the ability to augment the reach geospatial definition to include the functional attributes along with more accurate positioning (Figure 4).

Using data sources and methods that will be further explained in this report, reaches have been aligned to proper official geospatial channel definitions that are maintained by the USACE and enhanced to a custom network resolution. The archived data can then be associated to the custom network to be used by tools to provide accurate representation of commerce movements.

Figure 2. Five docks over a 50 mile section of Black River, AR.

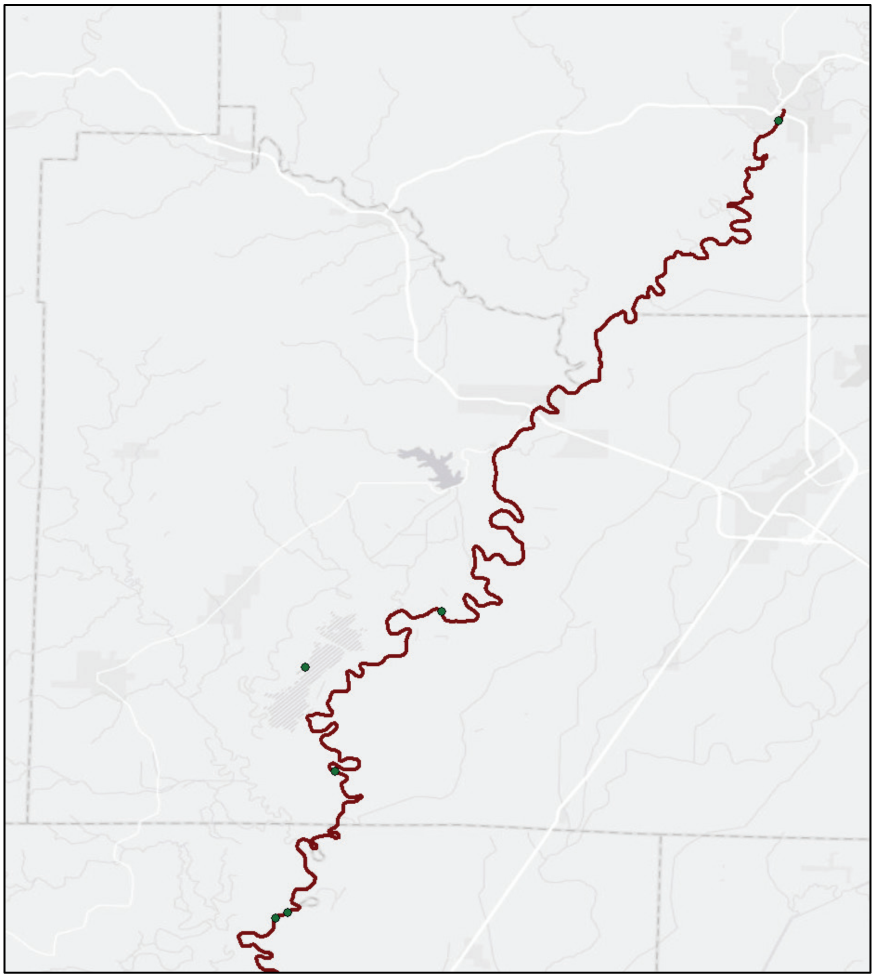


Figure 3. WCSC single reach definition for Milwaukee, WI.

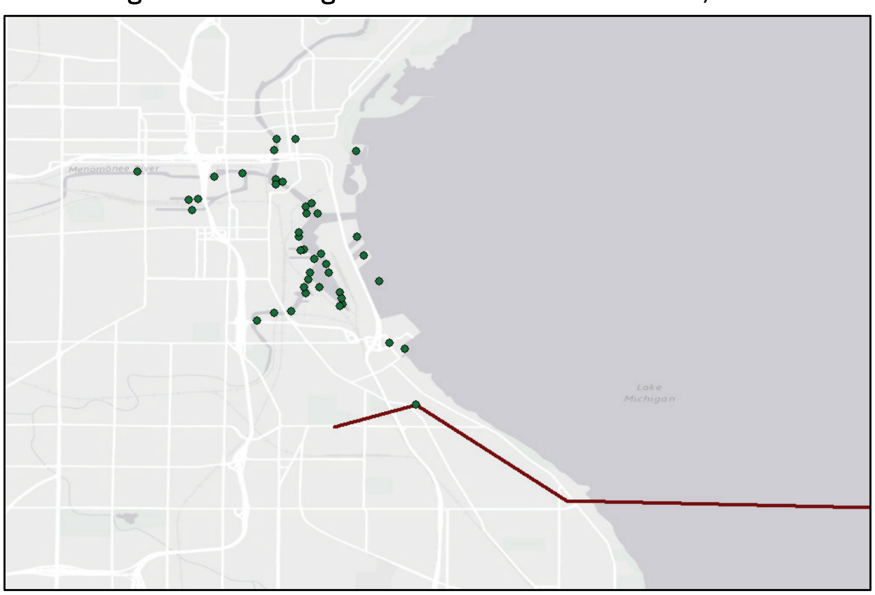
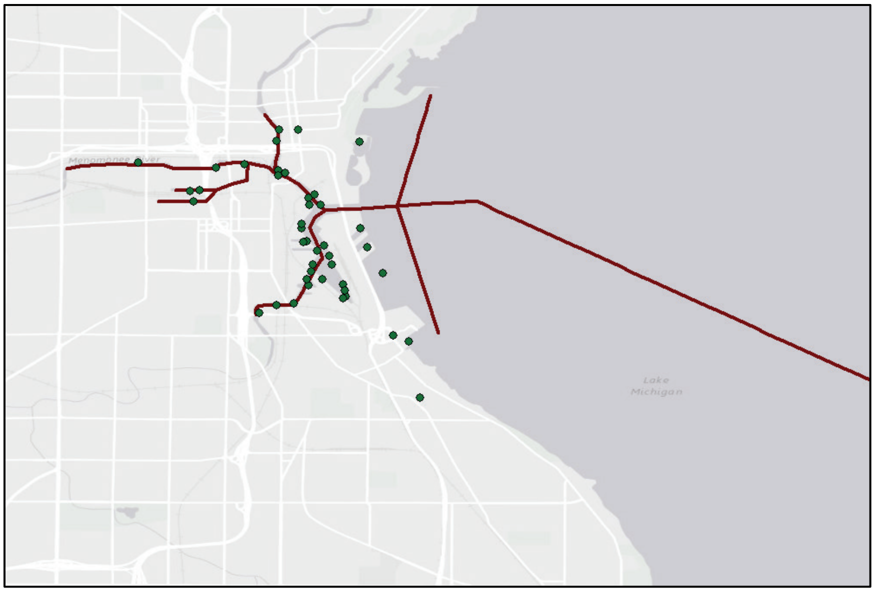


Figure 4. Augmented reach definition for Milwaukee, WI.



