Composing enterprise mashup components and services using architecture integration patterns

Yan Liu\textsuperscript{a,b,*}, Xin Liang\textsuperscript{b}, Lingzhi Xu\textsuperscript{c}, Mark Staples\textsuperscript{b,c}, Liming Zhu\textsuperscript{b,c}

\textsuperscript{a} Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA, USA
\textsuperscript{b} University of New South Wales, NSW 2052, Australia
\textsuperscript{c} NICTA, 13 Garden St, Ewleigh, NSW 2015, Australia

\begin{abstract}
Enterprise mashups leverage various source of information to compose new situational applications. The architecture of such applications must address integration issues: it needs to deal with heterogeneous local and/or public data sources, and build value-added applications on existing corporate IT systems. In this paper, we leverage enterprise architecture integration patterns to compose reusable mashup components. We present a service oriented architecture that addresses reusability and integration needs for building enterprise mashup applications. Key techniques to customize this architecture are developed for mashups with themed data on location maps. The usage of this architecture is illustrated by a property valuation application derived from a real-world scenario. We demonstrate and discuss how this state-of-the-art architecture design method can be applied to enhance the design and development of emerging enterprise mashups.
\end{abstract}

1. Introduction

The concept of enterprise mashup stems from the mashup development of Web 2.0. Mashup enables individual users to retrieve content from several Internet sources to compose highly personalized applications on the Web. Mashup development has now been extended to the enterprise domain whereby a broad range of local or public information\textsuperscript{1} can be accessed through various mechanisms such as REST (representational state transfer) APIs or Web feed formats such as RSS (really simple syndication), ATOM (atom syndication format), and JSON (JavaScript object notation). Leveraging this kind of information brings opportunities to develop new situational applications (Mohammadi et al., 2009a) to improve business productivity or repurpose existing applications with value-added features. An example scenario in this context has been presented in our previous work (Liu et al., 2009): a property valuation service (PVS) company which uses an internal service to retrieve historical data about the valuation results of individual areas. Such data informs the house price valuation process when combined with other public information such as traffic and crime statistics. Mashup of both local and public data sources on a location map enriches the information set for a valuer\textsuperscript{2} to make an accurate decision on-site.

The current development of enterprise mashups is ad hoc and mostly from scratch. A developer usually chooses a preferred development environment such as Yahoo Pipes, Google Mashup Editor, Microsoft Popfly or QedWiki, and manually combines data at the syntactic level. Many similar mashups use the same tools and APIs but do not systematically share common pattern. This implies that less reusability is achieved. As mashups are emerging applications, there has been little research dedicated to proposing a systematic development methodology, and few best practises are well generalized from industrial practices.

The development of enterprise mashups is at the scale of enterprise integration, involving the issues of organizing flows of public and private data, and transforming data into consistent formats for mashup manipulation. In addition, the architecture of enterprise mashups should consist of reusable architectural components so that data sources or new logic for mashups can be extended.

Among software architecture principles, architecture patterns (or styles) offer the promise to help architects to identify combinations of architecture and solution building blocks that have been proven to deliver effective solutions in the past, which may provide the basis for effective solutions in the future. Although architec-

\textsuperscript{*} Corresponding author at: Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA, USA.
\textsuperscript{1} E-mail address: yan.liu@pnl.gov (Y. Liu).
\textsuperscript{2} http://mashupaustralia.org/.

\textsuperscript{2} A valuer is a professional that assesses the price of a property, usually during an on-site inspection.

0164-1212/5 – see front matter. Crown Copyright © 2011 Published by Elsevier Inc. All rights reserved.
ture patterns have been extensively studied and applied to various application domains of distributed and network systems, to our best knowledge their usage customized to enterprise mashups has not been explicitly studied, neither widely adopted in practice.

The contribution of this paper is identifying the essential architecture patterns and defining their interactions within a generic architecture for designing enterprise mashups. In this paper, we compose key architecture patterns developed in our previous work (Liu et al., 2009) using components and RESTful Web services in extensible design of mashups. In addition, we develop essential techniques to realize this architecture for a map with overlays, however, the approach is generic to other types of mashup presentation such as markers and annotation on a map. We demonstrate our architectural approach and techniques using a scenario derived from a real-world property valuation service used in the Australian lending industry.

The structure of this paper is as follows. We first introduce the motivating business scenario in Section 2 that helps to identify the needs for developing patterns for enterprise mashups. We then summarize the related work in Section 3. Three architecture patterns are discussed in Section 4 to address the needs of mashup design and development. The integration of these patterns are guided by a RESTful architecture in Section 5, followed by a demonstrating Web-based PVS application. This paper concludes with lessons learnt and future work.

