Aim4GDI: Facilitating the Synthesis of GDI Resources through Mapping and Superimpositions of Metadata Summaries

Trias Aditya · Menno-Jan Kraak

Abstract In the world of geospatial data infrastructures, geoportals are developed to facilitate access and use of geospatial resources, including data. The design and implementation of effective and efficient geoportals are becoming crucial. Providing possibilities for users to organize and integrate available resources can arguably enhance the process of data discovery and spatial analysis required, and hence, is the focus of this paper. We argue that the synthesis of summaries of distributed datasets, map and feature services through the geoportal can present a coherent view to support data discovery. For this purpose, the atlas metaphor is used as an indexing server and integration interface of summaries. Thematic maps and a storyteller view are accessible through the atlas metaphor. They can be used to provide a supporting context for data discovery and access purposes.

Keywords geospatial data infrastructures · metadata · data discovery · visualization · integration · atlas metaphor · RDF

1 Introduction

The growing use of web mapping applications and adoptions of specifications from Open Geospatial Consortium (OGC) are the two important factors that keep the pervasiveness of the Geospatial Data Infrastructure (GDI) across the local, national, and regional levels furthering. The GDI is intended to facilitate both users and providers' needs. Through the GDI, providers can publish two different types of data: offline datasets and online geospatial web content such as Web Feature Services (WFS), and Web Map Services (WMS). These GDI resources are meant to be discoverable and accessible to users. A portal or gateway facilitating publication of and access to those GDI resources is of high

T. Aditya (✉) · M.-J. Kraak
Department of Geo-information Processing, International Institute for Geo-Information Science and Earth Observation (ITC), P. O. Box 6, Enschede 7500 AA, The Netherlands
e-mail: aditya@itc.nl

importance [1]. As mentioned in an OGC paper [2], a geospatial portal is a (web) interface to a collection of online geospatial information resources, including offline data sets, online map, and feature services. This includes functionalities to provide clients of viewer, discovery, publisher, gazetteer, data extraction-manipulation, and style management.

The implementations of OGC specifications are a key premise for facilitating data and information sharing via geoportals. However, the design and implementation of usable geoportals is challenging. Considering the potential uses and richness in content of a GDI, a geoportal that enable users (novice and expert users of various domain applications) to profit from their access to the GDI is needed. For the success of a GDI, providing effective and efficient methods of access are urgently required [3]. The current practices of geoportals seem to be data oriented instead of demand-driven [4] and often provide limited support to enable users to effectively and efficiently assess the fitness for use of the data required. For example, the visualization and integration of the metadata, map, and feature services required in support of information discovery and access for decision support could be improved. A metaphor for an improved access and use of the GDI is required.

As a solution to the above, this paper proposes Aim4GDI, a web-based atlas interface as a metaphor to access the GDI.

2 Searching and browsing the GDI resources

Beyond the motivation previously mentioned, assuming the interoperability is in place, GDI has potentials not only as a means for information access, but also for offering decision supports and enabling geocollaboration [5]. In order to achieve these potentials, some interoperability impediments, like incomplete sets of mapping and chaining specifications and their tools compliance (see e.g. [6]), as well as usability drawbacks [7], [8], have to be overcome. Our discussion will focus on the improvement of the usability aspects of geoportal to support information discovery and access for decision-support.

Using geoportals, users commonly have possibilities to search metadata or browse directories of metadata, to load and cascade WMS, and to share data. From this perspective, the U.S. Geospatial One Stop—GOS [9], GeoConnections [10], Inspire Geoportal [11], and an open source catalog server-GeoNetwork.org [12] are some interesting examples of the active OGC conformant geoportals. To access the information and services through the search interface of the geoportals, users have to provide at least one search term related to location, attribute or time of the required data or map services. This approach is appropriate to support tightly defined discovery tasks. To exemplify this kind of task, let's look at Danny, a GIS technician who needs a specific dataset for an agriculture map-updating project. He could start the search process by submitting or selecting search terms related to his topic, scale, and area of interest to a geoportal. A common solution that Danny gets using today's geoportals is the display of search results in forms of a set of abstracts and thumbnails with links to view data and to review full metadata descriptions.

