20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

Scientific workflows in a geographic portal for web-based spatial analysis

Touraivane[a], **A. Couturier**[a], **G. Wattelez**[a] **and M. Allenbach**[a]

[a] *University of New Caledonia*
Email: arnaud.couturier@univ-nc.nc

Abstract: Our geographic portal, created in 2009, gives access to information collected during many years of scientific research. The information comes from the University of New Caledonia and its partners and includes field data, results of studies, analyses and scientific models on environmental, social and cultural issues. All the data concern New Caledonia, Vanuatu, Wallis and Futuna in the South Pacific.

The portal is based on a geographic server that implements OGC standards (Open Geospatial Consortium) on geographical information, including WMS (Web Map Service) and WFS (Web Feature Service), allowing information to be displayed on dynamic maps.

The purpose of this paper is to present the tools our portal provides to process specific scientific workflows on different topics. Unlike the static information that was previously available, the results of these workflows are not necessarily static and stored. Users can use the application with their own parameters, and can get the results of calculations in real time.

Two scientific projects are presented as examples of possible applications. The first is the assessment of the risk of fire start in New Caledonia. The increasing number of fires is causing substantial damage on the main land. A multidisciplinary team of scientists combined their knowledge to create a forecast model to estimate the risk of the start and spread of fires. The second project is about monitoring dredging for the creation of Vavouto harbor in New Caledonia.

The use of this tool in the two projects shows that incorporating workflows in a geographic portal is an efficient way to highlight and share their results. Incorporating workflows in such a framework makes it easier to interpret results, possibly by superimposing them on results of other workflows also calculated in real time or by using stored static data. This study underlines the importance of interoperability, as promoted by OGC.

Keywords: *Interoperability, OGC (Open Geospatial Consortium), dynamic maps*

1. INTRODUCTION

In 2009, the University of New Caledonia (UNC) created a geographic portal to compile, display, and share data on maps. The data were collected by the university itself and by its partners in projects on different topics including the environment, geology and oceanography, mainly in the South Pacific territories, i.e. New Caledonia, Vanuatu and Wallis and Futuna. Here we briefly present the means used for the construction of this portal.

In addition to the simple visualisation of geographical data, users can request geospatial and customizable scientific workflows. We also provide an overview of the incorporation of the scientific workflows in the portal and explain how they function.

2. THE GEOGRAPHIC PORTAL

2.1. Overview

The geographic portal was created in 2009 to store and display field data or results of scientific research on a map. Users can access a large amount of easily viewable information. Combining and comparing data from very different scientific fields is straightforward, so users can focus on their own topic.

2.2. Implementation

To make our new tool interoperable, we chose Open Geospatial Consortium (OGC) specifications to store and share spatial data. We chose the WMS (Web Map Service) standard to send a request to a map server to provide geo-referenced images. The WFS (Web Feature Service) standard is designed for the management of geographical features described in vector form. To comply with these standards, we used GeoServer software (http://geoserver.org/display/GEOS/Welcome).

2.3. Advantages of interoperability

The use of OGC standards enables an application to be interoperable with other applications that also comply with OGC standards. Interoperability is the ability to work with other existing or future systems provided there are no restrictions to access or implementation. Provided the data also comply with the specifications, this feature makes it possible to use data from different fields and different sources.

Merging information is made simple, thus allowing visualization and interpretation of the phenomena by experts. In addition, the map in the geographic portal is dynamic so the expert can obtain a general view or a partial view of the study area. When an abnormal phenomenon is identified, the map allows the viewer to understand and explain it. The explanation for the anomaly may depend on the incorporation of other data.

2.4. Introduction of scientific workflows in the geographic portal

Storing the outputs of scientific models is not always necessary, especially when the phenomenon to be modelled changes often (at least daily). In the case of a daily phenomenon, the user does not always need to identify its behaviour each day, but only on the days that are important to him/her. When outputs are stored automatically, even if the results are never requested for certain days, the model will run and its outputs will be stored every day. On the other hand, the use of results is ensured if they are provided to a user of an application, and in this case, computing and storing the results is worthwhile.

