Efficient Delivery of Forecasts to a Nautical Sports Mobile Application with Semantic Data Services

Ricardo Carvalho
Amorim
Faculdade de Engenharia
Universidade do
Porto/INESC-TEC
rcamorim@fe.up.pt

Artur Rocha
INESC Technology & Science
artur.rocha@inesctec.pt

Marco Oliveira
INESC Technology & Science
marco.a.oliveira@inesctec.pt

Cristina Ribeiro
Faculdade de Engenharia
Universidade do Porto /
INESC-TEC
mcr@fe.up.pt

ABSTRACT

Weather and sea-related forecasts provide crucial insights for the practice of nautical sports such as surf and kite surf, and mobile devices are appropriate interfaces for the visualization of meteorology and operational oceanography data. Data are collected and processed by several agencies and are often obtained from forecast models. Their use requires adaptation and refinement prior to visualisation. We describe a set of semantic data services using standard common vocabularies and interoperable interfaces following the recommendations of the INSPIRE directive. NautiCast, a mobile application for forecast delivery illustrates the adaptation of data at two levels: 1) semantic, with the integration of data from different sources via standard vocabularies, and 2) syntactic, with the manipulation of the spacial and temporal resolution of data to get effective mobile communication.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5 [**Information Interfaces and Presentation**]: User Interfaces; C.2.4 [**Distributed systems**]: Distributed databases

1. INTRODUCTION

The practice of outdoor sports is dependent on atmospheric conditions that are permanently monitored by an array of institutions. Sports directly related to the state of the sea, where sportspersons are physically exposed, such as surf and kite-surf, are greatly dependent on both weather and sea conditions and become much safer with accurate

forecasts for regions near the coast.

Several kinds of institutions collect weather and sea-related data: national and regional meteorology agencies, port authorities, military marine departments and research institutes belong to this group. As data are gathered on a systematic basis, there are well-established supporting infrastructures that coordinate the collection of data in the scope of their field of activity, making them usable for a plethora of application-specific purposes. These data can be generated by global observation systems such as satellite networks, from campaign vessels, or even from devices deployed on specific constructions at the seashore or in the sea, such as multi-parametric buoys (including, for instance, submerged sensors). Such devices are used to read atmospheric parameters such as wind direction, temperature and humidity, or sea surface ones like maritime agitation, wave height and wave period.

Another step in data usage is the execution of mathematical models to process sensor data and obtain forecast products, with spatially and temporally referenced values for selected variables. *In-situ* readings are used as reference inputs at given points, to fine tune the mathematical models used to predict behavioural changes for a specific time coverage. Global predictions are therefore complemented by higher resolution outputs near the coast.

Each data producer typically uses specific formats, which are well-known inside their community but significantly different from other communities. The implicit semantics for these formats is therefore hard to process in a systematic way. Furthermore, different institutions from different countries may be involved in data collection for a given region, leading to increased difficulties when processing the data for visualisation or other kinds of integrated processing. Nevertheless, institutions are increasingly adopting standard data and metadata encodings, interoperable interfaces and common vocabularies to facilitate the exchange of both data and metadata.

This work is based on data collected for the northern Portugal and Galicia regions, in the context of projects resulting from European cross-border collaborations. The data collecting agencies provide expertise on data collection and organisation, execute models on selected data to generate

forecast products on a regular basis, and identify requirements concerning new forms of data exploration. Moreover, the research institutes are engaged in modelling data and aggregating metadata to build robust storage and interrogation infrastructures and services, as well as in the development of new applications based on these services.

The paper provides a synthetic view on existing data collecting infrastructures for the target region and on the set of services designed to integrate data within a common semantic framework, leveraging on the data systematically collected and processed by several institutions. The semantic services are tested on a mobile application to assist users as they choose a suitable place to practice sports that depend heavily on sea and weather conditions. The development of the application is a challenge at two levels. The first concerns the interface to the semantic services. A successful implementation in this case shows that the services satisfy several general requirements. The second has to do with the data manipulation in mobile applications. Visualising forecast data in mobile devices has specific challenges concerning the data volumes, the fusion of data from several sources, and the interface design. The result is an application which is both rigorous on data presentation and appropriate for the intended users. The decisions concerning data manipulation and visualisation are detailed and provide insights for other applications dealing with geo-referenced forecast data.