Since a GDI might cover assorted topics of related geospatial data and maps, browsing through topic directories is a common approach for completing a loosely defined task. This task usually specifies no detailed requirements at the start of the interaction with the geoportal. In this task, the fitness of use is not simply depending on matching values of certain elements in metadata. Consider Lisa, a transportation engineer who requires datasets for her work designing a traffic survey in support of a road extension project. To get appropriate search results, she might need more than just simply defining format and area of interest. Through browsing activities, a user such as Lisa can page through the links offered

via the directory of topics or providers and find some data that can be used for her project. Currently, not all geoportals support browsing activities.

Regarding support for searching and browsing activities through geoportals, the existing set of displays does not facilitate users' needs to quickly compare and sort metadata out. Users must drill down each item to assess the matching of the metadata elements to their queries. Additionally, they also do not permit users to assess the trend and pattern of the data that are available for a specific topic of interest. Furthermore, current geoportals provide no contextual support to help users to browse (either by topic or by providers) the data and services available. Thus, in such setting, users only have possibilities to assess the use suitability based upon the utility or property of the data/ services, with for instance no background layers (e.g. thematic maps) are offered to assist the data suitability assessment.

In addition to the drawbacks mentioned above, the integration of the GDI resources (WMS, WFS, and metadata characteristics) is not solved thoroughly. Possibilities to combine different resources into one single interface would broaden the use of the data published. The GOS [9] provides a good example of the benefits of interoperability of GDI resources where publicly available WMS can be cascaded and displayed through the viewer. However, using the GOS and other geoportals, the possibilities to perform visual thinking by means of representing metadata and of juxtaposing the GDI resources (not only WMS) are still limited. By visual thinking, we envisage opportunities for users to study and compare for example, the density, pattern, and distribution of available datasets for a single or multiple topics in support of a "fitness for use" assessment. These prospects suggest the rise of the GDI to reach broader users and to be used more to enable information access and decision-making.

3 The atlas as an indexing and integration service

In our research project, we employ the concept of atlas as a metaphor. The rationale for this decision builds upon the familiarity and understandability of how-to-use atlas by (GDI) users. We specifically consider the national atlas since it is part of the national information assets and presumably has been introduced to and used by users during their school education or in practice. The general approach applied in web and book atlases is to provide an indexing mechanism on what maps and relevant documentation that are available for a list of specific topics. Further it brings a relatively uniform representation of maps and information to users. This scheme of visualization allows users to build comparisons and synthesis on a specific theme selected. To enhance the user's understanding of the context of the mapped themes, a storyteller view is developed. Through a storyteller view, supporting information related to thematic maps and GDI resources are accessible. Through such a map storyteller and GDI storyteller, users have access to relevant documents, graphics, images, and animation as well as related online content. They are designed to facilitate certain rhetoric of communication (see e.g. [13]) in support of browsing activities.

The GDI storyteller resources are to be created and maintained by the atlas editor. The data provider organizations or any registered users can also contribute to the content of the storyteller system by submitting content or graphics related to a specific thematic map. As an example, a user could submit a documentation regarding experiences of the use of a specific map service about road features to be displayed through the storyteller view under topic "Transportation" and listed under the map "road networks."

In some national GDIs, national web atlases have been positioned as one of nodes of the countries geospatial infrastructure, e.g., [14] and [15]. To use the atlas as a geoportal, we propose two vital functionalities to be built: index and superimposition schemes. The index

scheme refers to organization and management of GDI resources (such as datasets and WFS published) into summaries of metadata. These summaries are displayed as textual and graphical representations on top of the interface to support data and information access. The superimposition scheme deals with methods to allow users to overlay metadata items or subset of required feature services on top of thematic map to enable visual thinking during the searching and browsing process. The superimposition scheme can be seen as cascading GDI resources, thus here, the cascaded "graphics" are not limited only to WMS but also including metadata and WFS visualization on top thematic maps of the atlas.

The following will look closer at the metadata management of summaries and the atlas directory, query services, and mapping functionality of the Aim4GDI.

4 Metadata management

4.1 Metadata summaries

The GDI resources that are commonly published in the framework of the GDI are: proprietary datasets, WMS, and WFS. In the current practice of geospatial data management, these three types of data are documented using related metadata standards or specifications in the form of XML documents. Proprietary vector and raster datasets are documented for example using the International Standard Organization (ISO) 19115. Meanwhile, map services and feature services are required to implement OGC GetCapabilities interface to enable client (human-users or machine) to get a descriptive feedback in the form of XML about the service to be accessed.

