PUTTING DOCUMENTS ON THE MAP USING SEMANTIC ANALYSIS

How the NSW Office of Environment and Heritage put hundreds of documents on the map, using semantic analysis to automatically create maps from PDF files

Delivering Coastal Data - The NSW Coastal Information System

Leander Wiseman
Senior Natural Resources Officer
NSW Office of Environment and Heritage
Leander.Wiseman@environment.com.au

Stephen Lead
Co-Founder and GIS Expert
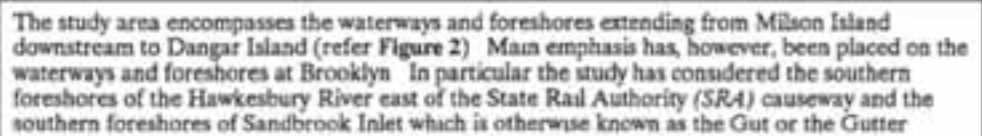Full Extent Pty Ltd
Stephen.Lead@FullExtent.com.au

The NSW Coastal Explorer – document discovery via a map interface

Background

Putting datasets on a map is relatively easy using Arca-GIS – provided there is an address, or a lat/long, or an existing GIS dataset with which to link the data. The problem arises when none of this information is available, potentially leading to a costly and time-consuming manual geolocation process.

This was the situation faced by Leander Wiseman, Senior Natural Resources Officer at the NSW Office of Environment and Heritage (OEH). OEH have a responsibility towards managing the coastline of NSW, and over the past 30 years have created thousands of reports including environmental plans, erosion reports, strategic management documents and feasibility studies relating to this work.

Many of these documents previously existed only in the form of bound paper reports which were difficult to access by OEH staff, and inaccessible to the public. OEH's goal was to make this information available to any interested party, in a

> The study area encompasses the waterways and foreshores extending from Milson Island downstream to Dangar Island (refer **Figure 2**) Main emphasis has, however, been placed on the waterways and foreshores at Brooklyn In particular the study has considered the southern foreshores of the Hawkesbury River east of the State Rail Authority *(SRA)* causeway and the southern foreshores of Sandbrook Inlet which is otherwise known as the Gut or the Gutter

Figure 1 - example text from a printed report

manner which was simple to use and which fostered information discovery. The ideal solution was to make the documents accessible via a map interface, so they could be found geographically and contextually by panning and zooming. The challenge was to find a way to geocode thousands of documents using minimal resources.

Introducing the NSW Coastal Explorer project

OEH commissioned Full Extent to build a prototype application to be called the **NSW Coastal Explorer**. The goal was to design a map-based interface to be used as the discovery mechanism for OEH's documents; specifically, those relating to the NSW coastline. The Coastal Explorer would also display traditional GIS datasets including the locations of beaches, estuaries and elevation datasets. An additional requirement was to display live feeds from the department's tidal- and rainfall-gauges.

The first step in the process was to scan the printed reports, which Leander organised via an external agency, before providing Full Extent with PDF versions of each report.

Since the reports were predominantly related to specific areas along the NSW coastline, they were strewn with references to placenames. Note that many of these reports pre-date desktop publishing, and were typed

Our goal was to determine the locations of each placename, allowing us to visualise any place mentioned in the reports as a pin on the map, and therefore to find and discover reports using a map interface.

Geocoding the reports using Text-to-map

Full Extent's **Text-to-map** product was used to read through the reports and find placenames, utilising a semantic search function. Rather than testing every word as a potential placename, the Text-to-map search recognises placenames by their context *within the sentence*. This allows it to very quickly parse a large document and extract the placenames. The semantic analysis approach also helps to avoid ambiguity - for example, evaluating Orange (the town or county) vs orange (the color).

The next step was to geocode the placenames to determine their coordinates, allowing them to be shown on the map. The output of this process was a KML layer containing markers for each placename reference, and a link back to its page in the report.

Building the mapping interface

Full Extent's **Mapping viewer** was used as the front-end for the Coastal Explorer, as it afforded very quick deployment with minimal coding required. The Full Extent Mapping viewer is based on Esri's ArcGIS Server JavaScript API, which meant that adding OEH's existing ArcGIS data layers was simple.

