|  |
| --- |
| Use Case Name: Finding hydrology 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 Horsburgh (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.*  Find hydrology data co-occuring in space-time region with selected events |
| demonstrateshow ontology can e a) discovery of hydrology data related in time and space to event, and b) apply a threshold to other similar hydrological data to potentially expose a similar environmental event. The environmental event in this example is a wildfire. |
| **Goal**  *The goal briefly describes what the user intends to achieve with this use case.*  The user is an environmental or ecological scientist who wants to examine the patterns of chemical constituents of streams in response to wildfires. S/he intends to find (using a web browser) hydrologic and chemical data where wildfires have occurred., so will constrain searches to a specific geographical region and timeframe, and look for measurements of specific chemicals.  The use case will be extended to illustrate that ontology can be further used to potentially detect a signature of an event (like a wildfire) in other similar data. Patterns in this data might be suggestive of discrete events in other data.  This use case is based on an actual published study (see Resources), which examined flux of nutrients in watersheds affected by a wildfire in the Santa Barbara coastal area (see definitions under Notes).  We will annotate the data with info that scientists stated about the data, but that is not actually part of the data’s description. For example, in the paper (see resources), scientists say that a feature is in the data (high instantaneous ratio of runoff:precip) and is indicative of fire. |
|  |
| **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 or 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, based on occurrence of fire  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)   #3 is enabled by using region and timeframe selection of #1 and #2 to query chemistry data with spatio-temporal proximity within some range, which requires lat/long for both measurement and fire event.  More formally, the user is making a query that: for region **r** during timeframe **t** for measurement of chemical constituent **c** using units **u.** |
| ***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.*  Definitions  Flux is transport of some chemical, which can be measured by the change in concentration of the chemical. This is measured as substance per area per time. A space, time, and substance dimension. E.g., temperature flux, light flux.) Wikipedia: “The rate of flux of a property over time”. In this case water in the carrier (liters/sec). Export of suspended sediment mass flux since you are weighing the amount of sediment.  “3010 ha” is hectare (10,000 meters squared) hectare is a standard measurement of a spatial extent.  SWEET:runoff: flow of water …..across different media , over land surfaces, etc.  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.   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)  Questions at Level of Dataset Metadata  ------------------------------------------------  1) What data is available on the long-term collection of rainfall on the south-facing slope of x?  2) What data is available on streams in (a) Santa Barbara County, (b) San Jose Creek, (c)  San Jose Creek within tidal flux region or ocean terminus (“mouth of stream”)?   * Note, a list of options can be made available to the user through partonomic reasoning. Also, there is a reasoning challenge with terms like “mouth” which are vague and somewhat operational.   3) What other data or data sources are out there which might help comprehend Santa Barbara County data, via DataONE member nodes?  Questions at Level Requiring Dataset Metadata and also Data  ---------------------------------------------------------------------------  4) For location a and timespan t, what is the (a) average stream discharge (or flow stage height), in units y (e.g., volume per time) and (b) the flux for chemical constituent c?  5) What are the inputs needed to calculate flux of nitrate over a period of time t from mouth of stream?  6) Given similar measurement data across several collection agencies, can we determine if the protocols are similar enough such that the measurements can be accurately compared?  Example Answers  -----------------------  1,2) dataset a,b,c  3) project a, dataset b  4a) 33,000 liters/sec  4b) 15,000 micromoles /sec/meters squared  5) Nitrate concentration in stream water, stream water delivery rate, watershed area of stream.  6) Yes.  Justification for Answers  ------------------------------  1) Matching of metadata for rainfall data and for a specific location  2) Based on geographic location and waypoint encoded for the datasets  3) Matching of metadata on types of studies and location for other projects  4) Based on retrieval or calculation (Available as collected or derived data?)  5) Based on provided definition of how flux is calculated.  6) Based on how the protocols are defined.  How Semantics was Used  --------------------------------  1) Not needed, just matching on available metadata.  2) Using partonomic information available, perhaps through a service like GeoSPARQL  3) Not needed, matching on available metadata.  4a,b) Depends on “Justification for Answer”  5) Facilitated by a definition of flux over time with respect to inputs for calculation  6) Inference is used to evaluate similarity in protocols. |
|  |

***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 |
|  |  |  |  |  |

**Notes from working group – 2013 April 15-18**

Can we inter-compare operators to detect change (nitrate concentration from x to y) for example for a rainfall event.

Detect events:

When you see a sufficient contrast it may be indicative of a burn event.

What are signatures of burn event? E.g., change in soil type, change in suspended sediment.

Suspended sediment may be caused by a rainfall or a burn event.

Suspended sediment coupled with changes in water chemistry is indicative of a burn event.

(nitrogen or other nutrients were measured in this paper.)

fecal coliform levels increase after rainfall due to runoff.