2. Motivating business scenario

In this section, we introduce a business scenario for enterprise mashups—a property valuation service derived from our collaboration with an organization in the Australian lending industry (Zhu and Thomas, in press).

PVS is one of the key services in the mortgage lending process. It provides a lender or a broker with an accurate valuation of the real estate in a loan application. PVS is usually provided by an independent valuation firm. When a valuation firm receives a request from lenders or brokers, the valuation process starts, and requires a number of tasks to be completed. The business goal of the whole valuation process is to reply to customers with an accurate property valuation and with minimal response delays.

In the valuation process, three roles cooperate to accomplish this task, namely administrator, valuer and manager. The administrator contacts the owner of a property, schedules the inspection time and assigns a valuer to do the valuation on site. Once the valuer accepts the job, the valuer conducts an on-site inspection, completes the valuation form and sends it back to the valuation firm office. A manager receives the valuation form from the valuer and approves it before the valuation result is sent to the customer.

Although the whole business process is well-defined and the responsibility of each role is clear, the actual interactions between different roles are subtle, and achieving the business goal of guaranteed responsiveness is non-trivial. The outcome depends on the availability of the selected valuer and their working experience in the target suburb. Assigning the job to the wrong valuer (either unavailable or unfamiliar with the market in the area) is likely to incur delays caused by switching valuers and back and forth communication through phone calls, emails or faxes. It could also result in less accurate valuations due to a lack of expertise for special purpose valuations such as commercial property valuations. To avoid these risks, the administrator needs to take into account the valuers’ record in terms of the valuation quality and quantity. This information is available, but administrators need central IT support to access such information.

The lack of flexible and customizable software tools is one major barrier to improve the business process of PVS. One system used in the valuation industry is called valuation management system (VMS). Currently VMS only satisfies basic business requirements and does not provide customizable functions to the internal business process of individual valuation firms. For example, there is no function to allow administrators to quickly access valuers’ historical records for particular suburbs.

Enterprise mashups with customizable information overlays provide a feasible solution to address this problem. The mashup architecture is motivated by presenting information to facilitate the tasks in the PVS process. It consists of data or information from three sources: public information on suburb boundaries, corporate records about valuers, and public information on housing prices. The architecture supports a mashup with overlays for administrators to easily view the available valuers for an area. From the valuer’s point of view, the mashup provides a Web-based valuation form so that the valuer can complete the form while simultaneously accessing information related to the property value, such as the traffic statistics.

3. Background

There are large range of mashup technologies and development tools. One study (Clarkin and Holmes, in press) categorizes them based on their role in the lifecycle of a mashup. In our previous paper (Liu et al., 2009), we broadly categorize them into two groups, the mashup builders and the mashup enablers.

Mashup builders are tools that help developers produce mashup presentations such as IBM QEDWiki and Yahoo! Pipes. These tools can even allow non-developers to compose simple mashups by connecting data sources and operations (such as filtering, sorting, selecting) widgets to create new composite applications. Some other approaches (Zhao et al., 2008) only separate user interfaces from business logic components and allow quick evaluation of the results. However, these mashup builders do not allow finer arrangement of the internal structure of the mashup application, making systematic reuse, performance tuning and loose coupled design difficult. These composite applications are often brittle and at the mercy of data source changes. As argued in (Daniel and Matera, 2009), mashup components do not substantially differ from common web applications and might benefit from traditional web engineering methods and instruments. (Daniel and Matera, 2009) showed how to migrate from traditional web applications to mashup applications but without proposing a reusable structure of mashup applications. Our approach uses the thematic notion and combines this with design patterns to allow a more modular design of the mashup composition beyond data manipulation.

On the other hand, mashup enablers provide functionalities to mashup builders by accessing unstructured data, and making internal and external resources available. For example, Kapow Mashup Server, Openkapow, Tarpipe and Feed43 belong to this category. The primary goal of these mashup enablers is to have pre-built data source adaptors (such as RSS, ATOM, RMDB, Messages and Application-specific APIs) and actuators (such as email/post to other data services’ update functionality). Although they do not impose a black-box internal design of the mashup composition like graphic mashup builders, they also leave the design completely to developers to build from scratch. Other works focus on the conceptual taxonomy of mashup
components (Sheth and Larson, 1990) or project management issues for situational mashups (Mohammadi et al., 2009b). Our approach fills this gap by providing a new and reusable internal architecture from a thematic point of view leveraging existing patterns.