Organization of this report

This report is organized into six chapters:

- Chapter 1 introduces the situation at hand and the purpose of the report.
- Chapter 2 provides the philosophy behind how to enhance the current system while maintaining a reference into the raw tabular data.
- Chapter 3 investigates the methodology utilized to process the data into the refined augmented system.
- Chapter 4 provides a use case scenario of the methods employed.
- Chapter 5 summarizes the report and application uses.
- Chapter 6 introduces future work that will be researched for inclusion in the next data enhancement effort.

2 Philosophy

There are three aspects of the system that are essential and assumed to be available and accurate in order to enhance the results that can be accessed. The results will only be as good as the reach definitions that the route traverses and the docks where the movement entered or exited the network. If both are properly defined, any resolution is possible. Valid analysis also depends on the assumption that all movement details are recorded properly in the provided route string and docks associated with any movement.

Reaches

Reaches are the first key to being able to spatially represent commerce. They are officially defined and maintained in the NWN. The NWN is limited by the number of reaches per project location and the resolution of coordinates that geospatially represent the reach. The number of reaches and the resolution of the coordinates can both be increased to provide a better geospatial representation of commerce.

Reaches are given a unique Linknum, which associates them with a parent link. This link is the most important part of the reach definition. Movement routes are reported by links. Reaches can be added to the system and assigned the same parent link to associate to the route string and a unique Linknum in order to disseminate the movement throughout the network with well-known shortest path routing algorithms at the shipping/receiving link. Reaches that are added need to have a flag, known as the ignore bit, to represent whether it needs to be considered when calculating throughput when a movement does not stop at the link, but rather transits through the link. This occurs when adding reaches that represent a side channel to a parent link that should not get credited with tonnage that never actually touches the reach, but rather just transits through the other reaches associated to the link.

The resolution of reach coordinates is imperative to the process of snapping docks to the nearest coordinate of a reach. Increasing the number of coordinates creates a system where a true nearest neighbor search can be used to calculate the actual distance traversed on a reach by a movement.

Docks

Docks are the second key to spatially represent commerce. Docks are officially defined and maintained in the MD+. The MD+ is limited by maintenance of the dataset and dissemination of updates in location and contact information. Docks are able to be entered without any spatial information and cannot be mapped, but still associated with movement for tabular results. Most have been provided valid coordinates and the remaining necessary docks for routing can be manually geocoded using known owner or address information.

Docks are given a unique location and dock ID pair, which can be used to associate movements to a dock. The coordinate is also imperative to find the nearest reach coordinate to utilize routing algorithms to determine actual reach traversal of movements at the shipping/receiving link.

Movement details

Movement details are the final key to spatially represent commerce. Movements are transcribed into tabular data tables from cargo manuscripts by WCSC.

Details records show various attributes of the movement, of which the shipping/receiving dock and route string are the imperative part for routing (Table 1). LOC_S and DOCK_S create the unique ID for the shipping dock, and LOC_R and DOCK_R create the unique ID for the receiving dock. ROUTE is the provided route of links that WCSC has recorded.

Table 1. Sample WCSC detail records.

LOC_S	DOCK_S	LOC_R	DOCK_R	ROUTE
73512	10	73013	527	*6161
84940	4	86100	10	-8885>8894
75520	1	73013	535	-6159#6160<6161
14311	249	99999	0	-3350#3239<9999
99999	1	675	122	+9998!3736>3737

The shipping/receiving dock is linked back to MD+ through the unique ID to retrieve the associated coordinate for processing. The only anomaly in dock IDs is when the shipping/receiving dock is 555 or when the shipping/receiving location is 99999. When 555 is reported, this is a "ghost" dock that does not actually exist and is used to consolidate movements. When 99999 is seen, this is a location that does not actually exist and is used for foreign traffic that does not have a domestic shipping/receiving dock to reference.

The route string is divided into five character sections. Each section is associated to a direction (Table 2) and link. The first character is the direction, and the next four characters is the link.

Table 2. Movement direction indicator.

Character	Direction
#	Through Down bound
!	Through Up bound
-	Leaving Down bound
+	Leaving Up bound
<	Entering Down bound
>	Entering Up bound
*	Local Down bound
$	Local Up bound

The link can be associated to either an actual link, super link, or SD link. Groups of links where a movement must pass through multiple links with no break in transiting movements can be represented by a single, super link or SD link. The movement is disseminated to multiple actual links using a master link definition table provided by WCSC (Table 3). This rollup of links is required due to character limitations of the route string storage space at the time the current processing system was instantiated.

Table 3. Sample super link to link relationship.

Super Link	Link
2276	2200
2276	2202
2276	2204
2276	2206
2276	2208

Each link can represent one or more reaches (Table 4). A flag, ignore bit, has been added to the WCSC reach definitions to define reaches that are not to be transited if the link is not part of the shipping/receiving portion of the route string. This allows for side channels that are not currently represented by the NWN to be added, but not credited with movements that did not dock there. They will only be considered when processing the shipping/receiving portion of the route string.

Table 4. Sample link to reach relationship.