2. DISTRIBUTED FORECASTS

At a high level, the Nauticast application uses several kinds of information resources, namely data, operational models and resulting products. Data consists either of basic sensor data or those resulting from processing. By models we mean mathematical models devised by experts in forecasts, which are run on raw or refined data. Products designate selected outcomes of models which are exchanged between experts or disseminated to the public. An example of a product presentation is the detail of a page of forecasts for surfers presented in Figure 1. The horizontal axis has time periods (12H00, 18H00, 00H00, 06H00). The bottom part of the figure shows values for three variables, wind, wave period and Surf Quality Index, in three distinct locations. The top part aligns this information with values for tide and surf heights. The resources used in the data services and in the mobile application were assembled in the context of the RAIA Oceanographic Observatory [13] as a result of a coordinated cross-border effort by a set of institutions (a meteorological agency, a navy technical department and several research groups) to interoperate their systems, build a higher-level service-based infrastructure, and improve the knowledge of the maritime environment for the region of the Northwest Iberian Peninsula.

The set-up of the distributed infrastructure required some adaptation of the partner communities, adopting good practices both from the Earth and Ocean Sciences (EOS) and the Geospatial Communities (OGC) [13]. It takes advantage of existing standards and aims to contribute to the implementation of the European INSPIRE Directive [4] by 2020, as is legally required from EU member states. All the partner institutions are committed to make their data promptly available after running the forecasting models. There are however several obstacles, ranging from the lack of investment on data infrastructures to the absence of actual applications for some datasets that leads to their neglect.
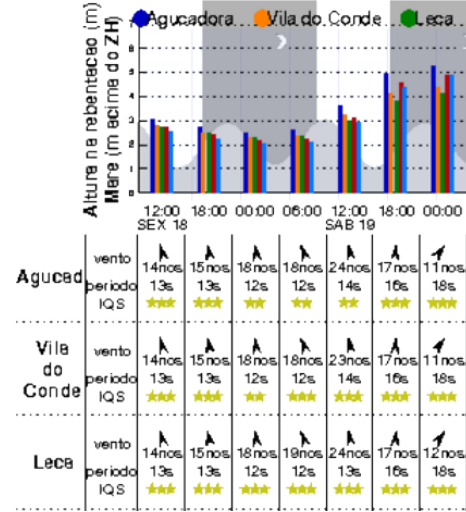

Figure 1: Wind, wave period and Surf Quality Index data, a detail from `http://www.hidrografico.pt/previsao-surf-porto.php`

For the surf and kite-surf predictions, the app relies on atmospheric forecasts by MeteoGalicia (Spain) and on more specific nautical forecast products from operational oceanography provided by Instituto Hidrográfico, a division of the Portuguese Navy. To tackle semantic issues derived from encoding at the dataset level, the partners agreed to comply with the Climate and Forecast (CF) conventions (CF version 1.4 or above) to specify which data each variable represents, as well as their spatial and temporal properties. In doing so, each party can keep their own localised preferences concerning *long names* and *descriptions*, which naturally arise as two countries with different languages are involved, while adhering to the same *standard name* to facilitate integrated processing and visualisation.

All institutions agreed on the open access to semantically rich data through a geographically distributed infrastructure of standard services and data formats. Each institution has its own systems and is in charge of supporting the corresponding services and data.

To provide forecasts, the operational models usually generate outputs as NetCDF [8, 5] files that are harvested by a THREDDS Data Server (TDS) [9], thus making them immediately available to prospecting clients. TDS implements standard interfaces (APIs) for data and metadata access services. APIs such as Web Map Service (WMS) [3], Web Coverage Service (WCS) [14] or Network Data Access Protocol (OPeNDAP) [2], among others, can be harvested by a Catalogue Service for the Web [11], and accessed by other services or systems, thus enabling the creation of added value applications through the fusion of datasets which would be hard to relate otherwise.