OEH specified that the site must be simple and intuitive, with no prior GIS experience assumed in the user. To this end, a deliberate decision was made not to use a Table Of Contents, as this can easily lead to confusion for novice users.
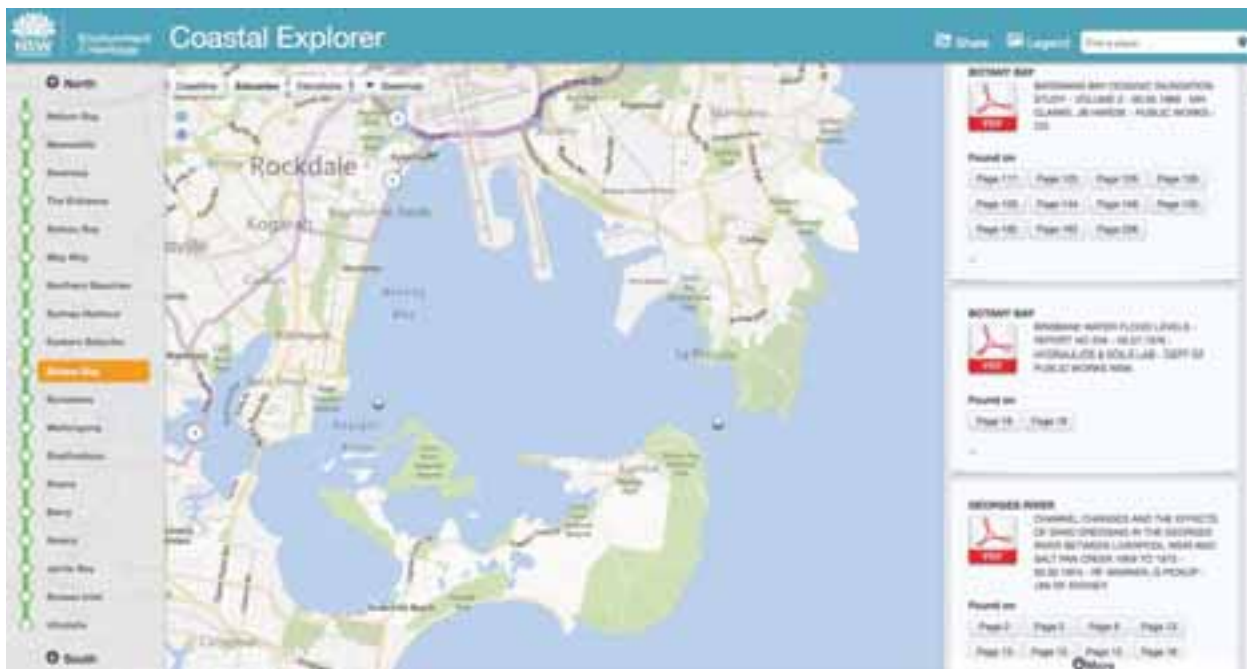

Figure 2 - the mapping interface is uncluttered and intuitive

Instead, the concept of "maps" has been used, whereby a **map** is configured to include a number of **layers**. For example, the Coastline map holds the beaches, erosion and tidal gauge layers, plus any reports relating to the

coastline. The Estuaries map holds the estuary monitoring stations and rainfall gauges, plus any reports relating to estuary locations.