Link	Linknum	Name	Ignore
2426	242604	Adams Bayou Channel, TX (Mile 0.0 To Mile 2.3) - 006 Adams Bayou Channel (242604)	1
2426	242605	Adams Bayou Channel, TX (Mile 2.3 To Mile 3.4) - 006 Adams Bayou Channel (242605)	1
2426	242606	Adams Bayou Channel, TX (Mile 3.4 To Mile 9.2) (242606)	1
2426	242607	Cow Bayou Channel, TX (Mile 00.0 To Mile 15.2) - 004 Cow Bayou Channel (242607)	1
2426	242608	Cow Bayou Channel, TX (Mile 15.2 To Mile 20.9) - 004 Cow Bayou Channel (242608)	1
2426	242608	Cow Bayou Channel, TX (Mile 20.9 To Mile 22.3) - 005 Orange Field Turning Basin (242609)	1
2426	242600	West GIWW (Mile 266.1 To Mile 266.5) - 001 Sabine River Channel Section B to Orange TB (242600)	0
2426	242601	West GIWW (Mile 266.5 To Mile 269.2) - 001 Sabine River Channel Section B to Orange TB (242601)	0
2426	242602	West GIWW (Mile 269.2 To Mile 272.5) - 001 Sabine River Channel Section B to Orange TB (242602)	0
2426	242603	West GIWW (Mile 272.5 To Mile 276.8) - 002 Neches River to Sabine River (Section B) (242603)	0

3 Methodology

Identification indexing

Integer indexing versus full text indexing is the primary focus for optimization of query processes on the database and allowing for future data enrichment of the NWN (Windand 2012). By breaking the route string into individual entries in the database with unique Linknum assignments, the database can be optimized to better search for individual areas of interest without knowing what the actual parent link is represented by the route string. One to many mappings will be introduced from the parent link in the route string to these unique Linknum representations for each channel on which commerce exists. The effects of this new breakdown will allow for future data enrichment processes that link the NWN to other systems for future deployment of additional reporting mechanisms as new data is introduced to the system without having to manually search by a string to make the connection.

The idea behind this process is that by creating a single, unique Linknum, there is now the ability to use this as the indexed key that can be referenced by the current and any future datasets that are collected. This will allow for an interface to be developed on top of the new data dictionary that will allow novice users to access the data needed without having to understand the numbering and naming convention of channels or know the specific docks that they are interested in as these connections can all be made via a geospatial search for the nearest neighbor channel in the system. This indexing will be a powerful mechanism necessitating the alignment, augmentation, and resolution attributes discussed in the following sections.

Alignment

Alignment of reaches to USACE official channel definitions can be achieved by merging the NWN with eHydro, a navigation channel condition software portfolio used to record hydrographic survey data for channels.

In eHydro, USACE creates spatial footprints (Figure 5) of channels for survey documentation. These footprints are aligned to the official definitions of the channel and annually maintained by USACE.

Figure 5. eHydro reach definitions for San Diego, CA.

The NWN provides geospatial links for associated channels for recording the movement data. This data is often not geospatially correct (Figure 6) and needs to be aligned for proper analysis (Figure 7). Alignment is achieved by adding, splitting, and moving NWN links in accordance with the eHydro definitions to create a merged network that can accommodate the needs of both programs and provide a crosswalk for commerce and channel conditions. Once aligned, the original NWN definition of reaches (Table 5) is now expanded to append new eHydro representation of reaches (Table 6) to allow for enhanced geospatial analysis.

Figure 6. NWN alignment with eHydro for San Diego, CA.

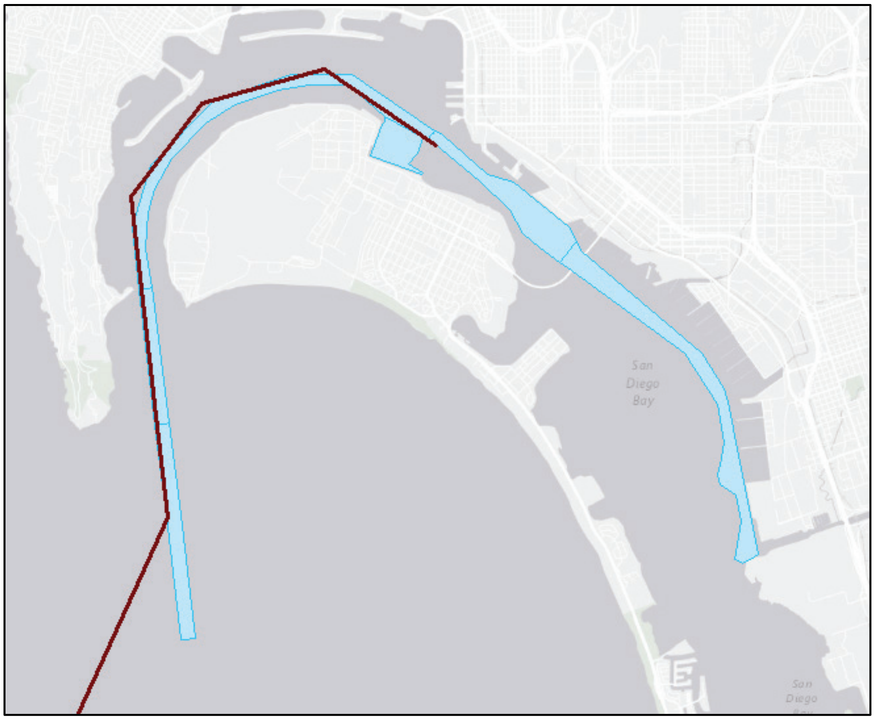


Figure 7. Re-aligned NWN with eHydro for San Diego, CA.

