ISPRS Journal of Photogrammetry and Remote Sensing 114 (2016) 294–305

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Integrating geo web services for a user driven exploratory analysis ☆

Simon Moncrieff [a,b,*], Ulanbek Turdukulov [b], Elizabeth-Kate Gulland [b]

[a] Cooperative Research Centre for Spatial Information (CRC-SI), Australia
[b] Department of Spatial Sciences, Curtin University, WA, Australia

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form 14 January 2016
Accepted 19 January 2016
Available online 18 February 2016

Keywords:
Geovisual analytics
Server side
Health research
Geo web services
User driven analysis

ABSTRACT

In data exploration, several online data sources may need to be dynamically aggregated or summarised over spatial region, time interval, or set of attributes. With respect to thematic data, web services are mainly used to present results leading to a supplier driven service model limiting the exploration of the data. In this paper we propose a user need driven service model based on geo web processing services. The aim of the framework is to provide a method for the scalable and interactive access to various geographic data sources on the web. The architecture combines a data query, processing technique and visualisation methodology to rapidly integrate and visually summarise properties of a dataset. We illustrate the environment on a health related use case that derives Age Standardised Rate – a dynamic index that needs integration of the existing interoperable web services of demographic data in conjunction with standalone non-spatial secure database servers used in health research. Although the example is specific to the health field, the architecture and the proposed approach are relevant and applicable to other fields that require integration and visualisation of geo datasets from various web services and thus, we believe is generic in its approach.

1. Introduction

Geographic information technologies are evolving from stand-alone systems to a distributed model of independent web services. Many international organisations, governments and businesses on all levels encourage distribution of spatial data as geo web services using Open Geospatial Consortium's (OGC) specifications.[1] As a result, OGC services are presently a common way of accessing spatial data on the web.

OGC provides a number of standards to handle spatial data online, such as Web Map Service (WMS) and Web Coverage Service (WCS) for manipulating raster data and Web Feature Service (WFS) for handling vector data, and to access spatial operations online, such as Web Processing Service (WPS). Broadly speaking, WMS can be used by data providers for data presentation, WFS and WCS can be used for providing access to data, and WPS for transforming the data resulting in a derived product. Each web service responds to text-based requests with optional parameters such as output format and data filters, accessed via applications which build requests from software or user inputs. But in such cases, WMS, WFS and WPS geo web services are pre-configured to provide a presentation of spatial layers that are each a fixed combination of data and processes at a certain spatial scale, at given time frames and with a defined number of attributes. Thus, a service provider essentially decides which datasets, or views on a dataset, a user should be able to access.

The users of these geo web services need to explore and interact with spatial information in order to make timely decisions. Often such exploration is fluid across multiple spatial and temporal scales and attributes, requiring researchers to switch from a local to a global view and back in order to gain insights into different dimensions of the dataset. For example, a health researcher might need to understand environmental factors that influence disease outbreaks affecting certain age groups at specific income levels in a subset of suburbs and compare those with the global indicators for other area types such as county, state or country.

With spatially contextual data, such as health records which impose point or area locations on other attribute values, a user driven model is desirable. Such a model should incorporate ad-hoc queries, guided by a data administrator to determine relevant filter

☆ The research was sponsored by the Cooperative Research Centre for Spatial Information (CRC-SI).

* Corresponding author at: Department of Spatial Sciences, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia. Tel.: +61 892662288.

E-mail addresses: s.moncrieff@curtin.edu.au (S. Moncrieff), ulanbek.turdukulov@curtin.edu.au (U. Turdukulov), E.Gulland@curtin.edu.au (E.-K. Gulland).

[1] OGC standards – http://www.opengeospatial.org/standards/l, accessed 12 December 2014.

http://dx.doi.org/10.1016/j.isprsjprs.2016.01.015
0924-2716/

parameters and analyses for processed outputs. This combination of ad-hoc queries and administrator controls is necessary to cater for both the need for data exploration and the inability to describe data adequately at a given spatial scale without prior processing of the data. Further, a user driven model should facilitate the ability to refine a query and generate a result within a Web GIS environment using a single interface to the dataset, enabling users to interact with spatial data via various web services without knowledge of the complexities of data integration. The required web service delivery approach to achieve this would necessarily combine elements of existing OGC web service types.

In this paper we propose such a user need driven web service model based on GeoServices REST specifications[2] and combine these with the principles of visual analytics – a science of analytical reasoning facilitated by interactive visual interfaces (Cook and Thomas, 2005). The aim of the framework is to provide a method for scalable and interactive access to various geographic data sources on the web and visualisation methodology to rapidly integrate and visually summarise the properties of a dataset. The framework extends the WPS approach to enable ongoing exploratory and user driven spatial analysis across multiple attributes, datasets, and spatial and temporal scales. It can be used to rapidly explore alternative visualisations, and can also produce results conforming to OGC standards for compatibility with further processing in WPS workflows or software applications.