Workflows may also require a large number of parameters, which provide more results or possible visualizations. In this case, storing the results is not possible since the model would have to run with all possible combinations of the parameters.

3. STRUCTURE OF WORKFLOW SERVICES

3.1. A server dedicated to scientific workflows

The scientific processing software is hosted on a dedicated server, from which the required data is accessible. During each user request, the server receives and processes the request using a PHP (Hypertext Preprocessor) script that takes the user's parameters into account and starts the requested workflow. Once processing is complete, the PHP script retrieves the information that has been computed.

3.2. Processing software

In our examples, the processing software we chose was R (R Development Core Team, 2011) but the operating principle would be the same if other software had been used.

When the processing script is run, it fetches the data needed to run the model, it implements the model, and saves the results in temporary files, and additional information in supplementary files. These are the files the PHP script reads to send the information to the user's web browser. For spatially defined results, the output files can be pictures or text files describing geographic features in GML (Geography Markup Language).

3.3. How results are displayed on dynamic maps

The display of model results functions in the same way as the display of stored data, except that the data is not stored in a database, but in temporary files that are created while the process is underway. Addresses of files containing the results are transmitted to the user's web page. Thanks to the AJAX (Asynchronous JavaScript and XML) concept, the file content is sent to OpenLayers, which is responsible for interpreting and displaying geographic features.

3.4. Interactions between components for a workflow

Figure 1 shows how the components interact, so that scientific workflows can run according to a user request and their outputs can be displayed in real time on the map in the geographic portal.