Institutions apply multiple numerical models such as Wavewatch III (WW3) [12], Wave Model for Coastal Regions (SWAN) [1], Regional Ocean Modeling System (ROMS) [7] or Weather Research and Forecasting (WRF) [10], with distinct enforcements, spatial and temporal resolution, to obtain forecasts for different uses. Each output of a model execution generates a downscaled dataset of multi-variable spatio-temporal data.

Even if the spatial extent of the forecasts is the same, a software application may have to display gridded data at very different scales and possibly with time gaps for which not all forecasts are available. Apart from these issues, since high resolution models result in large volumes of data and transferring large datasets to mobile devices is not cost-effective due to data fees, the implementation of an intermediate preprocessing service is essential in this case.

In the context of forecast data processing, dealing with several, large datasets with high-resolution data hinders direct access by a mobile device. A pre-processing routine was developed to address this issue and make it possible to deploy the created dataset to the mobile devices displaying it to the end users. Nevertheless, as it is common for several institutions to collaborate and produce specific parts of the dataset, the fusion of these parts has to be executed prior to their refinement.

3. SEMANTIC DATA FUSION

The integrated visualization of multi-variable spatio-temporal data in mobile devices poses interesting issues concerning the high volume of exchanged data and the amount of processing power required to properly render them. Data from diverse provenances and with different resolutions needs to be merged and tailored before being sent over to users' devices. On the course of this section an approach to solve these issues and make data systematically available to the end-users will be detailed.
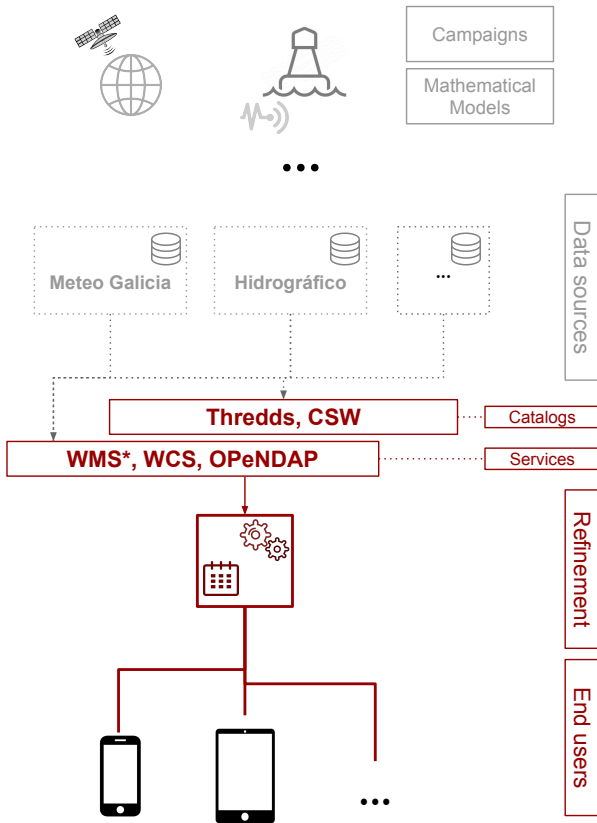

Figure 2: Outlined representation of the data flow, starting from the data sources until the end-user

3.1 Handling data diversity

As each device needs to retrieve the same resources to display, a service to pre-process data and make them available to the clients was created. All data is thus gathered and centralised in this node, that also handles the different spatial and temporal resolutions that are retrieved from the geographically distributed centres.

This centralised node has the additional advantage of reducing the dependency from multiple distributed services by also becoming a cache service, enabling data access even if the original data services are momentarily down.

Figure 2 shows the two main institutions involved in this particular scenario: Meteogalicia[1], providing data related to the weather conditions, such as wind speed and direction; and Instituto Hidrográfico[2] responsible for the sea-related variables—wave period, significant height and wave direction. In both cases, the forecasts are generated by running the mathematical models and enforcing them to other datasets coming from global forecasting model outputs, satellite imaging, buoys deployed overseas or various campaign vessels.

