|  |
| --- |
| Use Case Name: Finding population data co-occuring in space-time region with selected events |
| Point of Contact Name: Patrice Seyed (UNM/RPI), Deborah McGuinness (RPI – dlm @ cs.rpi.edu)  Contributors: Jeff Horsborgh (UTAH), Margaret O’Brien (SBC/LTER), Mark Schildhauer (NCEAS) |

|  |
| --- |
| **Use Case Name**  *Give a short descriptive name for the use case to serve as a unique identifier. Consider goal-driven use case name.*  Finding population data co-occuring in space-time region with selected events |
| **Goal**  *The goal briefly describes what the user intends to achieve with this use case.*  The user (an environmental or ecological scientist using a web browser) intends to find data that will help her understand how various chemicals in bodies of water affect various populations of organisms. Therefore, firstly, the scientist intends to find available data that, constrained to a specific geographical region and a specific timeframe, provides measurements on a specific chemical and a specific organism population. More usefully, the scientist would like to find, given measurements of a specific chemical, what is the proximity to nearby organism populations, and how does population count trend with trends in the concentrations of the chemical.  Given the way that people describe data which is granular and how they do research, which is not. We need to be able to connect the data to the research for formally.   1. formalize the knowledge , and then in connection to the data to facilitate search.   Figure how what it is about the data that says you can do [something] (e.g., hurricane tracking) with it.  ---   1. Collect datasets (LTER, USGS/EPA, ORNL-DAAC available if appropriate) and any gaps in the data (Margaret, Patrice, Line) 2. Collect data on geospatial features relevant for use case, including watersheds, locations wildfires occurred (The data contain point information of collection locations) (Margaret, Patrice) 3. Identify Vocabulary Requirements *Immediately* Relevant for This Use Case    1. Quickly inspect existing community-based ontologies, identify for reuse, extension (Patrice, Mark, Jeff) 4. Develop ontology class definitions for detecting instances of fire or storm events (Mark, Hilmar, Patrice) 5. Convert datasets into RDF applying vocabulary identified in #2. 6. Demonstrate queries described in slide 4 and 5. |
| **Summary**  *Give a summary of the use case to capture the essence of the use case (no longer than a page). It provides a quick overview and includes the goal and principal actor.*  The use case is ultimately a facilitation of an environmental and ecological scientist understanding, over time, trends between chemical concentrations and organism population counts. |
| **Actors**  *List actors, people or things outside the system that either acts on the system (primary actors) or is acted on by the system (secondary actors). Primary actors are ones that invoke the use case and benefit from the result. Identify sensors, models, portals and relevant data resources. Identify the primary actor and briefly describe role.*  Primary: Environmental and Ecological Scientist |
| **Preconditions**  *Here we state any assumptions about the state of the system that must be met for the trigger (below) to initiate the use case. Any assumptions about other systems can also be stated here, for example, weather conditions. List all preconditions.* |
| **Triggers**  *Here we describe in detail the event or events that brings about the execution of this use case. Triggers can be external, temporal, or internal. They can be single events or when a set of conditions are met, List all triggers and relationships.* |
| **Basic Flow**  *Often referred to as the primary scenario or course of events. In the basic flow we describe the flow that would be followed if the use case where to follow its main plot from start to end. Error states or alternate states that might be highlighted are not included here. This gives any browser of the document a quick view of how the system will work. Here the flow can be documented as a list, a conversation or as a story.(as much as required)*  1)   User selects some region of interest  2)   User selects for a specific timeframe the data was collected from that region 3)   User selects for measurements on a specific chemical and its concentration (e.g., nitrogen, arsenic, and other nutrients or toxins, and coliform bacteria or other disease-threatening microbes) taken in streams (different measurement units depending on the chemical) 4)   Users selects a measurement dimension of interest 5)   User selects from a list of organisms  #5 is enabled by using region and timeframe selection of #1 and #2 to query population data with spatio-temporal proximity within some range, which requires lat/long for both measurement and population count data.  More formally, the user is making a query that: for region **r** during timeframe **t** for measurement of chemical constituent **c** using units **u**, what is the distance (**?d**) to the closest population of a organisms **o** and what is that organisms’ population count (**?c**) over **t**? |
| ***Alternate Flow***  *Here we give any alternate flows that might occur. May include flows that involve error conditions. Or flows that fall outside of the basic flow.*  3) User selects from a list of measurements on chemicals and their concentrations provided by the system based on proximity to region/time of selection #1 and #2.  This flow pre-calculates spatio-temporal proximity in advance, and user is able to select from a list of chemicals and their concentrations on record that are “closest”. Any of 1-5 can be refined in this manner. |
| **Post Conditions**  *Here we give any conditions that will be true of the state of the system after the use case has been completed.* |
| ***Activity Diagram***  *Here a diagram is given to show the flow of events that surrounds the use case. It might be that text is a more useful way of describing the use case. However often a picture speaks a 1000 words.* |
| **Notes**  *There is always some piece of information that is required that has no other place to go. This is the place for that information.*  Initial Data  For initial sample data, we are identifying three sources of water-related data.  All will cover Santa Barbara county and come from:  1.     Holdings from the Santa Barbara Coastal Ecosystem (SBC) provided by Margaret O’Brien. 2.     Holdings from CUASHI provided by Jeff Horsburgh 3.     Holdings used in SemantAqua a water quality portal developed by RPI with data from USGS and EPA.  Access provided by Deborah McGuinness  Some Background  \*Margaret O’Brien is the Information Manager at the Santa Barbara Coastal LTER site (aka SBC), one of 26 LTER (Long-term ecological research) sites in North America.  She has a number of domain researchers (aquatic ecologists, hydrologists, oceanographers, etc.) contributing data to a repository, that is based on \*potentially\* comprehensive, and valid metadata in EML  (EML is a formal XML schema for ecological and more generally, scientific metadata).  EML metadata are used broadly throughout the LTER and at NCEAS, but are different from the metadata used by the CUAHSI framework.  \*\*CUAHSI is a consortium of hydrological researchers that contribute their data to the confederated HIS database system.  Jeff Horsburgh is deeply involved with HIS and CUASHI.  CUAHSI also contains a lot of hydrological information that might be complementary to the data that Margaret's researchers are collecting, for example, in streams within Santa Barbara County.  Similarly, the EPA or other agencies might \*also\* be collecting data on water quality in streams in Santa Barbara County.  \*\*\*Towards the goal of environmental monitoring, this project uses identified pollution events relative to user chosen regulations.<http://tw.rpi.edu/web/project/SemantAQUA>  We will inspect different dimensions on measurement across the projects, as well as protocols applied, and consider how best to use an ontology across these dimensions.  Example question(s) |
|  |