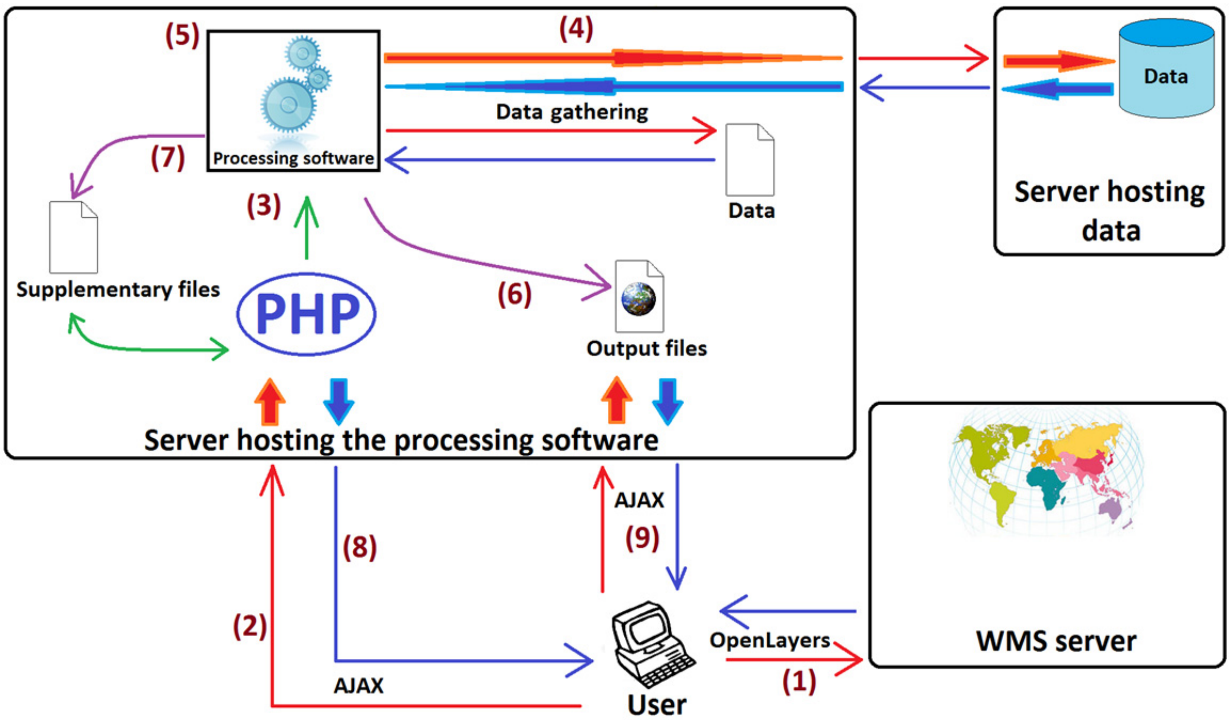

Figure 1. Interactions between components for displaying scientific workflows outputs in real time

After access to the geographic portal page, a WMS request is sent to the WMS server to obtain geo-referenced images of the study area. The answer leads to the display of the base map on the dynamic map (1). For each workflow, the user sends a request containing the parameters required for processing to the server dedicated to scientific calculations (2). PHP receives the request and starts the process software (3) specifying any parameters entered by the user. The process software collects the required data (4) on a remote server, or on the processing server itself. Then calculations begin (5), this may be a model whose results are stored in files, some of which describe the results (6) while others provide additional information (7). The addresses of the files are sent to the user's web browser (8). The content of the file is retrieved from the web page by the AJAX concept (9) and forwarded to OpenLayers, which is responsible for the presentation of the outputs.

The geographic portal is available on http://grimm.univ-nc.nc/geoportail/. Currently, we offer workflows on two different topics, which we now describe as examples of what is possible.

4. TWO IMPLEMENTATIONS

4.1. Fires in New Caledonia

Study context

The number, frequency and extent of fires are on the increase in New Caledonia. Following the huge fire that destroyed the *Montagne des Sources* in 2005, scientists from different fields joined forces to conduct a four-year study to better understand the origin and impact of the fires. In the framework of the "Fire and ecosystem biodiversity in New Caledonia" (INC) project, large amounts of data were collected and analysed, including satellite images containing climate data, and models were built of the risk of fires starting and spreading.

Interactions between ecosystems, human practices, climate and fires were analysed and a geographic information system (GIS) was developed to simultaneously monitor these components in space and time (http://www.espace.ird.fr/index.php/projets-71/49-projets/projets-en-cours/77-incendie-et-biodiversite-des-ecosystemes-de-nouvelles-caledonie). The savannah was identified as an environment conducive to fire starting and spreading, and relief was shown to facilitate the spread of fires. An anthropological study helped to better understand the use of fire by local populations, especially for agricultural purposes. A new seasonal drought index, based on *El Niño* events was implemented to assess the areas affected by forest fires across the New Caledonian territory. From these observations and other statistical tools, a risk model for the start and spread of fires was built for *Grande Terre* and *Île des Pins*. A biodiversity risk model and an erosion risk model were also built.

A model to assess fire risks

The model to assess fire risks provides a map showing risks levels represented with a colour pallet. All the input layers of the model, including remote sensing data (Normalized Difference Vegetation Index, burned areas) and climate data (6 indices including humidity index and temperature index on 17 zones), were standardised in a grid raster with a resolution of 300m. The risk of fire starting is assessed using a Bayesian network including the human, environmental and climate components; each of these components is itself a sub-model incorporated into the main model according to the Bayesian networks theory. A conditional probability law relative to the state of the variables of each component can estimate the probability of a fire starting for the component concerned. A total of 87% of model predictions based on daily data from a 10-year period (2000-2010) were correct. The impacts of fires on biodiversity and erosion are estimated by integrating respectively induced hazard and estimated soil loss in the assessment of the risk of a fire starting (Mangeas et al., 2013).

Process on the Web

Using the model described in Section 0 above, the geographic portal contains an application that calculates the risk of fires starting on *Grande Terre* and *Île des Pins* for each day, from the first day included in the study (January 1st, 2000) to today. For this purpose, a partnership was set up with *Météo France* New Caledonia for the automatic recovery of climate data, so that the weather indices for the territory required to estimate the risk of a fire starting, could be updated daily.