Institutions also deal with sensor networks and the concept of data fusion, i.e. merging data from heterogeneous arrays of sensors, has been discussed [15]. The resulting data is then made available either directly from the existing catalogs or through services that implement standard interfaces for established output formats.

To start with, the collection of the required oceanographic and atmospheric datasets has to deal with issues such as network delays and server maintenance, considering the geographic dispersion of the involved infrastructures. An application can take advantage of the built-in NetCDF remote access client[3] and query a remote file for specific variables or coverages. Preliminary testing with such approach revealed some issues when querying each of the fields using the remote access client, so the choice fell on retrieving the whole dataset for each forecast.

In the process of merging the datasets, some discrepancies concerning spatial and time resolution and repository organisation were found. As expected, the outputs of each institution show subtle differences in the geographical coverage, and the way this data is made available to the public follows local, organisation-dependent conventions. After starting to fetch data for the wind-related forecasts, we also noticed that 1) the time coverage for that data was different on datasets from different institutions; and that 2) the geographic boundaries were, in some cases, also different.

After retrieving the raw datasets, processing them into the single dataset requires normalisation of time and spatial resolution. To handle the existing diversity, the higher resolution dataset was retrieved, then the necessary data variables have been extracted along with their temporal and spatial boundaries. These were then used as criteria to query the remaining datasets. A set of points of interest have also been selected for displaying in the devices. Coordinates for these spots are passed onto the next stage to be associated with the values for the remaining variable.

3.2 Data resolution

---

[1] http://www.meteogalicia.es
[2] http://www.hidrografico.pt
[3] https://www.unidata.ucar.edu/software/netcdf/docs/remote_client.html

| (a) | (b) | (c) |

Figure 3: Application screen shots showing 3a) the initial sport selection; 3b) the simplified view with the quality index variable shown; and 3c) the detailed view of a specific region

Table 1: Comparison of the available data for two variables–Wave significant height (Hs) and wind Gust—extracted from both sources. Values are for December 8th 2015

[41.0, -8.9]	Reference Dataset	Weather dataset
Time / Variable	Hs	Gust
T00:00:00.000Z	1.9983872	11.65399
T01:00:00.000Z	–	11.65399
T02:00:00.000Z	–	11.554
T02:00:00.000Z	–	11.554
. . .	. . .	. . .
T06:00:00.000Z	1.9972923	11.35399

Although each institution applies internally defined spatial and temporal settings, a bigger issue emerges when considering displaying such data on mobile devices. As it is important to maintain the temporal extent of the forecasts, the spatial resolution has been reduced instead, in order to downsize the volume of the merged dataset. The goal is to facilitate the download of the dataset from the server, and more importantly, to ensure that a larger number of devices is capable of loading such dataset.

The THREDDS data server has a sub-setting service for NetCDF-based files that allows developers to get specific portions of a dataset, either specifying a bounding box, a temporal coverage or both. Additionally, it is also possible to define, for each of the requested variables, a stride— n— that samples each n-th value of the selected variable. To merge the datasets, the spatial resolution is reduced to match a grid size that is appropriate for all the forecasts

and good enough for the desired usage. This means obtaining a common geographic reference, with a specified spacing for all datasets, allowing quick evaluation of the results of resolution changes in the device's interface.

Concerning time coverage, the existing differences between datasets rely, at the moment, on one providing forecasts every 6 hours and others having more frequent data, decreasing that spacing up to 1 hour intervals. Again, the first dataset is used as a master and the corresponding data is retrieved from the other institutions. As shown in Table 1, actual data is only extracted when the involved datasets provide values for the needed variables. In this case, we retrieve two points (highlighted) from this block. From the users' point of view, this still ensures forecasts accuracy, and keeps the involved data exchange to the minimum without compromising visual quality. As these models are very accurate, applying a dataset for half of the time coverage yields little to no accuracy loss for the kind of users and applications we are dealing with.