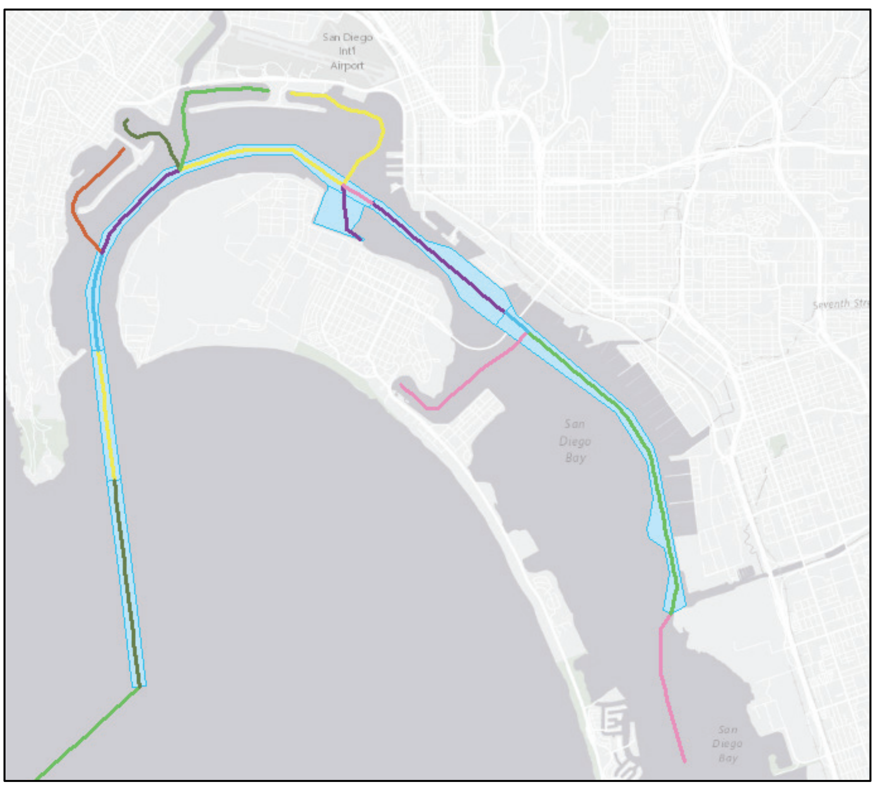


Table 5. NWN link for San Diego, CA.

Link	Linknum	Name
8696	869600	SAN DIEGO HARBOR, CA
8695	869510	PACIFIC DEEP WATER SPINE ACCESS

Table 6. Augmented NWN links for San Diego, CA.

Link	Linknum	Name	Ignore
8695	869500	San Diego Harbor, CA Deepwater Access (869500)	0
8695	869501	San Diego, CA - Approach Channel (869501)	0
8695	869502	San Diego, CA - Entrance Channel (869502)	0
8696	869600	San Diego, CA (Mile 00.0 To Mile 01.1) - North Bay Channel (869600)	0
8696	869601	San Diego, CA (Mile 01.1 To Mile 02.4) - North Bay Channel (869601)	0
8696	869602	San Diego, CA (Mile 02.4 To Mile 04.3) - North Bay Channel (869602)	0
8696	869603	San Diego, CA (Mile 04.3 To Mile 04.7) - North Bay Channel (869603)	0
8696	869604	San Diego, CA (Mile 04.7 To Mile 06.6) - Central Bay Channel (869604)	0
8696	869605	San Diego, CA (Mile 06.6 To Mile 06.9) - South Bay Channel (869605)	0
8696	869606	San Diego, CA (Mile 06.9 To Mile 10.5) - South Bay Channel (869606)	0
8696	869607	San Diego, CA (Mile 10.5 To Mile 12.2) (869607)	0
8696	869608	San Diego, CA - Shelter Island Yacht Basin (869608)	1
8696	869609	San Diego, CA - Americas Cup Harbor (869609)	1
8696	869610	San Diego, CA - West Basin (869610)	1
8696	869611	San Diego, CA - East Basin (869611)	1
8696	869612	San Diego, CA - Aircraft Carrier Turning Basin (869612)	1
8696	869613	San Diego, CA - Glorietta Bay (869613)	1

When reaches are added, they are all assigned the same relative link (Table 5) from the NWN and assigned a unique Linknum (Table 6) for the augmented network.

The additional reaches will be used to associate to appropriate link assignments in the route string to get the movement in the general vicinity of the shipping/destination. Then, the docks will be snapped into the network to find the nearest reach for processing the actual reaches traversed by the movement in the parent link association.

Resolution enhanced

Resolution of coordinates is achieved by calculating how long the line segment is and then how many points need to be added in order to make a maximum custom distance between coordinates. A distance of 528 ft has been chosen for the purpose of this documentation.

By adding the coordinates along the line string, the nearest coordinate for where the movement enters the network has been moved from the first coordinate in the line string (Figure 8), to the middle of the line string (Figure 9) creating a more accurate estimate of distance transited for the movement distance.

Figure 8. Low-resolution snapping coordinate for San Diego, CA.

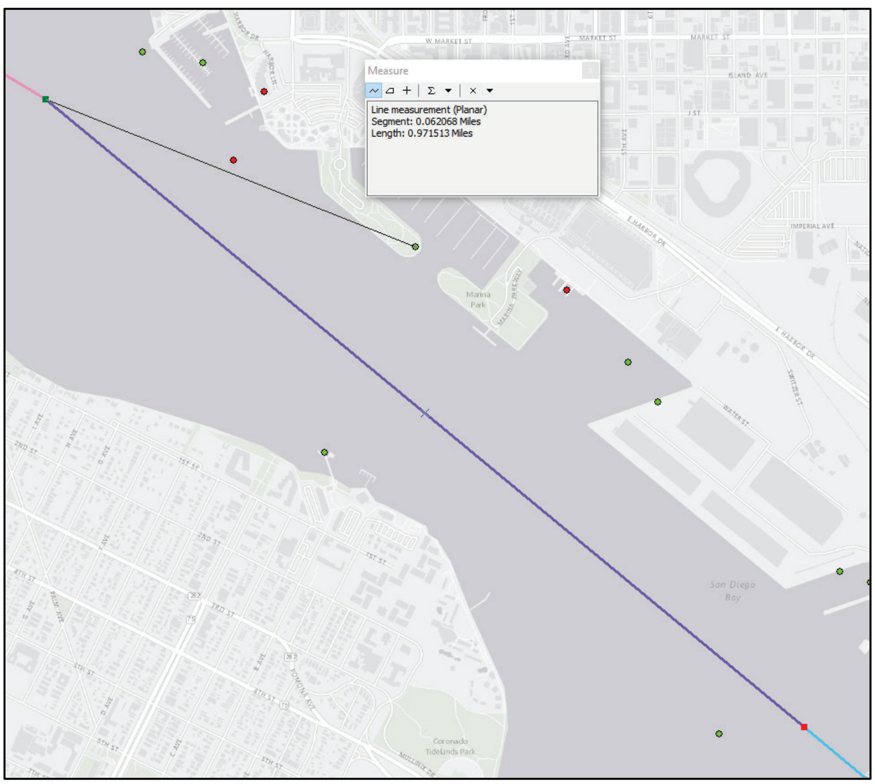
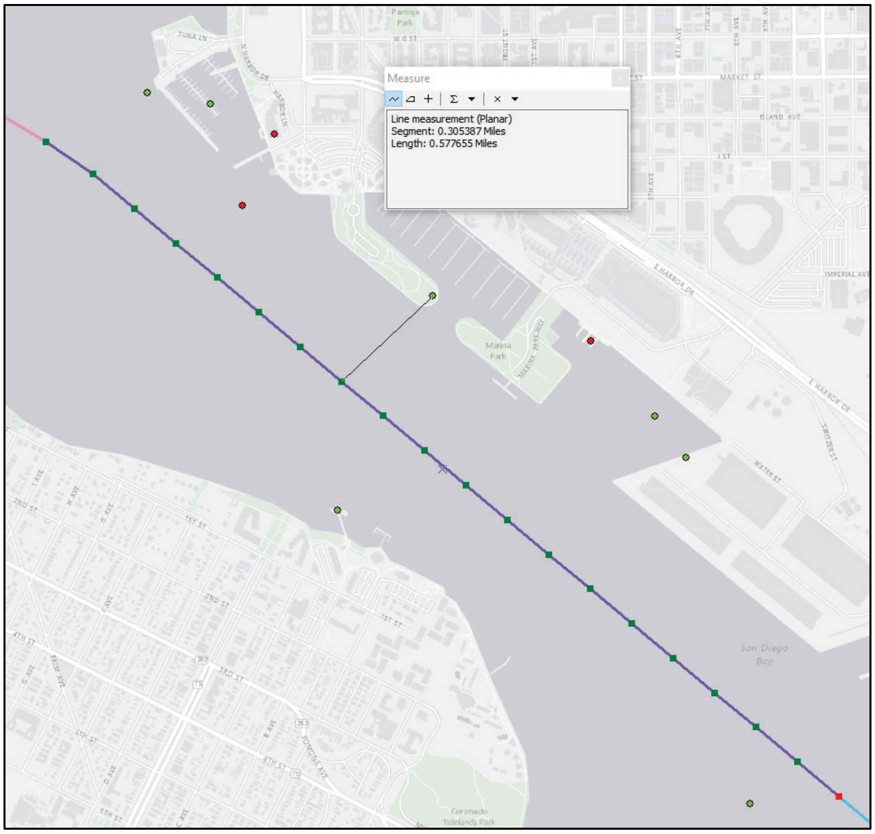