In the application, the user can choose the day and one of the three outputs provided, i.e. the risk of fire starting, the risk for biodiversity, or the risk of erosion, as well as the form in which the results are displayed (as raster images or in vector form with contour lines or polygons). Results are displayed directly on the map in the geographic portal. Different levels of colour represent different degrees of risk. Other information can be added to the map depending on user needs. **Figure 2** is a screenshot showing an example of fire risk map.

4.2. Environmental Management Plan for monitoring dredging in Vavouto Bay (New Caledonia)

Study context

The production of nickel is the main source of wealth in New Caledonia, which has around 30% of reserves of nickel ore in the world. The project to mine the nickel deposit in the Koniambo massif (*Province Nord*) carries high hopes for balancing economic and social development in New Caledonia. To make this project possible, a harbor had to be constructed at Vavouto. A deep navigation channel had to be dug in the lagoon for the transport of construction material for the nickel processing plant and subsequently for the export of finished products (Touraïvane et al., 2011).

However, these works are taking place near areas classified World Heritage of Humanity by UNESCO, due to the exceptional biodiversity found there (Allenbach et al., 2010). That is why in 2005, INERIS (*Intitut National de l'Environnement Industriel et des Risques* - http://www.ineris.fr/) called for an environmental management plan to monitor the impacts of the construction works on this fragile environment.

One of the parameters used for monitoring the impacts of dredging is water quality as measured by turbidity. Measurements were taken by fixed buoys at 26 sites in the vicinity of the Vavouto Bay. From these measurements, a database was built using a relational schema that contains all the information concerning water quality measurements and other parameters such as depth, water temperature and weather conditions. With all these measurements, it is possible to estimate turbidity at each point of the study area, including Vavouto Bay. The chosen interpolation method was kriging.

Interpolation

Kriging is the optimal method of assessment in the statistical sense; it is based on the analysis of correlations between different spatially distributed points (Gratton, 2002). An additional advantage of kriging is the possibility to calculate the estimation error, which gives the degree of confidence of the assessments.

Kriging the area was made with measurements taken on the stations. A minimum of six different stations was required to perform the interpolation on a given day. If this minimum was not met, data from nearest days (past and future) were added until at least six different stations. We used ordinary kriging because it has been found to be a good concession between complexity and accuracy. The exponential type was selected for the variogram γ, i.e. $\gamma(h) = \sigma^2 \left[1 - e^{-\frac{h}{a}}\right]$ where σ^2 is the reached level (maximal variance without nugget effect) and a is the distance for which the level is reached. Variogram parameters are automatically selected in order to minimize the error between the theory variogram (exponential type) and the experimental variogram (obtained from measurements).

Interpolation of our study area produced a map of selected variables, particularly water quality. Using this map, the expert can get a general view of the monitored area and determine if the risk of impacts is high on certain days.

Web-based tools for the environmental management plan for Vavouto harbor

One of our applications is gathering information collected at the sampling sites in the database to enable estimations by kriging and to display outputs on the map in the geographic portal. The user selects the day and the variable to be displayed (turbidity, wind speed, etc.) and runs the calculations. The results are then automatically displayed on the map. **Figure 3** is a screenshot showing results of turbidity interpolation on January 23rd, 2009.