4. END USER DATA VISUALIZATION

To address the constraints in visualizing fairly large forecast data through limited interfaces, we implemented a process to merge separate datasets and assemble smaller ones with negligible loss of accuracy. To display this data, we used well-accepted representation models in data visualisation. An example of such representations are flux curves, proposed in the context of wind data visualization by *Beccario*[4]. Despite their visually-rich interfaces, these methods assume an underlying layer of datasets with high spatial res-

---

[4]For a real-time demonstration of this visualization model, please refer to: `http://earth.nullschool.net/`

olution, and therefore the visualization of this kind data in constrained environments is often limited to basic, tabular representations.

4.1 Weather data visualization in NautiCast

Although simple interfaces are flexible and can display significant amounts of data, their lack of precision becomes visible when considering single points in a map, instead of larger geographic regions. Animations have been proposed to provide a richer visualization in small devices, representing different variables in a map view [6]. This allows, for instance, the display of wind motion and speed using moving particles, and the use of static arrows to represent wave characteristics.

Figure 3 shows the application interface. The application starts by highlighting the two target sports for this particular scenario: surf and kite-surf, as shown in Figure 3a. Cached data is immediately available in the mobile device at application start. Background tasks check for updated datasets. If they are available, the dataset is automatically updated. The application is organised around the forecast conditions for specific spots. Figure 3b, for instance, has information for the surroundings of the "Leça" beach. Selected spots are popular due to their favourable characteristics for the practice of surf, kite-surf or other nautical sports.

Data for each spot is a detail of the larger dataset, representing a summary obtained by mathematical models and empirical formulations applied to those spots. Such detailed data includes a set of selected variables and their corresponding values for the spot and, in case of a surf forecast, a rating of one to three stars. The stars correspond to the Index of Surf Quality (IQS), computed by Instituto Hidrográfico considering the variables that contribute to the overall quality of the spot, for a particular time instant. This rating provides a global, easy to grasp information from the surf forecast for that specific spot.

Users can have a more detailed view of a selected region (see Figure 3c), encompassing several spots, where the remaining data—sea and wind conditions—are included. This view shows the merged pre-processed dataset with image resources extracted from the available WMS web services. These images are downloaded together with the dataset and the device's screen resolution is considered to generate appropriate images. This allows a richer visualization, complementing iconographic symbols and flux curves with a background image that shows different colours depending on wave heights, using a pre-defined scale.

4.2 Weather visualisation initiatives

Delivering forecast data for public information has been explored in other applications. Several projects propose optimized interfaces and refined data to deliver forecasts to the general public. Services such as AccuWeather[5] can display weather information on a variety of interfaces, including mobile devices. Nevertheless, their approach is generic, and therefore more specialised information, relevant to specific sports, is not considered.

On the other hand, initiatives such as WindGuru[6] and WindFinder[7] offer information at a deeper level, in some cases targeted at nautical sports. Analysing the methodology used to harvest and process weather data, these projects apply similar approaches to build global forecast datasets. This includes the selection of a set of spots—well known for their suitability for this kind of sports—with indicators rating their current conditions in a predefined scale.

Despite their successful delivery of forecast data to the users focused on a specific sport activity, existing applications rely on globally generated data for their predictions. This means that data generated and optimised locally by dedicated institutions is not considered in the process. This allows a broader set of places with existing data in the global mesh, but does not take advantage of a wealth of more detailed and geographically distributed datasets. Prediction accuracy can be improved with local information, but the use of data from different sources call for a different approach. Institutions that collect and process local, high-precision forecast data, can be included in the overall workflow, ensuring a higher resolution dataset whenever possible and, at the same time, providing better predictions for known spots.

5. CONCLUSION AND FUTURE WORK

Generic forecast data is often organised in complex structures and consists of large datasets. The direct delivery of this data to the end users does not facilitate an easy interpretation and poses technical problems when the reading devices have limited resources. Among researchers in weather and ocean forecasts, standards for data and metadata harvesting are already in place, and are increasingly important for the goal of creating an environment of interoperable, worldwide services. Data services can demonstrate their importance when considering data visualization, which requires an array of output formats for resources such as images and data subsets.

We describe a generic approach to deliver forecast data to the end users, focusing on the specific scenario of nautical sports that are heavily dependent on weather conditions.