Switching from Coastline to Estuaries is accomplished using simple toggle buttons, which remove and replace layers automatically while maintaining the current map scale. This allows a large number of layers and functionality to be abstracted behind the relatively simple buttons. The Full Extent Map viewer handles the switching mechanisms between maps.
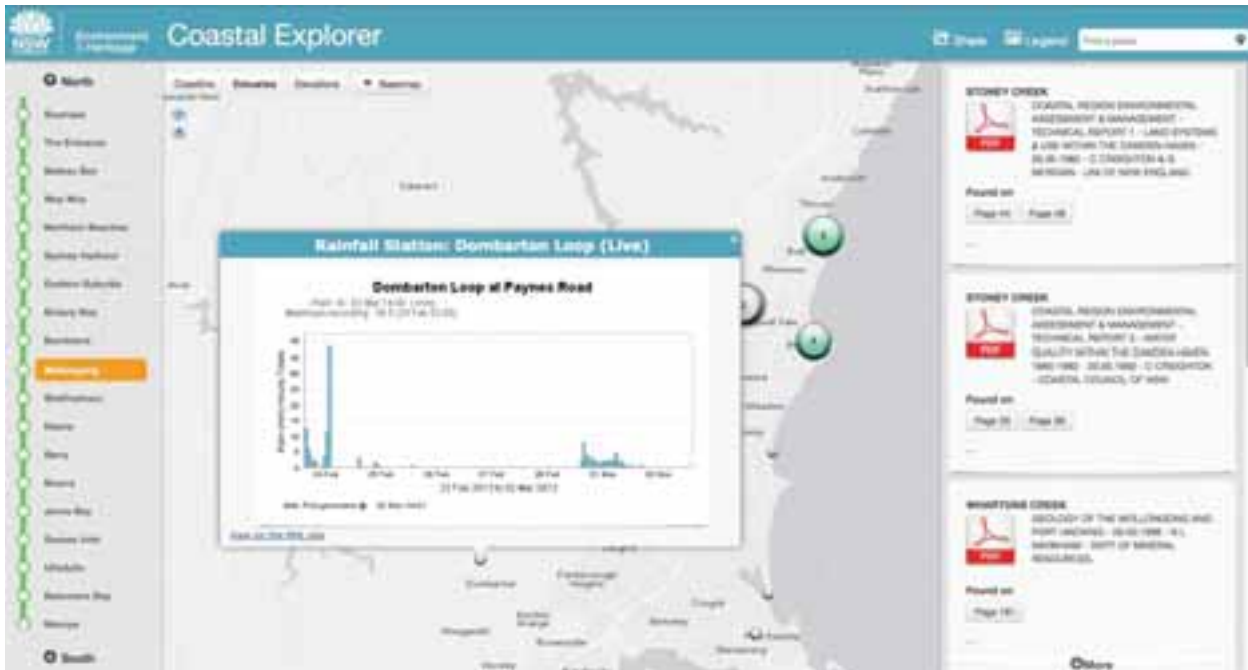

Figure 3 - the Estuaries map shows live tidal gauges

The NSW Coastal Explorer is designed to highlight datasets and reports located along the NSW coastline, so an innovative custom navigation widget was designed to help guide users to relevant locations, while still allowing freedom in where to navigate. The navigation widget is shown at the left of the map screenshots.

The navigation widget highlights the current map extent, and allows the user to navigate quickly to other key locations along the coastline by clicking directly on a location name. This allows the user to navigate from Merimbula (on the south coast of NSW) to Byron Bay (on the north coast) in seconds, without tedious searching or panning. The widget is updated if the user navigates the map window manually, so it always shows the user's current location.

Integrating the reports

Once the reports were processed by the Full Extent **Text-to-map** semantic engine, they were available as a KML layer, which was added to the map. The reports are shown in the panel on the right-hand side of the screenshots:

Report documents are added to the panel when they fall anywhere within the current map extent, and removed from the panel once the map moves away from that area.

Users may open a report directly on the page which mentions the placename (as shown in the "Found on" button in the screenshots) or download the reports for offline reading.

Benefits

The key benefit of the NSW Coastal Explorer is that vast quantities of valuable information, previously locked in these reports content and inaccessible ,even to OEH staff members, will be available to all of OEH and the general public. Moreover, these reports will be easily discoverable using the intuitive mapping interface, which helps users to discover spatial relationships between documents.

For example, a user unfamiliar with the Cronulla area may not think to look up nearby areas such as Garie Beach or Bundeena, since these areas are not related when using an alphabetical index. Using the map-based interface, it is easy to find any reports for the nearby areas.

Putting the documents on the map also allows the user to discover information about the area they are researching, by checking the tidal- and rainfall-gauges or viewing the estuary monitor station information.

The Full Extent Text-to-map semantic search allowed OEH to unlock value from their exsiting content and data by geocoding hundreds of reports with very little effort, avoiding countless hours of tedious manual data entry and making feasible a project which would otherwise have been too time-consuming to undertake.

Authors:

Stephen Lead
Full Extent Pty Ltd
Co-founder and GIS expert
Stephen.Lead@fullextent.com.au
www.FullExtent.com.au

Leander Wiseman
Office of Environment and Heritage
Senior Natural Resources Officer
Leander.Wiseman@environment.nsw.gov.au
http://www.environment.nsw.gov.au/estuaries

The NSW Coastal Explorer project is currently available as an internal prototype application only, but OEH hopes to launch it to the public shortly. In the meanwhile, please contact Stephen for further information.