(What else makes people sick besides fecal coliform?)

can we correlate levels of chemicals and rainfall.

In southern Carolina water quality is dependent of rainfall.

The difference in runoff was less with more storms spaced closely together

There exists a shape file for the burned area, maybe from the fire departmet.

Instantaneous runoff is the hourly rate of runoff measured over part of a day.

Can we get San Pedro USGS measurement data.

Describe indicators as what might have happened.

Instantaneous run-off could be an indicator of burned event.

We can infer rainfall events.

Could use the Melack data as a filter to find indications of burn data.

How do we calculate instantaneous runoff?

Watershed area (table 1) you can calculate instantaneous runoff from

Discharge in liters per second, divide by area of the watershed. Centimeters per time.

Is the instantaneous run-off the maximum instantaneous runoff?

The query for fires occurring is particularly interesting

We can use the varied measured variables to see what is the result that wre the consequences of a fire or rain effects. We know the coincidence of fire, rain, and data collected together.

**1)data is useful for predictive of certain events occurring.**

**Come up with ways to suggest the hypothesis that fire occurred. (attempt in OWL)**

**2)tagging/annotating so users will know there is measured hydrology data used in wildlife and rainfall events.**

**Queries to Support:**

1. “Do we have evidence to support that fire occurred at location x?”
2. “Give me water quality measurements from streams where there is some indication that fire occurred with that watershed ” (watershed drains into a stream)
3. “Where was there a spike in instantaneous runoff?” (instantaneous runoff is caused by storms, the consequence of a wildfire followed by a storm leads to instantaneous runoff, other causes of instanteous runoff include harvesting soil) (requires some statistical analysis that may not be in the original data tables.)
4. “What is the impact of fire of water quality?” (what are the indicators, domain knowledge, but also what does the Melack paper conclude.)
5. “I know of a burn event at a form at bounding box x, how did that affect water quality and is the effect similar to studies?”

\*\*\*suspended sediment is bad for salmon.

Fire: nitrogen, instantaneous runoff -> indicative that fire occurred.

“Peak

instantaneous discharge (m s**\_**1)was determined by dividing

a storm**’**s maximum 5-min discharge rate (m3 s**\_**1) by the

watershed**’**s contributing area (m2).”

Based on wildfire/hydrology paper, enable the sort of searches that are only possible from aggregating the data and making the conclusions that the researchers made.

Enable queries that map to the abstractions that are added by the authors that are not available directly in the data.

A flood event exists based on a "significant" change in water flow.  
Can we describe this criteria formally to "derive" there are instances  
of the 'Flood Event' class?

Different people have different conceptions of what an "Event" is. (Do we care for this work? Only to the extent that it supports query answering for this extension.) If we have a formal notion of a flood event we can query for : "Stream Chemistry data before and after flood event" Also formalize fire event to query: "Stream chemistry data before and after fire event". Exercise of going through paper and mapping important concepts for our  
model to community ontologies (envo, sweet, obo-e, etc.)

Monitoring streams – flow, chemistry, and what else. What motivated those efforts?

Consolidate focus on what stream ecologists have actually reported on. Wildfire evoking changes that affect kelp ecosystems. (Streams don’t catch fire, watersheds catch fire, streams transport sediment from watershed to the nearshore areas.)

What is a stream ecologist interested in analysizing? Look at the paper and backtrack from the study and understand the underlying semantics.

There are rich terms in the study paper that are not in the data.

Given the way that people describe data ( granular) and how they do research, which is general and integrated. 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 hurricane tracking on it.

---

ACTION ITEMS

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.

MOB’s notes

Annotate the data with info that a scientist found out about the data, but that is not actually part of the data’s description. So with the fire-hydro use case, scientists say this feature is in the data (high instantenous runoff:precip) and is indicative of fire. Patterns in the data can be suggestive of discrete events. So this is entirely appropriate for a separate annotation doc – ie, does NOT belong in metadata.

Watershed is usually ‘big’, or the larger unit = land area that drains to point. ‘Catchment’ (or ‘drainage’) is more often applied to the smaller unit, like sub-catchment. Drains to a stream-reach

Right now, it would be hard to find this data in dataone.

Lines of questioning:

I am interested in stream metabolism. What data is there in dataone that are useful to me?

I know that I need certain measurements (dissolved oxygen, ph, conductance, solar irradiance), and they have to be co-located in space and time.

Give me the datasets to help determine whether or not there was a fire.

1. Simplest answer: here is a shapefile that describes a fire’s perimeter
2. Next: here is a time series of shapefiles at a place that show changes in land cover, and where there is some ash.
3. Most complex: here is data that has measurements of watershed features that are often affected by a fire, and that might indicate a fire has happened.