



Voyager Search
Presents the
NGA Data Depot

A response to the
Disparate Data Challenge
Stage 1

Matt Harrison
Voyager Search
mharrison@voyagersearch.com
703-231-8106



Overview

In response to NGA's Disparate Data Challenge (DDC), [Voyager Search](#) is pleased to present an integrated search solution we are calling the NGA Data Depot. When it comes to national security, the cost of missing information and wasting precious time is too high to bear. We care deeply about solving this problem and are honored to participate in this challenge. Further, we appreciate the opportunity to show what our software, Voyager, is capable of accomplishing through the clear lens of a competition and demonstration. This document describes how we approached the problem and we encourage you to watch our [video](#) (<https://youtu.be/tztPjYswlCY>) and then explore the [DDC demo site](#) (<http://voyagerdemo.com/ngaddc1/>). Access instructions and a report of the formats that we have indexed have been provided through the submission process.

The screenshot shows the Voyager Search interface. At the top, there is a search bar with the placeholder "What are you looking for?" and dropdown options "Within" and "Where is located?". Below the search bar, there are two tabs: "Recent searches" and "Saved searches". Under "Recent searches", there are entries for "earthquakes", "bathymetric contours", and "earthquake intersects". The main area features a world map with various data overlays. On the left, there is a sidebar titled "Featured" containing several data items with small thumbnail images and detailed descriptions. On the right, there is a sidebar titled "NGA Data Depot" listing categories like "Web Services", "NITF Data", "Imagery", "Humanitarian Data", "Other", and "Raw Data" with their respective item counts. The bottom of the page has a footer with the word "UNCLASSIFIED".

Homepage



NGA's Challenges

We started this project by examining the content repositories so that we could better understand the content itself. NGA has many systems with a wide array of content types scattered among them. That makes finding the right content like looking for a needle in many haystacks. To compound the issue, the content varies widely as does the metadata associated with each piece of content. Finally, users are required to learn how to use many sites to conduct their searches. This means that relevant relationships are hidden and content is missed. **Voyager has solved these issues providing a single point of search across disparate formats and content repositories.**

NGA is looking for a solution that:

1. Is easy to use and presents results in a cohesive, understandable way.
2. Provides contextually relevant information.
3. Can search everything at once providing global utility.
4. Is fast and flexible, accounting for less than perfect inputs, but does a good job with specialized content.
5. Offers a cost-effective alternative to custom solutions but which also operates in secure environments.

Voyager's Approach

Voyager addressed all five of these challenges. The content ranged from file based data to web services to web pages. Voyager indexed these assets and their metadata quite easily. When we indexed HTML pages, we also harvested any files we found, stored them locally and added those things to the index.

Then we examined the results. To help make it easier for users to search and filter, we decided to enhance the index with labels — Content Category, Data Description and Data Link. These labels can be found as filters and reflect NGA's content repositories and groupings. We added steps to our indexing pipeline to accomplish this but did not modify the source content in doing so.

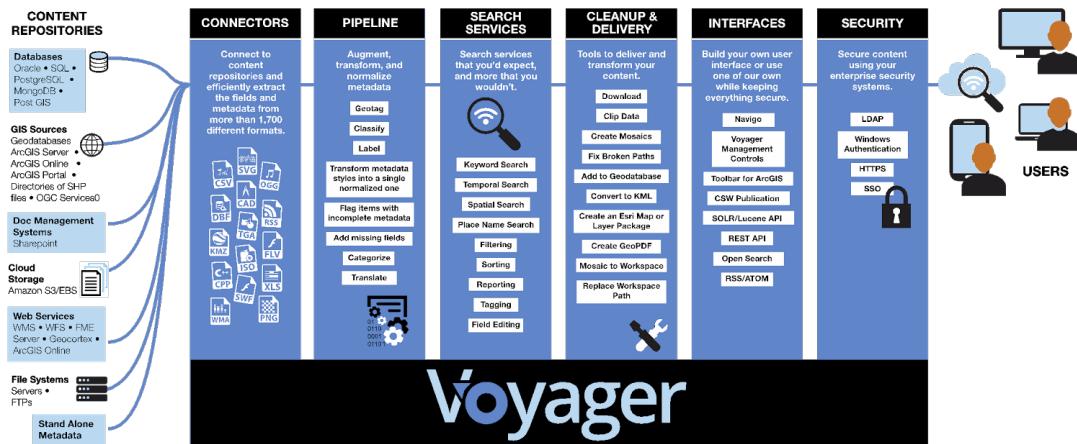
When we were satisfied with the results, we configured the homepage. The “NGA Data Depot” section contains links to pre-defined searches and the “Featured” items show the range of content indexed. We selected the most valuable fields to show on the search results pages so that users could scan the results and use filters in the sidebar to drill down finally, we added all of the metadata about each item to the

Voyager

detail pages. In the header you will see “Saved Searches” that allow users to jump between the content repositories or look for a specific type of content. Users have the ability to create their own saved searches based on their workflows or needs. Saved searches can be shared with other users or specific user groups.