In this work, we extract those types of geospatial metadata as summaries after the metadata are published. Here metadata are to be published by data providers into the atlas system via a registration web page. A metadata summary is a concise description of GDI resources. This summary consists of XML elements outlining title, abstract, usage, access, area coverage, topic, and temporal information of GDI resources. Introducing the metadata summary between metadata standards /specifications and end-users may produce redundancy of datasets' descriptions. However, this must be done, as our focus is to improve the search, representation and selection of data and service via the Aim4GDI. Given the nature that providers across the GDI implement various standards, to present uniform and concise representations of the GDI resources to users and to link them to the maps in the atlas, a special type of metadata is needed. This dedicated metadata repository is intended to support a strategy for "overview first, zoom and filter, then details on demand" [16] in the context of geospatial data discovery.

The metadata summary is expressed as Resources Description Framework (RDF) and serialized as XML format (named as RDF/XML [17]). The rationale to use the RDF/XML encoding was motivated by two considerations. First, RDF is a W3C specification that was designed to provide a standard for metadata of web resources[18], [19]. The data being described is seen as directed label graph of the triples (subject–predicate–object). As such graph model, it provides advantages in terms of processing and querying of datasets [20]. In the Aim4GDI's case, it gives straightforwardness to reconstruct the relationships between topics and corresponding thematic maps with their supporting information and related GDI resources for integration and information retrieval purposes. Second, the atlas structure (or directory of the atlas) is constituted of selected topics, in which for each topic, it can include

some relevant thematic maps. Having such structure, a map in the atlas can be seen as an indexer of thematically related GDI resources as well as the atlas resources (including relevant charts, graphics, images, and textual descriptions to the map). In this graph-like construct, the directory data is better represented using RDF/XML than using XML (for tree-like domain application) or relational databases (for table-like domain applications) [21].

In our experimental environment, providers are required to submit the web address of the XML metadata of proprietary datasets and the web address of GetCapabilities as well as GetFeature and GetMap accesses of WFS and WMS. Further, an important requirement at the registration step is that the provider is required to specify the linking of the metadata or WMS/WFS to a specific thematic map in the atlas. This linkage approach is the key to organize the GDI resources into the atlas system (indexed into a particular map within a specific topic in the atlas directory).

As mentioned earlier metadata will be processed as summaries after the providers register their data through a simple registration web page. To handle the datasets' registration, a simple Java servlet is used to process the web address of the metadata submitted. Subsequently, metadata summary items are created out of it with help of an Extensible Stylesheet Language Transformations (XSLT) template *"xml2summary."* The stylesheet template is applied against the XML metadata to generate geographic, topic, temporal, linkage to a thematic map, usage, and accessibility elements. The generated RDF/XML summary used the Dublin Core namespaces [22] (i.e., dc-elements and dc-terms), topic category of geospatial metadata standards (ISO 19115), and proprietary namespaces suited for this application (metadata summary). The resulted elements (excerpted) are shown in the following:

```
<ms:item rdf:about="http://overijssel.nl/gis/transport/dataxml/buszone.xml">
<!– Resource general Information –>
  <dc:title>Overijssel bus zones</dc:title>
  <dc:description>....abstract information....</dc:description>
  <dc:created>1999-08-01</dc:created>
  <dc:creator>Province Overijssel</dc:creator>
  <dc:publisher>Opleiding van GIS Overijssel</dc:publisher>
<!– Geographical coverage–>
  <ms:geocov rdf:parseType="Resource">
    <dcterms:spatial>
        <dcterms:Box>
            <dcterms:northlimit>539914</dcterms:northlimit>
            <dcterms:eastlimit>267933</dcterms:eastlimit>
            <dcterms:southlimit>459816</dcterms:southlimit>
            <dcterms:westlimit>185302</dcterms:westlimit>
        </dcterms:Box>
    </dcterms:spatial>
</ms:geocov>
<!– Topical coverage–>
<ms:topicov rdf:parseType="Resource">
    <dc:subject>Transportation</dc:subject>
    <iso19115:topicCategory>Transportation</iso19115:topicCategor>
</ms:topicov>
</ms:item>
```

Summaries of online feature and map services are aggregated in the same way. For WFS and WMS to be included, it is required within the registration to specify the title, abstract, description (in case its GetCapabilities has no sufficient information about title, abstract, and description) and the designated association to a specific map. In addition, for a WFS submission, it is required to specify the access node of DescribeFeatureType, and GetFeature requests. Meanwhile for a WMS to be included, the provider should specify the GetMap access node. Following the submission, the web address of each request is accessed. Subsequently, stylesheet templates "wfs2summary" and "wms2summary" will summarize the items based on the responses outputted. For a WFS type of resource for instance, the resulted summary will specify the title, abstract, geographic bounding box, topic, linkage to a thematic map, and the web access of the GetFeature request as an RDF data.