Other research has also put further structure on top of existing data. One major direction is the semantic annotation of data sources in the context of Web Services and RESTful Services (Sheth et al., 2007). Mashups pose new challenges as data is sometimes embedded in XHTML or other presentation structures, rather than being clearly separated. These researchers add semantics to data by annotating them with small domain ontology models. Such semantics could be based on themes. However, semantic web technologies do not provide architectural mechanisms that address how such annotated data should interact with the rest of the control, adaption and presentation layers. We consider semantic annotation to be a formal way of organizing data into themes, which supports our method.

As mashup is a type of integration, existing enterprise integration and composition patterns have been adapted to support mashups (Ogrinz, 2009). Such patterns focus on the interfaces and interactions among different sources rather than the internal architecture. Liu et al. (2007) described an architecture to build Web service composition using mashup techniques. Cetin et al. (2007) proposed a service migration strategy in favor of mashup properties – composition of heterogeneous resources. Zou and Pavlovski (2007) introduced accountable and reliable enterprise mashups. Other ways of designing mashups exist for specific domains and applications, such as searching (Braga et al., 2008) and business processes (Xu et al., 2008).

Our work is complementary to these methods, yet provides a more generic architecture solution of building themed mashups by means of architecture patterns. In our approach, the data retrieval and thematic mashup generation perform as individual services and are seamlessly integrated together using Web-based techniques in the RESTful style. Functions such as semantic filtering and service selection can be added to extend the features of this architecture without affecting the mashup generation.

4. The integration patterns

The integration need of enterprise mashups is combining data and functions from more than one source. The data and the output from any function are of myriad forms and formats. They can be stored at cooperate information systems, available from public domains using RSS feeds or Web services calls. They can be accessed by various protocols, such as DB drivers, RSS, REST, HTTP (Hyper-Text Transfer Protocol) or SOAP (Simple Object Access Protocol).

From the architecture point of view, the integration encompasses three core aspects: the dataflow procedure, the composition of heterogeneous data and interfaces, and different views on the same set of data. We develop an architectural solution for composing mashups guided by integration patterns. Fig. 1 depicts the conceptual level architecture. This approach initially exploits three patterns, namely Pipes and Filters pattern, Data Federation pattern and Model-View-Controller pattern. These patterns by nature are suitable for realizing the core aspects of the mashup architecture respectively: composing data flow, transforming heterogenous data formats, and enabling multiple views on data. The composition of these patterns adapts the RESTful architecture style addressed in the next section.

To apply these generic patterns to address specific mashup needs, our architecture approach follows three steps:

1. identify key components for mashup development,
2. map the key mashup components to individual pattern elements and structure,
3. carefully design the intersecting points of patterns to reduce the risk of cross-cutting concerns.

We demonstrate our approach for each integration pattern as follows.

4.1. Pipes and filters

The dataflow in the conceptual architecture (see Fig. 1) has a procedural nature. Many mashups leverage the Pipes and Filters integration pattern to create the logic for data combination. The implementation of Pipes and Filters pattern consists of a series of interconnected components. Each component performs a specific function such as filtering unnecessary data, joining records, or forking streams. The connection between components denotes that the output from one component is the input to the other component.

In the context of our mashup architecture, the data retrieval service consists of pipes of composite components that produce data as inputs to the mashup generation part. The data retrieval service devises a number of Filter components that are attached to each data source. For example, domain.com.au provides comprehensive housing prices for cities and suburbs in Australia, however the mashup may only consume data of the City of Sydney covering several suburbs. Therefore a Filter connecting to the data source in URL extracts data just for those suburbs and aggregate them for the area of the City of Sydney.

A Filter component can be further decomposed into several Pipes and Filters as shown in Fig. 2. It contains a data loader to load in the data, connects to a couple of components to trim unnecessary data of various kinds, and finally converts the data to the required format by the mashup component. For example, Australian Bureau of Statistics (ABS)\(^1\)\(^3\) provides comprehensive information of the geographic data of Australian territories, which can be used to produce

\(^{1}\) http://www.abs.gov.au/.
the area boundary on Google Maps. At the moment only a zip file is available for downloading, and the format is in the GIS format. However Google Maps only renders map data in the KML (Keyhole Markup Language) format, which is used to display geographic data in an Earth browser such as Google Earth, Google Maps, and Google Maps for mobile. Hence the format converter is attached at the end of the filter component.