Figure 9. High-resolution snapping coordinate for San Diego, CA.



This increase in resolution allows for a more accurate actual distance traveled, which will be compounded over millions of yearly records and will amount to a more accurate ton mile calculation as a representation of channel segment effectiveness.

Dijkstra shortest path algorithm

Dijkstra's algorithm is a shortest path algorithm that has been utilized to calculate the final reaches traversed on the shipping/receiving link segments of the route string (Lehner and Grill 2013; Rhoades 2017; Hofner and Moller 2012). The algorithm is used to transit from the nearest reach, with the associated parent link, to the shipping/receiving dock to the destination portion of the route string. The network utilized by the algorithm is created in some aspect in every route variation and can represent a single link, multiple links, or nearest neighbor transiting reaches in the network. Each associated reach becomes a node, and endpoints of neighboring reaches are represented by an edge in the network diagram with the weight being the length of the two reaches combined as a worst case scenario.

Single link (5-character route string)

Route strings of only 5-characters are representative of movements that are internal to a single link and never leave that portion of the network. The link number provided is the shipping and receiving link and all reaches associated to that link number will be loaded into an individual network for a Dijkstra's shortest path calculation.

The nearest reach to the shipping and receiving docks will be the start and destination nodes in the network. The shortest path will then be calculated. The shipping/receiving nodes will be recorded with outbound/inbound accordingly and the remaining nodes of the shortest path will be recorded as transiting. If a movement never leaves an individual node, it is recorded as intra-system.

Double link (10-character route string)

Route strings of 10-characters are representative of movements that are comprised of two link numbers. The link from the first 5 characters is the shipping link and the link from the last 5 is the receiving link. All reaches associated with either of these links will be loaded into an individual network for a Dijkstra's shortest path calculation.

The nearest reach associated with the shipping link to the shipping dock will be the starting node in the network. The nearest reach associated with the receiving link to the receiving dock will be the destination node in the network. The shortest path will then be calculated, the shipping/receiving nodes will be recorded with outbound/inbound accordingly, and the remaining nodes of the shortest path will be recorded as transiting.

There are rare occasions where the shipping and receiving node can be the same. This only occurs when a new reach has been added as a side channel and assigned to two distinct parent link numbers. This occurs when there is a natural change in link numbers in the NWN and the new side reach must be represented by either parent link in the general vicinity. In this occasion, the movement never leaves an individual node and is recorded as intra-system.

Multiple link (15-character or more route string)

Route strings of 15 characters or more are representative of movements comprised of multiple links. The link from the first 5 characters is the shipping link and the link from the last 5 is the receiving link. If the shipping/receiving link is 9999, this is considered a bogus link and is representative of deep water where a foreign shipment departed/entered the NWN system. In the case of 9999, no processing is done on the associated shipping/receiving link. The remaining middle links are the transiting links. Routes of 15 characters or more are processed in three distinct steps; shipping link, receiving link, and transiting links.

All reaches associated with the shipping link along with a single reach that is associated with the transiting links and is the nearest neighbor sharing an endpoint with the shipping link are loaded into an individual network for a Dijkstra's shortest path calculation.

The nearest reach associated with the shipping link to the shipping dock will be the starting node in the network. The single reach associated with the transiting links will be the destination node in the network. The shortest path will then be calculated, the shipping node will be recorded with outbound, and the remaining nodes, excluding the final node that is associated with the transiting links, will be recorded as transiting.

All reaches associated with the receiving link along with a single reach associated with the transiting links and is the nearest neighbor sharing an endpoint with the receiving link are loaded into an individual network for a Dijkstra's shortest path calculation.

The nearest reach associated with the receiving link to the receiving dock will be the starting node in the network. The single reach associated with the transiting links will be the destination node in the network. The shortest path will then be calculated. The receiving node will be recorded with inbound and the remaining nodes excluding the final node that is associated with the transiting links will be recorded as transiting.

The remaining reaches associated to the transiting links with the exception of reaches flagged as ignore in their definition are recorded as transiting.

4 Illustration

Alignment

The NWN must be aligned to be able to include locks and various structures of significance.