In the time allotted, **Voyager was able to make 27 of the 28 repositories outlined in DDC searchable. The some 200,000 items indexed include almost 100 different formats.** (See: *Data Challenge Table for complete list and counts.*) Both the text from the HTML pages as well as the items contained on those pages have been made searchable using keywords. Items with a spatial reference also support spatial search. When we indexed HTML pages, the related items found on those HTML pages can be discovered under the “Relationships” tab on the “Details page” or simply by searching by keyword.

From Confusion to Clarity and Control with Voyager



Our team was able to do this work quickly and easily because we started with our COTS software that has many of the tools required to deliver the data challenge.

- Extensible connector framework to ingest content source locations
- Pipeline to support enhanced indexing
- Extractors that read 1,800 different file formats
- A configurable user interface called Navigo that make search and discovery easy with little end-user training required
- Visualization tools for previewing map content
- Processing framework to deliver the data



How to Use NGA's Data Depot Site

New users will easily understand how to navigate our user interface. Navigo was inspired by Amazon and other shopping sites. On those sites, users enter keywords to search for what they want. Examine the search results. Refine those options. Select items to view more details. Then, decide what to add to their cart to purchase.

Navigo works similarly. Users enter a keyword into the search box on the home page and they can also run a spatial search entering a place name by querying the map (via point, line and polygon search and buffering). They are then taken to a page of search results with cards (or table or grid view) and an associated map that shows search results and usually a thumbnail of the document item (or HTML page). A map is presented on the left showing the bounding box of any spatial information, but users can also turn on the heat map layer to see concentrations of content that match that query. They can use the filters on the right side to and/or the map drawing tools to fine-tune the results set. Then they can click to view the details for any record in the index. If there is something that they want to get, they can click to download it. If it is a map service, those services can be previewed on the map or opened in a more robust mapping tool like ArcGIS Online.

The screenshot shows the Voyager Search interface. On the left is a map of North America and parts of South America with a red rectangular buffer drawn around the United States and Mexico. The search bar at the top contains the query "open from Syria". The results page displays 282,980 items. The results are shown in a grid of cards:

- HDX**: Daily Summaries of Precipitation Indicators for Algeria. Includes a pie chart of UNHCR's populations of concern originating from Syria.
- LIDAR**: Humanitarian Data from Syria (2014).
- Duquesne-Heights_HLZ_hlz**: UNHCR's populations of concern originating from Syria.
- dust_rounded_2**: UNHCR's populations of concern originating from Syria.
- ETOPO 1m Hillshade w/ Elevation**: UNHCR's populations of concern originating from Syria.
- Economy**: UNHCR's populations of concern originating from Syria.
- L_3301s**: UNHCR's populations of concern originating from Syria.
- LOGOS LEAPS**: UNHCR's populations of concern originating from Syria.

Each card includes a "Save search" button, a relevance dropdown, and a filter button. At the bottom of the results page, there is a "UNCLASSIFIED" section.

Search Results

Screenshot of the Voyager search interface showing a detailed view of Lake Tuz.

The page title is "Voyager Search" and the URL is "voyagerdemo.com/ngaddccrawlers/navigo/#/show/0d158c17-ec69-4d36-9a1c-72395929930f?disp=default".

Lake Tuz

Landsat

<http://landsat.visibleearth.nasa.gov/view.php?id=76769>

Lake Tuz (Tuz Gölü meaning Salt Lake in Turkish) is located slightly west of central Turkey. It is one of the world's largest hypersaline lakes with a surface area of approximately 625 square miles. In this image Lake Tuz is white, bright cyan, blue and green. Vegetation is green and bare ground is tan. This Landsat 5 image was acquired June 2009. It is a true color image using TM bands 3, 2, and 1. Lake Tuz falls on Landsat WRS-2 Path 176 Row 33.

Details **Relationships**

Last Indexed	Sep 16, 2016 12:03:48 PM
Format	Landsat
Description	Lake Tuz (Tuz Gölü meaning Salt Lake in Turkish) is located slightly west of central Turkey. It is one of the world's largest hypersaline lakes with a surface area of approximately 625 square miles. In this image Lake Tuz is white, bright cyan, blue and green. Vegetation is green and bare ground is tan. This Landsat 5 image was acquired June 2009. It is a true color image using TM bands 3, 2, and 1. Lake Tuz falls on Landsat WRS-2 Path 176 Row 33.
Bbox	-115.29181 -35.86667 150.18822 54.21667
Title	Lake Tuz
Dataset Details Url	http://landsat.visibleearth.nasa.gov/view.php?id=76769
Path	http://landsat.visibleearth.nasa.gov/view.php?id=76769
Name	Lake Tuz

Details