***Resources***

*In order to support the capabilities described in this Use Case, a set of resources must be available and/or configured. These resources include data and services, and the systems that offer them. This section will call out examples of these resources.*

**Data:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data | Type | Characteristics | Description | Owner | Source System |
| (dataset name) | Remote,  In situ,  Etc. | e.g. – no cloud cover | Short description of the dataset, possibly including rationale of the usage characteristics | USGS, ESA, etc. | Name of the system which supports discovery and access |
|  |  |  |  |  |  |

Dataset:

Paper DOI  
<http://dx.doi.org/10.1002/hyp.9508>  
  
Figure 1 shows these stations (only the codes here, names are in paper, and in data):  
Precipitation:  
GV201  
GV202  
HO201  
HO202  
  
stream water chemistry and Stream flow  
GV01  
GV03 (chemistry only)  
ON02  
HO00  
  
  
precip data  
GV201  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.4005>  
GV202  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.4006>  
HO201  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.4007>  
HO202  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.4008>  
  
Stream flow  
GV01  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.3007>  
ON02  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.3010>  
HO00  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.3008>  
  
  
water chemistry  
<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.6>

Hi Margaret,  
  
We're trying to examine the data which presumably informed Figure 2 in the Coombs/Melack paper--  
  
Figure 2 depicts the difference between instantaneous measures of stream runoff in burned vs. unburned sub-catchments along Arroyo Hondo Creek.  
  
We're trying to figure out which data set(s) might have informed this analysis.  The one Arroyo Hondo stream discharge data set that you sent us a link for:  
  
HO00 <http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.3008>  
  
Seems to only include measurements from one station, as there is no "station ID" column in the data (e.g. for HO21 and HO31 stations)? Seems like these are all just for HO00...  
  
I searched the SBC data archives and couldn't find any additional data about Arroyo Hondo stream discharge data from other stations.  
  
BTW-- I'll proceed to the next graphs and see if I can figure out which data informed them...I think you have found the data for Figure 3... but semantic descriptions of the data that informed Figure 2 would be useful.  
  
Thanks!  
  
Mark

**Modeling Services**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Owner | Description | Consumes | Frequency | Source System |
| (model name) | Organization that offers the model | Short description of the model | List of data consumed | How often the model runs | Name of the system which offers access to the model |
|  |  |  |  |  |  |

**Event Notification Services**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Event | Owner | Description | Subscription | Source System |
| (Event name) | Organization that offers the event | Short description of the event | List of subscriptions (and owners) | Name of the system which offers this event |
|  |  |  |  |  |

**Application Services**

|  |  |  |  |
| --- | --- | --- | --- |
| Application | Owner | Description | Source System |
| (Application name) | Organization that offers the Application | Short description of the application portal | Name of the system which offers access to this resource |
|  |  |  |  |

**Other resources**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Resource | Owner | Description | Availability | Source System |
| (sensor name) | Organization that owns/ manages resource | Short description of the resource | How often the resource is available | Name of system which provides resource |
|  |  |  |  |  |