The channel coordinates must be moved and the lock relocated (Figure 10) to allow for the accurate geospatial dissemination of the movements. Locks are not currently associated with a link and must be assigned a Linknum and link from the associated parent link and is assigned tonnage as if it was a transiting through reach.

Figure 10. Alignment of Ohio River McAlpine Lock and Dam.

Extension

The NWN must be extended to be able to include areas of interest that are currently represented by docks and are not associated to links.

This is done by adding additional channel segments to the network (Figure 11). The extended side channel must be assigned a Linknum and link from the associated parent link and is assigned tonnage as if it is only in consideration for docked tonnage and does not include any through tonnage.

Figure 11. Extension of Detroit River to include Rouge River.

Resolution

The NWN is limited in many locations to have a few and sometime only one reach representing an entire harbor.

These locations must be adapted to properly represent channels of interest that are not currently represented by useful links (Figure 12). The additional channels (Figure 13) must be assigned a Linknum and link from the associated parent link and is assigned tonnage as if it is only in consideration for docked tonnage.

Figure 12. Low resolution of Buffalo Harbor.

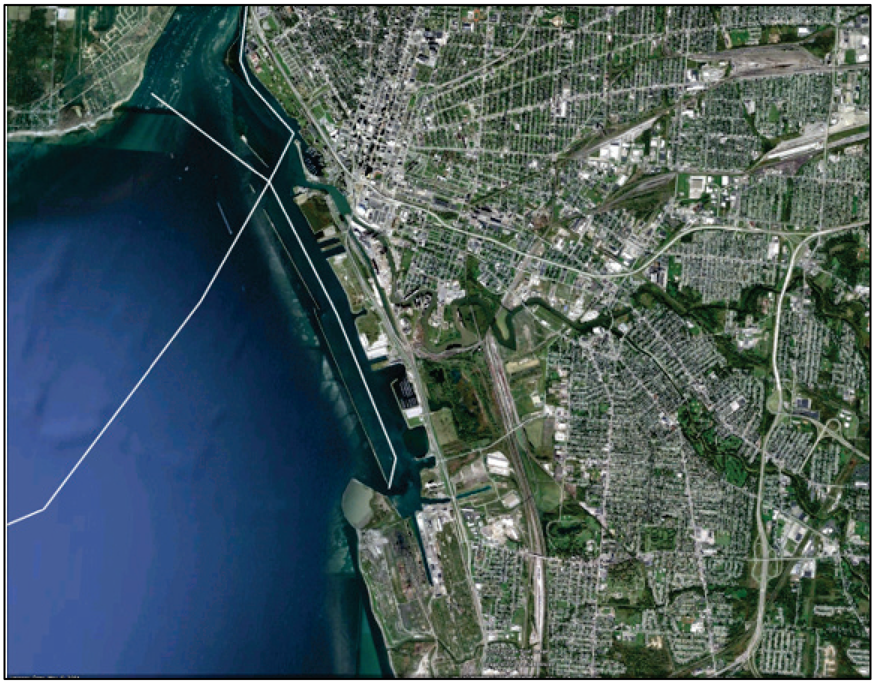
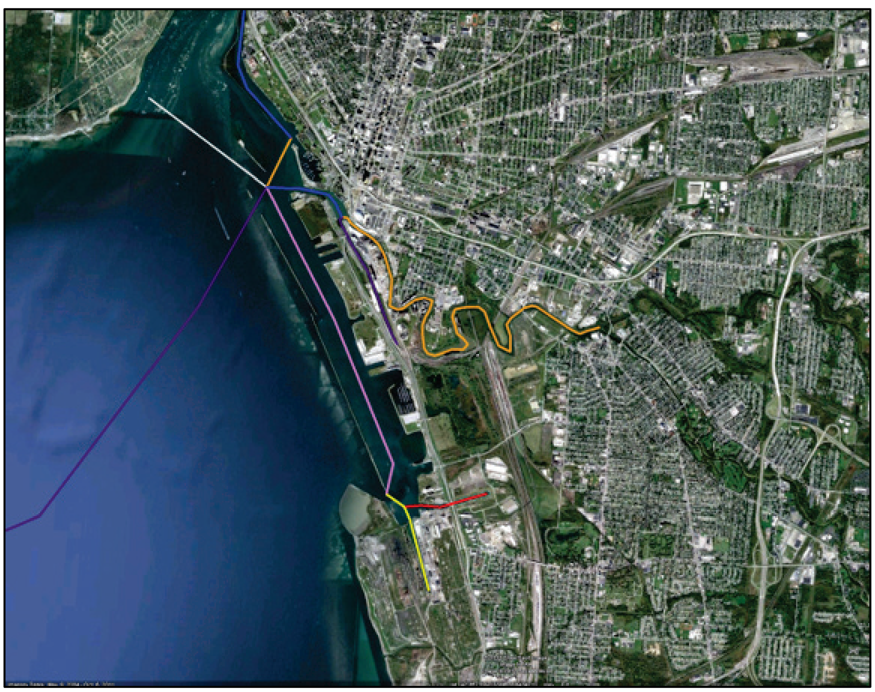


Figure 13. High resolution of Buffalo Harbor.



Route processing

A sample route string will be processed with hypothetical docks showing how links are processed in an expanded network.

A hypothetical route (Table 7) and segment of the augmented network (Figure 14) are considered.

Table 7. Detroit River sample movement.

Route
+6166!6165!6164!6163!6161!6160>6159

Figure 14. Detroit River sample movement.

The first step is to split the route into three sections; shipping, receiving, and through (Table 8).

Table 8. Detroit River sample movement links.

Section	Links
Shipping	6166
Receiving	6159
Through	6165,6164,6163,6161,6160

Shipping link processing

Next, the shipping link is analyzed. This is done by creating a Dijkstra's network of all reaches associated to the shipping link and the first instance of a reach in the through links that shares an endpoint with the shipping link (Figure 15). For this sample, the shipping dock is represented with a yellow pushpin. The shipping Linknum is found by searching for the nearest reach to the shipping dock, which is "616600." The network is then processed through the routing algorithm to find the shortest path between "616600" and the through link "616500." Once the analysis is completed, the path is recorded with the movement for all reaches traversed (Table 9). The reaches removed by the routing analysis and ignoring the through reach that will be traversed in a later step are represented in Table 9 with strikethrough text.