Similar to the Filter component, the Data Gen component also consists of several Pipes and Filters shown in Fig. 3. It transforms the data into a different scale or measure for the presentation purpose. Statistical data such as the median property prices of suburbs are normally presented by different color or density. The Mapping Function in Fig. 3, for example, maps the median property price to color units in order to overlay the price information on the map. This Mapping Function is different from the Format Converter, which concerns with the structure and format of the data. The component of Scaling Function deals with perceptual scaling that can be necessary along the pipes. Since there can be different methods of presenting map overlays, such as symbols marked on location maps in 1D (height), 2D (area) and 3D (volume), the Scaling Function converts the symbol representation between dimensions.

The advantage of using the Pipes and Filters pattern is the simplicity in modeling data flows. Each pipe or filter in the flow is explicitly defined and their integration is driven by the data flow. Data sources are filtered, processed and combined to produce new mashed-up information. In addition, it is easier to maintain the mashups built using this pattern since the mashups execute exactly as they are designed following the flow paths at runtime.

4.2. Data federation

The Data Federation pattern aims to efficiently federate both structured and unstructured data from multiple disparate sources. It helps to mitigate the burden of aggregating, correlating and correcting relevant yet heterogeneous data. This pattern supports data operations against a transient or virtual view. At design time, views of the data sources are declared and the relationships between such views are defined as operations including joins, unions, projections, selections and aggregations. Programs (such as stored procedures) or software tools (such as CodeGlide Fusion and Denodo) can be applied to define operations and produce views.

The relation between the Data Federation pattern and the Pipes and Filters pattern is illustrated in Fig. 1. The Data Federation pattern can be realized by several Pipes and Filters, each providing the Mashup Component with an interface to access the data. The data remains under the control of the source systems and are retrieved on demand for federated access.

Data federation concepts and techniques were extensively researched in the last two decades. Now this pattern has found yet another use in mashup architectures. In the context of mashup, one variation introduced to the Data Federation pattern is data redundancy. The original Data Federation pattern creates virtual views without introducing data redundancy. For mashups, however, data are often reused since the outputs from the Pipes and Filters produce the data in a more suitable structure and format to be processed by mashup components. Therefore, these data can be stored and reused later without going through the data processing Pipes and Filters every time they are needed. Thus introducing some level of redundancy can both simplify the design and improve its responsiveness.

4.3. Model-view-controller pattern

Mashups follow the model-view-controller (MVC) pattern by nature as they render data to present views according to user inputs. In the MVC pattern, the Model has access to data from a storage or a service provider site, and organizes the data in a structure to be used by the Controller. The Controller accepts the browser input, figures out what to query the model for, and decides which view to use and what data to send to the View. The View accepts input from the controller and generates outputs in HTML, XML or JavaScripts to the user’s browser.

There are two ways to construct a mashup architecture following MVC: sever-side mashup (Ort et al., 2011a) and client-side mashup (Ort et al., 2011b). In the deployment of a client-side mashup, the logic of composing map overlays is performed by Javascript at the client browser. This architecture can use Google Mashup Editor, XML databases and Javascript libraries such as jQuery. This simple deployment has drawbacks due to the browser’s limits to handle complex logic and large amounts of data. Alternatively, a server-side mashup allows the integration of data contents at the server side. The MVC pattern is constructed as the server-side mashup in our architecture.

To illustrate the MVC structure, we adopt the sever-side mashup diagram from (Ort et al., 2011a) and annotate the diagram with the key artifacts involved in the MVC pattern. As shown in Fig. 4, a View of the mashup is presented to the end-users through a Web page that contains GUIs. The user interacts with the GUI components and triggers Javascript invocations (through Ajax) to the server side such as loading a page with the mashup embedded. The logic of producing the mashup Web page is guided by the Controller. The Controller at server-side system uses Servlets to dispatch the requests to an appropriate service class that realize the functions of Mashup Generation service in the architecture presented in Fig. 1. Mashup generation and data processing from other sites or services (as Models) are done by the service class as part of the Controller. The response is then sent back to the Servlet, which in turn sends the response back to the client browser. The browser’s JavaScript functions update the Web page and represent the information obtained from the response as the View to the end user.

The server-side mashup is more scalable when making use of the Data Federation pattern and the Pipes and Filters pattern. As discussed above, the Data Federation pattern for mashups introducing data redundancy whereby data read-once/used-many-times can be stored at the server side without accessing the other sites or services each time the data demanded.

5. Composing patterns in RESTful architecture

The conceptual architecture in Fig. 1 consists of two core parts: the Data Retrieval Service (DRS) and the Mashup Generation Service (MGS). In this paper, we focus on the RESTful architecture that allows the MGS to consume the thematic data produced from the DRS using open APIs on the Internet.