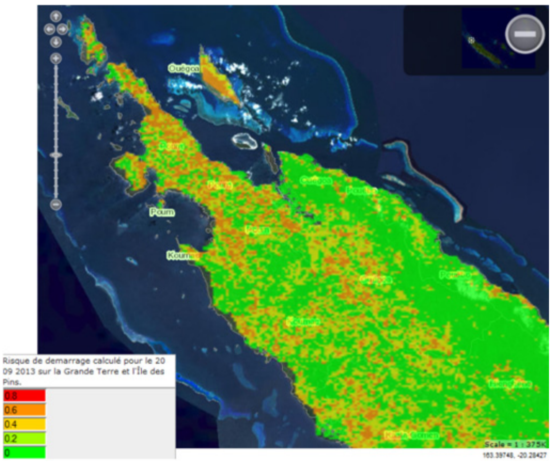

Figure 2. Map of the fire risk assessment on September 20th, 2013 displayed in the geographic portal

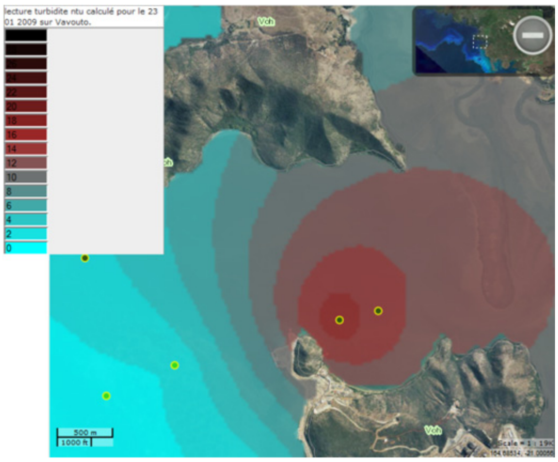

Figure 3. Map displayed in the geographic portal about turbidity in the vicinity of Vavouto harbor on January 23rd, 2009

One weakness of this application is that it is not possible to obtain a temporal overview of the situation without choosing several days and running the calculations for each of the days. But an expert is primarily interested in anomalies, such as peak turbidity, which only rarely occur. Consequently, we provide an easy access to the database to deliver a graph. The user selects the site of interest, specifies the distance around this site that includes the nearest sampling sites, and the variables to be taken into account. Results are displayed as static graphs and/or dynamic graphs made using the Google Charts API (https://developers.google.com/chart/), making it easier for the expert to establish which days turbidity thresholds are exceeded. After analysing the graphs, the expert can choose exactly which days to examine to ensure that works proceed in accordance with the preservation of the area's flora and fauna.

Unlike the database for the assessment of fire risk in New Caledonia, this database is not regularly updated. However, updating it is technically possible and would be a desirable feature of future environmental management plans. A daily update of a database would enable better monitoring of the area and more prompt and appropriate actions depending on the observations.

5. DISCUSSION

5.1. Advantages

The workflows described above are two examples among others. The results of each of these workflows are displayed on the same map. Merging information about each of these workflows is easy for the user because from his/her point of view, it is simply a question of stacking layers one on top of the other to make them visible (or not). If the workflow has been applied over a period of several days, the user can also see how the phenomenon evolves.

Other data can be added to the map. Indeed, the use of OGC standards ensures interoperability, and data from other sources can be used to complete the map. These additional data may help explain the phenomenon or certain anomalies. This is one way to showcase the research conducted by our different partners along with their results. Accessibility and usability are additional advantages since it is not necessary to code or use specific GIS software to access information or statistical workflows.

The use of dynamic maps allows users to obtain the information they need at the desired scale. The user can obtain a general view of an area or focus on a sub-area. This is particularly useful when there is need to target highly localized anomalies. The expert can then interpret them at the appropriate scale.

5.2. Limits of the method

One disadvantage is the time required for processing. In a web application it is unusual to have to wait several minutes for an output. But the volume of data that has to be collected and processed to obtain a result cannot be reduced. It is therefore important to inform the user that the requested workflow will require waiting. This is why workflow outputs are sometimes temporarily stored or put in cache, especially those concerning the risk of fire ignition.

5.3. WPS in perspective

Up to now, the workflows available in the geographic portal can be considered as "thematic", that is to say they are related to a specific problem in a specific context (estimating fire ignition in New Caledonia, interpolation by kriging of physiochemical and meteorological data in Vavouto Bay). The parameters sent by the user are limited (mainly the date and the name of a variable). But it might be useful to provide more "generic" workflows, i.e. workflows that are not necessarily related to the supplier's particular problem but that can be applied to data provided by the user; in other words, workflows with no specific context.