Our project has built on forecast data from several sources and a set of services providing an extra layer of abstraction on them. Projects RAIA (0520-RAIA-CO-1-E) and RAIA-TEC (0688-RAIATEC-1-P) set the baseline for this work and provided essential insights on how to tackle interoperability issues in this particular domain. Our goal was to test whether the existing services could support an application aimed at end users, using simple devices, and taking advantage of datasets with different spatial and temporal resolutions.

The workflow we adopted includes running daily routines to extract the appropriate datasets from the partner institutions and to merge them into a single, lower resolution and compressed dataset that can be easily ingested by smaller devices such as tablets or smartphones.

For the end user, focused on getting quick and visible feedback from the forecast data, we consider that there was little to no compromise in terms of both data quality and accuracy. In the end, the implemented workflow proved to be robust with respect to the diversity of data sources and institutions, namely in terms of temporal coverage and spatial resolutions. The same approach can be used in other application domains, where users depend on local forecasts for specific needs.

---

[5] http://www.accuweather.com/

[6] http://windguru.cz

[7] http://www.windfinder.com/

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] N. Booij, R. Ris, and L. H. Holthuijsen. A third-generation wave model for coastal regions: 1. model description and validation. *Journal of Geophysical Research: Oceans (1978–2012)*, 104(C4):7649–7666, 1999.

[2] P. Cornillon, J. Gallagher, and T. Sgouros. Opendap: Accessing data in a distributed, heterogeneous environment. *Data Science Journal*, 2(5):164–174, 2003.

[3] J. de La Beaujardiere. Opengis® web map server implementation specification. *Open Geospatial Consortium*, 2006.

[4] I. Directive. Directive 2007/2/ec of the european parliament and of the council of 14 march 2007 establishing an infrastructure for spatial information in the european community (inspire), 2007.

[5] B. Domenico. Ogc network common data form (NetCDF) core encoding standard version 1.0. *Open Geospatial Consortium*, 10:e090r3, 2011.

[6] R. F. C. Gomes. Visualização de dados oceanográficos espácio-temporais multivariáveis em dispositivos móveis. Master's thesis, Universidade do Porto, 2014.

[7] D. B. Haidvogel, H. Arango, W. P. Budgell, B. D. Cornuelle, E. Curchitser, E. Di Lorenzo, K. Fennel, W. R. Geyer, A. J. Hermann, L. Lanerolle, et al. Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the regional ocean modeling system. *Journal of Computational Physics*, 227(7):3595–3624, 2008.

[8] H. L. Jenter and R. P. Signell. Netcdf: A public-domain-software solution to data-access problems for numerical modelers. In *Preprints of the American Society of Civil Engineers Conference on Estuarine and Coastal Modeling*, volume 72, 1992.

[9] U. John Caron and E. Davis. UNIDATA's THREDDS data server. In *22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, 2006.

[10] J. Michalakes, S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock. Development of a next generation regional weather research and forecast model. In *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the use of high performance computing in meteorology*, volume 1, pages 269–276. World Scientific, 2001.

[11] D. Nebert, A. Whiteside, and P. Vretanos. Opengis® catalogue services specification. *Open Geospatial Consortium*, 2007.

[12] H. Tolman, M. Accensi, H. Alves, F. Ardhuin, J. Bidlot, N. Booij, A.-C. Bennis, T. Campbell, D. V. Chalikov, J.-F. Filipot, et al. User manual and system documentation of wavewatch iii version 4.18. 2014.

[13] B. Vila, A. Gómez, C. Cortizas, P. Dias, M. Hermida, M. Oliveira, A. Rocha, and X. Méndez. RAIA Observatory: Visualization of Oceanographic Data under INSPIRE Directive. In *Actas das 2as Jornadas de Engenharia Hidrográfica*, pages 369 – 372. Instituto Hidrográfico, 2012.

[14] A. Whiteside and J. D. Evans. Opengis® web coverage service (wcs) implementation standard. *Open Geospatial Consortium*, 2008.

[15] A. Wun, M. Petrovi, and H.-A. Jacobsen. A system for semantic data fusion in sensor networks. In *Proceedings of the 2007 inaugural international conference on Distributed event-based systems*, pages 75–79. ACM, 2007.