The function of DRS is to prepare the data of a certain theme (or so called thematic data) from multiple sources in a format ready to be processed by the MGS. A simple example of the thematic data can be the mean property selling prices for a specific suburb. The MGS takes the inputs of DRS and generates data overlays that can be rendered and presented in a client’s browser. Each part of the service incorporates the integration patterns to structure its key components. The integration of the two services requires open APIs between DRS and MGS be designed to best suit the data source. These open APIs have increasingly been implemented in a RESTful style following the software architecture design paradigm called Representational State Transfer (REST) (Fielding and Taylor, 2000).

The most important element of REST is a resource, which is the source of specific information. Resources can be made available in any format, such as a HTML page, an image, or an annotation file for DNA sequence. A resource is uniquely identified on the Web by a URI (Uniform Resource Identifier). To access a resource, REST uses the HTTP protocol to send URL requests and receive responses through HTTP. Common HTTP methods of GET, PUT, POST and DELETE provide operations on resources by defining respectively list/retrieve, replace/update, create and delete operations for resources.

To demonstrate the usage of a RESTful data retrieval service, an example of The Suburban Trends mashup is introduced as follows. The Suburban Trends mashup combines publicly available online resources from the Australia Bureau of Statistics, Australian Institute of Criminology, the NSW Bureau of Crime Statistics and Research, and Google. One of Suburban Trends’ REST APIs returns comprehensive information associated with the State Suburb Code (SSC) including the suburb’s centre coordinates, the suburb’s extent coordinates (the two coordinates that form the suburb’s bounding box), the SSC for the ten closest suburbs as shown in Fig. 1. These coordinates are the primary inputs for the MGS to produce the skeleton of the territory shape (more details will be presented in the next section). The sample HTTP request and response using REST APIs are illustrated in Fig. 5.

Other REST APIs from Suburban Trends provide thematic data as valuable inputs to enrich the mashup for the Property Valuation Service, such as socio-economic rankings, safety ratings and yearly crime statistics for New South Wales state since 1995. The RESTful architecture enable Data-as-a-Service for enterprise mashups by providing easy programming access to applications that may produce diverse, application specific mashup views of the same set of data.

If the thematic data is produced through existing enterprise legacy systems, the DRS may have already been implemented the interface exploiting traditional enterprise integration technologies such as Java Message Service (JMS) or Simple Object Access Protocol (SOAP) Web services. In this case, the RESTful architecture is still applicable with extra integration patterns involved such as Message Endpoint (Hohpe and Woolf, 2003). With the help of the Message Bus integration pattern (Hohpe and Woolf, 2003), the interface to access the thematic data can be wrapped by a message endpoint. A REST request/response can be delivered to/from the message endpoints by a message channel such as JMS or HTTP. The technical details are out of the scope this paper. The best practises to utilize these patterns have already been well studied and documented such as enterprise integration patterns in (Hohpe and Woolf, 2003) and (Chatterjee and Webber, 2003).

The simple architecture style of REST brings a number of benefits to the design of highly flexible, low-overhead REST APIs to access collections of thematic data:

1. Support for unified development of Web-browser based interfaces to data source. From the development point of view, REST APIs can be invoked using different programming languages such as C++, Python, Java, C# and so on. Thus, new features that further manipulate the data obtained through REST APIs can be developed in a compatible manner to existing applications.
2. Support for encapsulation of data sources. Access to data is well defined in REST APIs, which can be designed to produce different viewpoints of the data in the responses.
3. Lightweight. Compared to SOAP-based APIs in the complex SOAP message format, REST APIs are lightweight in terms of both communication and computing overhead. SOAP is well recognized for its poor performance when the SOAP messages contain large payloads (Davis and Parashar, 2002).
4. Support for a layered architecture. REST allows the introduction of other computing layers between the Web browser and the server hosting the data to help with more sophisticated pro-

---

6.1. Inherent overlay.

6. Architecture deployment and evaluation

6.1. Web-based property valuation

The architecture integration patterns for composing mashups are applied to the property valuation service process discussed in Section 2. The benefits of integrability and reusability are inherent from these patterns and the REST style. By following a scenario-based architecture evaluation approach, we have selected typical integration and reuse scenarios as the illustrating examples throughout the section. We describe our experience of using the new approach to developing the mashup for the Web-based PVS system.