Such generic workflows could apply to different fields, and be of varying complexity. If the purpose is to process geographic features, for instance with intersections or differences, workflows would be simple and very general. In that case, the processing time would only be a few milliseconds. We could also offer kriging as a generic workflow. The user would provide a set of points with a value associated with each point, the boundary of the zone to which the method is to be applied, and other parameters related to the desired type of kriging (simple, ordinary, or universal kriging). The output could take the form of the workflow outputs described in Section 0 above.

Other workflows, also generic but related to a specific domain are also possible. For example, in pedology, it would be possible to compute the "C-factor", i.e. the soil cover that limits the impact of raindrops. The C-factor is mainly determined by land use that provides information on the density and height of the canopy, the density of the lower strata, and of the humus layer. One possible application would be the following: the user uploads a land use map with which the processing server calculates the C-factor.

OGC provides specifications for workflows grouped under the WPS (Web Processing Service) standard. In the same way as for WMS, WFS and WCS standards, WPS was primarily established for geo-spatial processing. It specifies how to query a server about its capabilities (GetCapabilities), a particular workflow (DescribeProcess) and how to request a workflow execution (Execute). This standard was developed to enable interoperability between workflows and of course, between the inputs and outputs of workflows. Using WPS facilitates the processing of the data provided by a user, because the input data, i.e. the description of geographic features, must comply with the OGC standards. Presented workflows are not yet available in an OGC WPS but it is a work in perspective.

6. CONCLUSION

In this paper, we described how thematic workflows were incorporated into the UNC geographic portal, which already provided access to static data. Compliance with OGC standards such as WMS and WFS and the use of standard descriptions of geographic features make interoperability possible, meaning data from different partners can be made available and displayed simultaneously; and scientific workflows can be incorporated easily.

This geographic portal provided access to field data and to the projects results. It now also enables scientific workflows to be performed in real time, allowing the user to monitor a phenomenon day by day. Research works can be highlighted and their contents made easily accessible to non-experts; in other words, viewers do not have to learn a programming language or to have special software to access information and workflows.

The service could be improved by proposing not only thematic workflows directly related to a geographical context and lead studies, but also more generic workflows of varying degrees of complexity. Some workflows could even be associated with topics such as geology and ecology. The OGC WPS standard was created to explain how to provide such workflows. The user would then be able to perform the desired workflow of their own data (geographic features) and view the results on the dynamic map.

REFERENCES

Allenbach, M., M. Mangeas, and Touraïvane (2010). Le suivi environnemental du dragage de Vavouto (Nouvelle-Calédonie). Un chantier de dimension mondiale à la porte d'un site classé au patrimoine mondial. *XIèmes Journées Nationales Génie Côtier - Génie Civil.* Les Sables d'Olonne: Paralia CFL, 2010. 573 - 582

Gratton, Y. (2002). Le krigeage : La méthode optimale d'interpolation spatiale. *Les articles de l'Institut d'Analyse Géographique*, 2002.

Mangeas, M., C. Gomez, T. Curt, T. Ibanez, J. Munzinger, P. Dumas, M. Despinoy, C. Hély (2013) Wildfire risk for main vegetation units in a biodiversity hotspot: modelling approach in New Caledonia, South Pacific. *(in submission)*.

R Development Core Team (2011). R: A language and environment for. Vienne: R Foundation for Statistical Computing, Vienna, Austria.

Touraïvane, M. Allenbach, M. Mangeas, and C. Bonte (2011). Monitoring the Turbidity Associated with the Dredging in Vavouto Bay in New Caledonia. Paper presented at 19th International Congress on Modelling and Simulation, Perth, Australia, 12-16 December 2011 (http://mssanz.org.au/modsim2011)