Web Services for Environmental Data Access for Popular Applications

Rafael Duarte Coelho dos Santos, Rogério B. Andrade and Nandamudi L. Vijaykumar
Assoc. Lab. for Computing and Applied Mathematics
National Institute for Space Research
Av dos Astronautas, 1.758, Jd. Granja,
São José dos Campos, CEP 12227-010
São Paulo - Brasil

Eduardo Batista de Moraes Barbosa
Weather Forecast and Climate Studies Center
National Institute for Space Research
Rodovia Presidente Dutra, Km 40, SP-RJ
Cachoeira Paulista, CEP 12630-000
São Paulo - Brasil

Abstract—The Weather Forecast and Climate Studies Center (CPTEC) of the Brazilian National Institute for Space Research (INPE) is responsible for weather and climate forecasts and weather data collection and dissemination. Any user can access the data through CPTEC’s site to get data for his own applications, but must search, collate and process the data himself, which is not always practical. In this paper we present our ongoing development project that will allow access to the data through an unified collection of web services, that, through composition, allows end users of the data to create more complex, specialized queries with little effort.

I. INTRODUCTION

One of the main activities of INPE is the collection, processing and dissemination of data and information on weather, climate, space weather, etc. This data is collected from several types of sensors related to multiple missions INPE has to execute, and are very important to study the various phenomena that may have a significant impact on the environment, agriculture, tourism, climate change, etc., and can also be used in a variety of educational and outreach applications.

Each mission has its own methods, protocols and policies, and often uses specific, non-shared computational resources for data processing and distribution. Some of the raw and processed data is made available, often through the World Wide Web, for the community in general, but usually only in the ways the missions’ administrators found adequate.

Informal interview with users of the data dissemination servers at CPTEC (the Weather Forecast and Climate Studies Center) at INPE indicates that they want answers to questions that may be answerable with the existing data, but requires different access mechanisms. In this paper we present an ongoing development effort to grant access to environmental data from multiple sources through web services, which does not require modification on the existing infrastructure and which can be used to answer simple, popular questions about the data and the composition with other services for more complex data.

It is expected that such infra-structure enables developing popular applications not only to attend specialists but also population in general that are more conscious and curious to understand several phenomena.

II. MOTIVATION AND CASE STUDY

The main motivation for our development project is to allow users to get answers to questions that were not considered when the original data collection and dissemination mission was planned and executed, but that could be answered by that data. Some of those questions are actually asked to CPTEC’s user service department, and a very reduced selection is shown below:

1) Which is the average rainfall for my city in October?
2) Where was the coldest recorded temperature in the winter of 1990 in São Paulo State?
3) Which is the highest ever rainfall in a day registered in Porto Alegre city and in Rio Grande do Sul state in general?
4) Considering cities close to the equator line, which has the highest average rainfall?
5) Which cities in Bahia state are similar with respect to their annual temperatures?

This list is of course non-exhaustive; there are many questions that involve both past data (from several different sensors), satellite imagery and forecast data. The list is also shown in order of apparent complexity (easy questions require fewer queries executions and extraction of the results), and questions were selected to cover mainly rainfall and temperature, obtained from the Data Collection Platforms (PCD) network which integrates the Brazilian Data Collection System [1], a well-known environmental data source.

All questions could be answered by queries to existing web-based data dissemination systems, but would require increasingly complex collation procedures, therefore are not easily reproducible or modifiable. To illustrate this problem, Figures 1 and 2 show how one can submit a query about precipitation in a location and time frame and get its results; these are complete and user-friendly, but are presented in HTML format which is not really adequate for further processing [2].
III. Proposed Framework

The proposed framework is a collection of web services to allow direct access to the data, working in parallel with the existing data dissemination tools. Web services allow machine-to-machine communication using standards, being loosely coupled, language and OS independent and (usually) efficient for transmission of small chunks of data between server and client.

Several small, self-contained web services were developed so far. These web services allow the extraction of data from the databases accordingly through some canned queries, which return single values or small arrays of data. These web services can be chained in workflows to answer more complex queries.

IV. Examples

One example is provided to illustrate the answering of question 1 with the framework, using Taverna [3] (an open-source suite of tools for the visual composition and execution of workflows) to create a simple workflow using the web services on the framework.

Figure 3 shows the visual workflow edited in Taverna to answer question 1. Execution is done in the Taverna environment, but the entire workflow could be embedded into other applications or even executed remotely. The workflow is relatively simple, since part of the processing of the query (calculating average measures over a time interval) is done in the server.

Workflows can also be developed using only code in any language that contains APIs (application programming interfaces) that can consume web services. Depending on the purpose workflows can also be deployed as new web services or as functions that extend the existing data dissemination systems.

V. Future Work

The present implementation is just a small subset of what was proposed as the scope of a research and development project funded by the Brazilian Space Agency (AEB) and Brazilian Research Council (CNPq), which eventually will serve data and metadata from several missions hosted at INPE’s computers, and allow the integration with other web services and frameworks hosted by other institutions.

Acknowledgments: The authors would like to acknowledge the grants provided by the Brazilian Space Agency (AEB) and Brazilian Research Council (CNPq), process number 560188/2010-2.

REFERENCES