6.2. Web-based valuation form

The mashup of the PVS has three main GUI panels as shown in Fig. 6. The left panel contains the property list that the valuer has inspected or will inspect. The right panel is a tabbed panel that contains the valuation form. This panel also contains a tab to show historical prices trend of the area that the property belongs to. The bottom panel is the map panel, which displays overlays on Google Maps such as housing prices and traffic status. The map highlights the areas that a valuer prefers to work on together with the performance ranking and the booking information. The coloring of the map is based on the median price level of the properties in the area.

The Web-based evaluation form is consistent with the paper-based valuation form. The form can be converted to the PDF format and sent to the valuation firm on the fly, which saves paper and the cost of postage. This Web-based valuation form has several other advantages. First, it saves time spent by the administrator on spotting the best available valuer in the target area. The administrator only needs to input the suburbs of the target areas and the map will be automatically produced with a marker to list available valuers. Valuers are displayed only in their working areas, together with their ranks and bookings information. Second, the valuer can leverage the historical prices of the valuation property easily. Moreover, different types of information can be chosen to highlight on Google Maps. For example, when a valuer clicks on the traffic entry in the valuation form, the map panel displays a map with traffic overlay.

6.3. Mashup development

The core of this Web-based PVS is the embedded mashup display. The mashup development follows the architecture patterns discussed in Section 4. We use the scenario of the valuer distribution in the area of City of Sydney (including several suburbs) as an example to demonstrate this architecture approach.

The key components to produce the mashup is illustrated in Fig. 7. There are four sources of data: the territory geographic data for producing shape boundary on Google Maps, the housing prices and traffic data from the public domain, and the valuer information from the corporate information systems. Thus four pipelines form the Data Federation pattern and produce the inputs to the mashup component. The overall mashup data flow follows the Pipes and Filters pattern. The interactions amongst the client browser, the Web server and the mashup component employ the MVC pattern and use a server-side mashup deployment.

6.4. Preparing data for mashup

Preparing data for mashup involves the tasks of accessing the shape files and transforming them to the KML format. KML is an XML format and uses a tag-based structure with nested elements and attributes to describe a shape in coordinates. The KML file does not contain any thematic data. Instead, it provides the base shape that thematic data can be presented as overlays on top of it. The list below shows a partial KML file describing a shape.

```
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Placemark id="1">
    ...
    <LineString>
      <coordinates>
        151.31515, 33.91634
      </coordinates>
    </LineString>
  </Placemark>
</kml>
```
In our case, the KML shape files are essential to produce overlaid information on location maps such as Google Maps. Google Maps APIs provide a way to create a customized overlay if a well-structured KML file is provided on a publically accessible server. The KML shape files are produced by the data retrieval service of the architecture, which consists of a number of data loaders devised to retrieve data from various data sources as shown in Fig. 1. The original shape files to produce the KML files can be obtained in two ways as discussed in Section 4: (1) downloading and processing the raw shape files from the Australian Bureau of Statistics (ABS), or (2) invoking REST APIs of Internet services that provide suburb coordinates. The former requires components that may use Geographical
Information Software (GIS) such as MapWindow to process the shape files as illustrated in Fig. 7. The latter can be implemented by toolkits that support client side HTTP standards. The sample code using HTTPClient to invoke the REST API is very straightforward as follows:

```java
   HttpClient httpClient = new HttpClient();
   httpClient.getHttpHost().setPort(80); // or HTTPS
```

To demonstrate the techniques of producing data for KML-based map overlays, we present the details of processing original shape files downloaded from ABS in a geographical spatial vector data format. The shape file specifies the coordinates of all subrub boundaries in Australia, including the fields for each suburb: (1) Name of the suburb; (2) State that the suburb belongs to; and (3) Coordinate, the list of longitude/latitude pair that represents the boundary of the suburb. The coordinate data needs to be aggregated for the relevant suburbs and the remaining data can be filtered out. To do this, the shape file is loaded into a Shape Filter. We use a COTS GIS software MapWindow that supports some editing functions on a shape file. For example, MapWindow has built-in SQL-parsing functionality so that the filtering task can be accomplished by an SQL script similar to Select.. from.. Where State = ‘New South Wales’. Then the data is converted to KML coordinates.

The GIS2KML Converter shown at the top of Fig. 7 leverages a third party MapWindow Plug-in called Shp2KML to convert the shape files to the KML files. The effect of the resulting KML file is depicted in Fig. 8, which shows the base shape of the suburb “City of Sydney”. The next step is to input the base KML file to the mashup component to merge overlays with map coordinates.