Figure 15. Detroit River sample movement shipping analysis.

Table 9. Detroit River sample movement shipping Linknums.

Link	Linknums
6166	616600,~~616601,616602~~
~~6165~~	~~616500~~

Receiving link processing

Next, the receiving link is analyzed. This is done by creating a Dijkstra's network of all reaches associated to the receiving link and the first instance of a reach in the through links that shares an endpoint with the receiving link (Figure 16). For this sample, the receiving dock is represented with a yellow pushpin. The receiving Linknum is found by searching for the nearest reach to the receiving dock, which is "615901." The network is then processed through the routing algorithm to find the shortest path between "615901" and the through link "616000." Once the analysis is completed, the path is recorded with the movement for all reaches traversed (Table 10). The reaches removed by the routing analysis and ignoring the through reach that will be traversed in a later step are represented in Table 10 with strikethrough text.

Figure 16. Detroit River sample movement receiving analysis.

Table 10. Detroit River sample movement receiving Linknums.

Link	Linknums
6159	615900,615901,~~615902~~
~~6160~~	~~616000~~

Through links processing

The final step is to apply the movement to all reaches associated to the through links that are not marked as ignore.

A side channel has been added to link “6161,” representing the Rouge River (Figure 17) that has the ignore flag set to true, and must be disassociated with the through tonnage since it is never traversed by the movement. The reaches removed by ignoring the through reaches are represented in Table 11 with strikethrough text.

Figure 17. Detroit River sample movement through analysis.

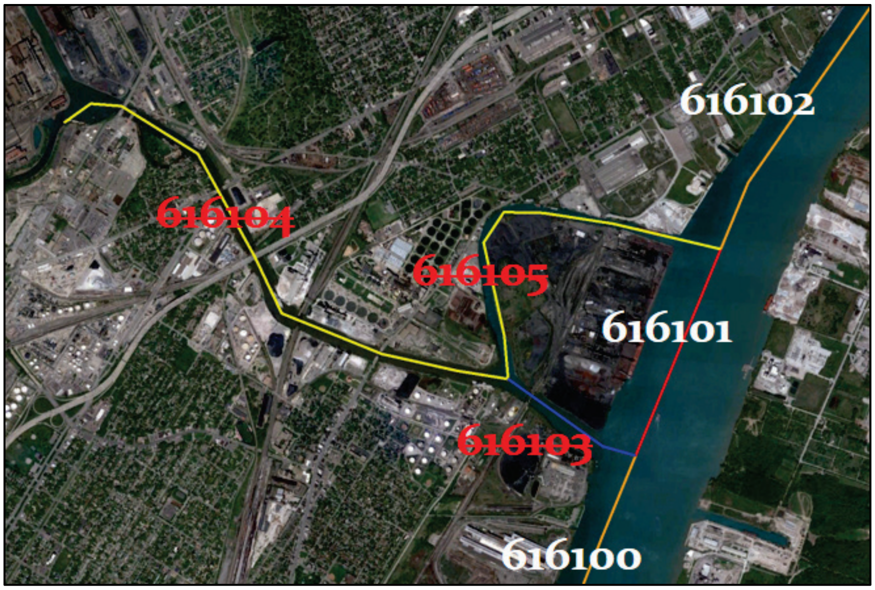


Table 11. Detroit River sample movement through Linknums.

Link	Linknums
6165	616500
6164	616400
6163	616300
6161	616100,616101,616102,~~616103,616104,616105~~
6160	616000

Routing processed

All links have now been analyzed and the appropriate Linknums have been credited with the associated movement. The movement is now increased in fidelity with the association of specific reaches that are refined for specific areas on interest.

5 Discussion and Conclusion

By using methods covered in this report, the route string can be broken down into individual channel movements to allow for quicker querying of the data for an area of interest in the network.

A test of a single project was done to analyze the effectiveness of the now integer indexed reaches versus the original full text indexed rout string.

The test environment was kept constant over all experiments (Table 12). The indexing methodologies introduced reduced the query times for the sample Milwaukee, WI project for a single year significantly (Table 13). The significant advantage gained in the "Entire Project" is due to the new format available since the route string has been broken down into individual Linknums; they can now be rolled up for pre-built project definitions. Utilizing the old, full-text search requires searching for multiple Linknums that represent the entire project in multiple contains conditions of the query. This is only compounded as a query for higher level rollups needed for USACE decision support methods is begun. The speed enhancement is also shown because, now, there are not additional conditions needed to query for individual docks as well as Linknums for individual defined areas of interest as seen in the high and low utilization searches.

Table 12. Test environment attributes.

Attribute	Value
SQL	SQL Server 2014
Year of Interest	2015
Data Entries (Full Text Indexed)	966,177
Data Entries (Integer Indexed)	16,613,045

Table 13. Milwaukee, WI average querying time.

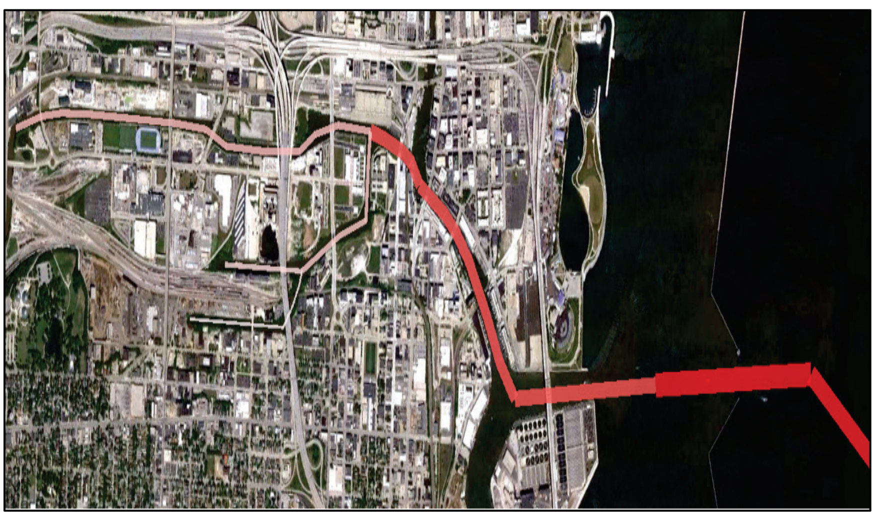
Area of Interest	Integer (s)	Full Text (s)
Entire Project	0.85	10.48
High Use Linknum (60940001)	4.23	15.91
Low Use Linknum (609400011)	0.77	14.83

Some additional advantages of the enhanced process network are that now distance transited is calculated on a more granular and accurate level. The channels can also now be geospatially referenced for plotting movement summaries for an area of interest.

