

Increasing Access to Certain North Carolina Environmental Data -- North Carolina Policy Collaboratory Project Update

Submitted on behalf of the North Carolina Policy Collaboratory by
W. Christopher Lenhardt, Brian Blanton
Renaissance Computing Institute (RENCI)
The University of North Carolina at Chapel Hill



1 December, 2017

Increasing Access to Certain North Carolina Environmental Data -- NC Policy Collaboratory Project Update

Submitted on behalf of the NC Policy Collaboratory by
W. Christopher Lenhardt, Brian Blanton
Renaissance Computing Institute (RENCI), UNC-CH
1 December, 2017

Introduction

This interim report has been prepared as required by the North Carolina General Assembly (NCGA) pursuant to Session Law 2017-209 (Section 20.1) to provide an update on activities to-date related to the four main tasks assigned to the North Carolina Policy Collaboratory (Collaboratory) per Session Law 2017-57 (Section 13.7) as amended by Session Law 2017-209 (Section 20.1). The Collaboratory has designated the Renaissance Computing Institute (RENCI) at UNC Chapel Hill as the principal investigator on this project, during which time RENCI will focus on the following objectives:

- Identification and acquisition of digital data relevant to environmental monitoring and natural resource management, including, but not limited to, the digitization of analog records.
- The creation of online public access to National Pollutant Discharge Elimination System (NPDES) and other water quality permits, permit applications, and relevant supporting documents.
- Creation of a system for electronic filing of applications for such permits and relevant supporting documents.
- The Collaboratory shall assess the feasibility of transferring these data to a central, searchable, and publicly accessible digital database as well as how and where the database could be managed.

Approach

In order to develop the proposal for submission in March 2018, we have identified several critical elements for which a more detailed understanding is needed, and for which specific tasks will be derived. These issues are:

- *Data identification/description*: Characterize the volume and complexity of the data and associated issues such as proprietary restrictions, licensing, and/or privacy concerns.

- *Functionality*: Identify specific capabilities needed for a permitting system or for a data access system, e.g. fields required as part of a search interface (geospatial index, keywords).
- *Interactions*: Identify potential interactions between data and functionality that need to be captured.
- *Digitization*: Assess the magnitude of the potential task, applying a records retention policy to determine how far in the past to digitize documents, and what are the available options and costs for digitization services.

The starting point for assessing these elements is to establish a baseline understanding of the landscape of current existing online, digital, and analog data, as well as for the interfaces to the data. This includes a deeper understanding of the processes currently used and deployed by the Department of Environmental Quality for gathering and managing the relevant data and information. We also need to better understand the volume and necessity for digitization of existing analog records. Finally, we need to identify a small representative set of user scenarios (use cases) that capture how end-users anticipate interacting with the system. Understanding these various elements is necessary in order to develop recommendations, options, and feasibility of transferring the data to a publicly accessible digital database and where this database should be located maintained.

Activities

When developing a proposal of this type, the urge to jump to a technology solution at this stage should be avoided. The information gathered in this context will inform technology choices later in the proposal development process.

Data and Information

To date, staff from RENC1 and the Collaboratory have participated in two preliminary meetings with representatives from DEQ and DEQ Information Technology staff. These meetings provided the opportunity to discuss the current situation vis-a-vis the provision of water permit data and related environmental data. These discussions have provided a good initial look at the current broad range of DEQ data holdings and data delivery interfaces, as well as initial information related to the non-digital water permit-related data and information. DEQ staff also demonstrated several online interfaces to permit and environmental data. These conversations also provided some initial indications of permit processes, special data restrictions (i.e. proprietary data), and indications of the scope of the potential digitization task. Our discussion has helped to focus our early work.

To supplement and extend our understanding of the currently available content, data and interfaces, we conducted an online survey of DEQ and related North Carolina state government agency websites that provide access to permit and environmental data. We assessed not only what type of data is available, but also the relevant water permit context, and a description of the type of interface.

Based on the information gathered to date, there is a wealth of environmental data and water permit data available via North Carolina state websites. The current challenge may be less a matter of the availability of relevant data, but rather the organizing of the information in a manner that addresses the needs as determined under the user scenario descriptive work.

As part of the full proposal development, we will be conducting a more thorough survey of the types of environmental data currently available and to survey the types of interfaces to those data to verify our initial assessment.

User Scenarios

We are developing a basic set of use case scenarios. The use case scenarios will provide the functionality specifications for an online system. These user scenarios need not be exhaustive, but should be representative. Once validated, they become the basis for more detailed proposal development. In an actual development scenario, these user scenarios would drive system development, engineering, and operations. Initial scenarios include:

Scenario 1: *A water resource manager or a water utility employee would like to check changes in the status of existing or new NPDES permits in order to identify the potential for previously unknown contaminants to be present in the water.*

Scenario 2: *A new permit is filed, reviewed, approved, and submitted to EPA.*

Scenario 3: *As part of a site survey pursuant to developing a permit application, an engineering firm needs to map the hydrology and flooding potential for the site.*

Scenario 4: *A scientist in the UNC system needs to find in situ flood extent data to validate a new flood forecast model.*

Scenario 5: *An AP Human Geography teacher needs 'real' data on the geographic distribution of human population and relevant natural resources to demonstrate the concept of an 'ecological footprint' for an urban area.*

Scenario 6: *A sport fly fisher wishes to assess stream and atmospheric conditions, as well as recent meteorological events to decide whether to spend some time on the stream.*

Scenario 7: *A natural resource manager, ...*

Scenario 8: *Oyster fisher wants to assess upstream rain and flood events for potential downstream impacts on oyster reef.*

Digitization

The activities related to the analog to digital conversion activity have been preliminary and consist mostly of seeking to determine the scope of analog records that might be candidates for digitization and the condition / complexity of the documents. Some initial research has also been done regarding available digitization services.

Other Activities

We also provided a briefing to the NC Senate Select Committee on North Carolina River Water Quality during the 3 October 2017 committee meeting. In addition to describing our approach as outlined above, we presented an overview of challenges related to making digital data available and developing online information systems of the type outlined in the legislation.

Next Steps

We will continue meeting with DEQ to validate our landscape analysis, vet our user scenarios, and increase our understanding of water permit record retention and related requirements to assess the scope of the digitization backlog.

We intend to turn our initial list of resources into an example catalog. This will require value added work to develop structured metadata for each resource record. In this catalog, a resource can be a dataset, a data collection, information objects such as permit resources, or web-based interfaces and applications. This value-add work will also support search functionality for a potential pilot interface.

We will work with DEQ to validate or modify the catalog as appropriate.

We will be engaging with other industry experts who have developed online permitting systems to discuss their experiences.

We also intend to consult digitization and information retrieval experts at UNC SILS and the UNC library system.

We will conduct a limited survey of environmental data systems hosted by other states.

Rough timeline of activities going forward:

December 2017

- Finish landscape analysis
- Refine use cases
- Vet landscape analysis and use cases with DEQ

January 2018

- Develop project implementation scenarios
- Assess technology options
- Develop costing scenarios
- Pilot application mockup?

February 2018

- Revise/refine proposal
- Revise/refine pilot application

March 2018

- Finalize report and all associated products

April 2018

- Submit final report to North Carolina General Assembly no later than April 1, 2018