One pipeline deals with extracting valuer working area distributions at the bottom of Fig. 7. All completed historical valuations are stored in a central corporate database and the existing VMS is connected to access this database. One filter component contains SQL stored procedures to extract the data of interest. Another component Criteria Filter is for valuer’s performance ranking. The ranking originates from the PVS manager with regard to individual valuation quality and quantity in the past. Such statistical results can be directly generated from an independent human resource system in a XML format. A XML Parser is used as the filter, which reads the XML file, processes the data and then matches the records of historical valuation results to discover valuers working in a certain area with a satisfactory record.

6.5. Producing map overlays

Eventually the four pipelines converge in the mashup component to produce a single KML file with map overlays that the Google Maps server can render. The overlay is visualized by colors. Different colors or color density can represent scales of the data. Hence the component of Color Mapping Function maps the data from one measure (price) to another (color). The mapping between the color and the housing prices is customized by the scale parameter of interval, shown in Eq. (1).

$$\text{Color unit} = \frac{\text{Max(housing price)} - \text{Min(housing price)}}{\text{Interval}}$$ (1)

The next step is attaching the color unit to each element (suburb) in the base KML file. It is straightforward to find the corresponding color level using the color unit, given the starting color is determined. Eventually, the mashup component composes a final KML file with map overlays built-in. The sample code of the final KML file is illustrated below.

```xml
   <name>CityofSydney</name>
   <placemark id="0">
      <name>Alexandria</name>
   ...
   <polygon>
   ...
   </polygoneBoundaries>
   <LinearRing>
   <coordinates>151.20134016, -33.9034390011261, 0 151.20181404...
   ...
   </polygon>
   <style id="C1">
   <LineStyle>
   <color>CC000080</color>
   <width>1</width>
   </LineStyle>
   <PolyStyle>
   <color>CCAA3198</color>
   <fill>1</fill>
   </PolyStyle>
   </style>
```

The resulting KML files are transmitted to the Google Maps server from an accessible site for rendering. As the files are normally of hundreds of megabytes in our case, optimizing the transmitting overhead is essential to make the mashup application responsive. Our solution is compressing the KML file using ZIP that helps to produce the size of the compressed file approximately a quarter of the original size. For example, the KML file containing all suburbs in New South Wales is 110.4 MB whereas the compressed KMZ file is 30.9 MB.

Using the Google Web Toolkit (GWT) based library called.gwt-Ext7, the compressed KMZ file can be rendered following the template below. The effect of the final display is illustrated in Figs. 6 and 9.

```javascript
MapPanel.addOverlay("http://www.kmz-serverhost.com/CityOfSydney_House Value.kmz");
```

---

7 http://gwt-ext.com/.
6.6. Producing different views

Three pipelines at the top of Fig. 7 deliver different views in the mashup following the MVC pattern. As shown in Fig. 9, the mashup enables users to select the menu item of interest, and the mashup switches between views. Since the communication is using Ajax asynchronous calls (see Fig. 10 for details), the end-user is not blocked for one operation and can continue the interaction with other functions of the Web page before the response returns. From our experiment, the user will experience some delay (no longer than loading a normal Google Maps page) the first time the map and overlays are loaded, and later refreshing of the content or switching the map between views is fairly prompt.

7. Lessons learnt and conclusion

In this paper, we propose an architecture solution that demonstrates new usage of architecture integration patterns in mashup architecture design and development. This architecture incorporates the state-of-the-art RESTful style for integrating key components that access and transform data to produce mashup overlays. The design of this architecture applies integration patterns and thus inherits their reusability when integrating mashup components and services. To realize this architecture, we also demonstrate the techniques implementing Google map-based mashups for the property valuation business service. Through our experience, we observe that component-based development helps to achieve good encapsulation of the operations required by mashups, and thus enables architecture patterns to structure key components in mashups. We summarize three aspects of lessons learnt through our experience.

7.1. Mashup deployment architecture

The MVC pattern can be deployed as either a client-side mashup or a server-side mashup (see discussion in Section 4.3). However it is not straightforward to make the design decision about which deployment architecture is suitable for a specific mashup scenario. We initially adopted the client-side mashup architecture, attracted by its simplicity and comprehensive tool support (Google Mashup Editor). However, as the mashup gets complicated with new features added, different operations needs to be performed on different data sets. Eventually the client-side mashup was no longer suitable and we migrated to a server-side mashup.
7.2. Using RESTful services

The affinity of REST and Mashup enables use and reuse of data sources for mashup development. Increasingly mashup applications publish processed thematic data for map overlays using REST APIs. Consuming these services using REST APIs and lightweight HTTP standards is widely supported by different languages and software tools. Mashup applications can benefit from reusable data already processed and so reduce the costs of development.