Application suites using this technology of processing the movement datasets are available in the Channel Portfolio Tool (CPT), where movements are used to rank and stack efficiency measures of channels, and Coastal Structures Management Analysis and Ranking Tool (CSMART) where channel and structure condition criteria can be ranked and stacked for comparisons.[1]

Using CPT, flow charts can be created to show the dissemination of movements to the appropriate channel movements within an area of interest (Figure 18).

Figure 18. Flow chart for Milwaukee, WI.



[1] https://www.cpt.usace.army.mil

References

Hofner, P., and B. Moller. 2012. "Dijkstra, Floyd and Warshall meet Kleene." *Formal Aspects of Computing* 24 (4-6): 459-476.

Khodakarami, M., Mitchell, K. N., and B. X. Wang. 2014. "Modeling Maintenance Project Selection on a Multimodal Transportation Network." Transportation Research Record, 2409(1), 1–8. https://doi.org/10.3141/2409-01.

Lehner, B., and G. Grill. 2013. "Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World's Large River Systems." *Hydrological Processes* 27 (15): 2171-2186.

Mitchell, K. N. 2010. "*A review of coastal navigation asset management initiatives within the coastal inlets research program (CIRP); Part 1: Coastal structures*." ERDC/CHL CHETN-III-80. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Mitchell, K. N. 2012. "*A review of coastal navigation asset management initiatives within the coastal inlets research program (CIRP); Part 2: Channel portfolio tool*." ERDC/CHL CHETN-IX-29. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Mitchell, K. N., and B. Skully. 2014. "Waterway Performance Monitoring with Automatic Identification System Data." Transportation Research Record Journal of the Transportation Research Board. 2426. 20. 10.3141/2426-03.

Mitchell, K. N., B. X. Wang, and M. Khodakarami. 2013. "Selection of Dredging Projects for Maximizing Waterway System Performance." Transportation Research Record, 2330(1), 39–46. https://doi.org/10.3141/2330-06.

Rhoades, M. 2017. *Network Analyst: An Introduction.* http://proceedings.esri.com/library/userconf/fed17/papers/fed_70.pdf.

Winand, M. 2012. SQL Performance Explained: Everything Developers Need to Know about SQL Performance. Austria: Markus Winand.

Unit Conversion Factors

Multiply	By	To Obtain
Feet	0.3048	meters
miles (U.S. statute)	1,609.347	meters
tons (2,000 pounds mass)	907.1847	kilograms

Acronyms and Definitions

Term	Meaning
CHL	U.S. Army Corps of Engineers, Engineer Research and Development Center, Coastal Hydraulics Laboratory
CPT	Channel Portfolio Tool
CSMART	Coastal Structures Management Analysis and Ranking Tool
eHydro	U.S. Army Corps of Engineers, Enterprise Hydro survey
ERDC	U.S. Army Corps of Engineers, Engineer Research and Development Center
ITL	U.S. Army Corps of Engineers, Engineer Research and Development Center, Information Technology Laboratory
IWR	U.S. Army Corps of Engineers, Institute for Water Resources
GIS	Geographic Information System
MD+	Master Docks Plus
NDC	U.S. Army Corps of Engineers, Institute for Water Resources, Navigation Data Center
NWN	National Waterway Network
USACE	U.S. Army Corps of Engineers
WCSC	U.S. Army Corps of Engineers, Institute for Water Resources, Waterborne Commerce Statistics Center
Link	National Waterway Network numbering schema for reaches which is defined by a 4-digit number and are used to define route segments
Linknum	National Waterway Network numbering schema for subsections of reaches which is defined by a 6-digit number with the first 4-digits equal to the link of the parent reach
Project	Pre-defined arbitrary grouping of channels or waterways
Reach	Pre-Defined channel or waterway or subsection thereof
Resolution	The distance between any two coordinates in the coordinate string and the density of coordinates or channels that represent an area of interest
Route	A collection of link, sd link, and super link numbers to include a direction character that represent the movement across reaches
SD Link	An arbitrary link that is not defined by a geospatial reference that associates multiple required route links that have no logical association to a link number to reduce route string length
Super Link	An arbitrary link that is not defined by a geospatial reference that associates multiple required route links that have a logical association to a link number to reduce route string length

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
December 2020

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

Data Enrichment and Enhanced Accessibility of Waterborne Commerce Numerical Data: Spatially Depicting the National Waterway Network

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Timothy W. Garton

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Information Technology Laboratory
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC/ITL TR-20-4

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Corps of Engineers
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Coastal Inlets Research Program, Funding account code U4357241, AMSCO code 060000

14. ABSTRACT
This report provides methodologies and processes of data enrichment and enhanced accessibility of Waterborne Commerce and Statistics Center (WCSC) maintained databases. These databases house tabular and statistical data that reports on The U.S. Army Corps of Engineers (USACE) Civil Works Division National Waterway Network (NWN), which geospatially represents approximately 1,000 harbors and 25,000 miles of channels and waterways.

WCSC is a division of The Institute for Water Resources (IWR). They have been tasked with the international collection, maintenance, and archival of all records involving commercial movements and commerce that occur on federal waterways. The current records structure is a large, tabular dataset and limited to the systems and processes put in place prior to the computing standards and capabilities available today. Methods have been tested and utilized to bring the tabular datasets into an optimized, modern geospatial network and expanded upon to create a higher resolution than previously maintained by the WCSC.

This report will expand upon the applied methodologies to optimize data queries and the overall enhancement of the data system to allow for linkages to various other sources of information for commerce data enhancement for decision support assistance.

15. SUBJECT TERMS
Waterways – Commerce – Statistics
Harbors – Commerce - Statistics
Data curation
Database searching
Statistical services

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER *(include area code)***
Unclassified	Unclassified	Unclassified	SAR	39	