We illustrate the environment on a health related case study that integrates the processing of health records and demographic data that include a spatial context. The processing in question relates to deriving statistical results based on underlying unit record patient information. Although the example presented is specific to the health field, the architecture and the proposed approach are applicable to any other field that requires rapid integration and visualisation of geo datasets from various web services and thus is generic in its approach.

The model proposed in this paper comprises a server side method enabling the dynamic exploration of, and interaction with, high granularity sensitive data using a software agent based approach. This approach comprises creating a modular data software agent that enables access to a data source through a discoverable web service interface, with an architecture that combines a data query module and processing techniques specific to the data source. The result is interactive data exploration, realised through flexible query and analysis, while ensuring the unit record data is not accessible from outside the data administrators' infrastructure. This approach would lead to decentralised, distributed data access due to the specialisation associated with the data, with each data provider being responsible for the dataset under their purview. In addition, a single themed portal communicating with web services for various datasets could be used to centralise data exploration, for example, through a selection of visualisations provided by the portal.

There are a number of advantages to the proposed data agent approach:

1. **Privacy**. While the raw data remains on the server, only the results of user driven queries are returned to the client. This allows access to processed results derived from potentially sensitive raw datasets without such data being transferred to the client, or indeed, without the data even leaving the data supplier's infrastructure. Further, compulsory rules and filters can be placed on the server, further restricting the output where necessary. This is especially important in privacy sensitive fields such as health research.
2. **Provenance**. Disregarding the privacy concerns, if the unit record data were published, the subsequent analysis and summarising could lead to possible misinterpretations. While this issue can be somewhat addressed through licensing, publishing an authoritative result could verify such outputs. Consequently, using the proposed method, access to statistical summaries controlled by the data administrator can be made available as the authoritative source of information for the given dataset. In this way, an administrator can enforce certain processing properties such as statistical validity and reference populations.
3. **Flexibility**. For health data analysis, a typical query can comprise a choice of one of 149 disease type groupings, with a choice of ages ranging from 0 to 85, over 3 gender types, choice of year ranges spanning 10 years, and at 8 spatial scales that easily results in approximately 574 million potential map layers, each comprising a different set of outputs such as disease counts and statistical summary information. For 11 years of data, the number of potential layers increases to 702 million, and for 12 years approximately 842 million layers can be generated, and so on. Data are often published in the form of predefined views, or outcomes; this constitutes a supplier push model. From the number of potential layers required, a supplier push approach of publishing all potential layers is not scalable where flexible data access is required as not all views can be accounted for. The approach of enabling a user to interact with spatial data on the fly, a user pull approach, can provide the required flexibility.

The layout of this paper is as follows: Section 2 outlines relevant approaches to integration of geo web services and combining those with principles of visual analytics; Section 3 introduces a case study that requires analytical and processing functionality using web services; Section 4 describes the main components of the architecture and role of the proposed components as well as the overall workflow associated with a spatial agent. Finally, Section 5 describes results and examples of visualisations that can be achieved using the resulting linked information.

2. Background

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces (Cook and Thomas, 2005). Card et al. (1999) provide a basic reference model for visualisation (Fig. 1). Visualisation is described as the mapping of data to a visual form that supports human interaction and understanding of data. There are three processes to support the understanding – data transformations, visual mapping and view transformations. Data transformation maps raw data into data tables with relational descriptions of the data along with metadata. Visual mappings transform data tables into visual structures that combine spatial substrates, marks, and graphical properties. View transformations create views of the visual structures by specifying parameters such as position, scaling, and clipping. Users can interactively change these transformations to perform their visual analysis tasks.

During data analysis, analysts engage in confirming or deriving hypotheses by interactively exploring data using various analytical activities such as summarising data, making predictions and identifying trends, patterns and outliers to derive new knowledge (Myatt, 2007). Exploration of data is a social process (Heer et al., 2007; Viegas and Wattenberg, 2006) since collaborators also have to be aware of their findings to avoid redundant rediscovery and loss of time by inadvertently repeating an analysis process. Many analysts collaboratively investigate the data with different analysis goals within an organisation. They need to review and share their findings as well as their analysis process. In the last case, commu-

[2] GeoServices REST Specification (ESRI) – http://www.esri.com/industries/landing-pages/geoservices/geoservices, accessed 12 December 2014.