7.3. Using other integration patterns

The three integration patterns presented are not exclusive in our architecture solution. Instead, they form the base structure to enable interactions with other integration patterns augmented for application specific mashup needs. For instance, the Data Federation pattern may contain a sub-scenario of locating a specific data service from multiple providers. This scenario may employ the JEE Service Locator pattern or Web Services UDDI service discovery pattern. In our envision, as the scale of the mashup application expands, augmenting the mashup architecture with more integration patterns is necessary. The design needs to focus on reducing the crosscutting concerns from multiple patterns.

7.4. Optimization

The aim of optimization is to enhance the user experience when interacting with the Web-based PVS. A mashup is user-centric, and thus the user experience is a key criterion for the design and implementation of a mashup. As the mashup component communicates with data intensively, optimizing data processing and transmission is essential to achieve efficiency and responsiveness for the benefit of the end users. Since the size of the KML file directly determines the waiting time that the end-user will experience, we focus on optimization techniques that can reduce the size of KML file and the number of bytes that have to be transmitted over the network. In addition to the general techniques to improve the structure of the KML file, we further compressed the KML file using ZIP\(^8\). Our experience demonstrated that zipped XML is normally only 1/4 of the original size. For example, the KML file containing all suburbs in New South Wales is 110.4 MB whereas the compressed KMZ file is 30.9 MB.

Another technique we employed is to reduce the number of decimal places for coordinates in the shape file. The more decimal places for coordinates, the more accurate the shape is. A coordinate with 6 decimal places has an accuracy of less than a foot. The accuracy level is sufficient to produce the base KML file in the PVS. This technique also reduces the size of the KML file. Finally, a big KML file is split into smaller files to transmit over the network one by one. All these techniques in combination can reduce the KML file size up to 80%.

Our conceptual architecture still has space to address advanced issues of mashup for enterprise applications. It remains as future work to extend this architecture to manage security and accountability\(^9\) (Zou and Pavlovski, 2007) when data are retrieved from third party services, data sources or applications.

Acknowledgements

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

References


Yan Liu is a senior research scientist at Pacific Northwest National Laboratory. Her current research involves software architecture, scientific workflow, model driven development, software performance engineering, and adaptive systems. She has extensive experience leading projects in domains ranging from enterprise applications, embedded systems, to scientific computing systems. She has published over fifty international journal and conference papers in these areas since 2004. Yan received her PhD in Computer Science from University of Sydney, Australia in 2004. Until December 2009, she was a senior researcher at NICTA (National ICT Australia). In 2008, she was nominated for NICTA Richard Newton Excellence in Research Award recognizing her contribution to NICTA’s vision and values.

---

A technical paper (Sakr, in press) provides detailed evaluation of a set of widely used XML compression tools.
Xin Liang is a software developer in Macquarie Group now. He has worked on a wide range of IT projects includes Content Management systems (CMS), Customer Relationship systems (CRM), Deal management systems etc. He likes programming on Microsoft technology stack and has special interest on client-side and server-side javascript. He was a visiting student in NICTA in 2008. He received his Bachelor of Computer Engineering (Honours Class 1) degree from University of New South Wales, Australia in 2009. In 2010, he received Macquarie Group IT Awards.

Lingzhi Xu is currently working in a commercial bank in Shanghai China. She has wide range expertise in business process analysis and Web-based development. She was working as a research engineer in NICTA from 2005 to 2009. In NICTA, she developed various software prototypes for research projects on software architecture knowledge management, software cost estimation, and business process analysis. She received Master degree in computer science from University of New South Wales, Australia in 2005.

Mark Staples is a Research Group Manager at NICTA, the leader of the Fraunhofer Project Centre in Transport and Logistics at NICTA, and has a joint academic position in the School of Computer Science and Engineering at the University of New South Wales. His research area is software engineering, especially in process improvement, architecture, and configuration management. He has published more than thirty scientific papers in journals and conferences. Prior to joining NICTA, he worked in industry developing safety-critical embedded systems and enterprise software infrastructure. He holds degrees in computer science and cognitive science from the University of Queensland, and a PhD in Computer Science from the University of Cambridge.

Liming Zhu is a senior researcher at NICTA (National ICT Australia). He holds joint academic positions and teaches software architecture courses at both University of New South Wales (UNSW) and The University of Sydney. His research interests include software architecture, service engineering and system development methodologies. He has published more than sixty scientific papers. He contributes to software architecture related standards and technologies at both ISO and several industry consortiums. He formerly worked in several technology lead positions in software industry before completing a PhD degree in Software Engineering at UNSW.