In the prototype, we make use of metadata of 70 datasets offered by different departments within the Province of Overijssel, Netherlands. Regarding map and feature services' experiments, the GeoServer and Mapserver, which both run in the local network, are used as the map and feature servers. For testing our concept, we only involve datasets and feature services that relate to two topics within the GDI: Transportation and Agriculture. Figure 1 shows the graph representation of a summary.

Metadata summaries are encoded as graph pattern of "subject-predicate-object" triples. Regarding summaries of GDI resources, they are differentiated by encoding them either as <ms:dataitem/> or <ms:WMSitem> or <ms:WFSitem> elements. As exemplified in the Fig. 1, the data of "regional bus zone" is a type of dataset item, and contained into the map of public transport.
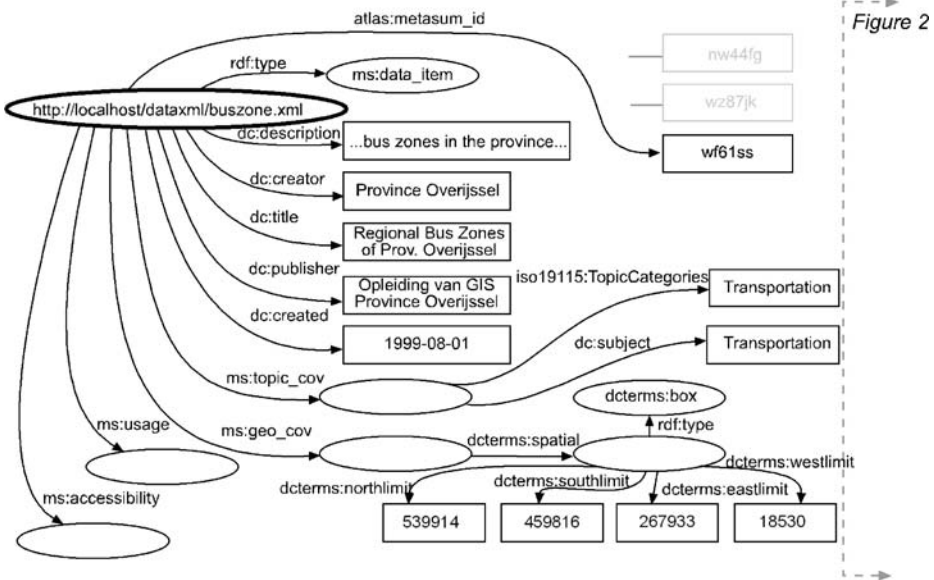

Fig. 1 The RDF graphs of metadata summaries

4.2 Atlas information structure (directory)

The atlas directory is developed to organize the content of the atlas. As mentioned before, within the atlas, users are expected to get access to the collection of maps and their supporting information as well as GDI resources. The directory indeed is designated to provide appropriate "yellow pages" for users to access and retrieve information required.

The topics listed in the Aim4GDI are based upon the TopicCategories element of ISO 19115. Each topic enlists related thematic maps. Within the directory file, each thematic map specifies the title, abstract, and the web address for display. In practice, the datasets of these maps are collected from various institutions and then compiled by the atlas editor. The map functions as the indexer of relevant atlas and GDI resources. This structure is intended to support browsing GDI resources through thematic map in order to provide contextual information for data discovery and decision-making [8]. This structure is also encoded using RDF/XML (Fig. 2 depicts the RDF graphs of the atlas directory).

5 Query services

The Jena web toolkit [23] is used in the system to provide query services to RDF data of the metadata summaries and the directory. Specifically, we use the ARQ query engine [24] to perform queries using SPARQL, the proposed standard for RDF query language [25]. SPARQL is built on the *triple pattern components*: a subject, predicate and object that construct a graph pattern. Any of these components or the entire graph pattern can be replaced by a variable. The query attempts to match the triples of the graph pattern to the model (i.e., RDF data). The query solution prescribes the matching of the variables over the model's nodes. For this, each query should use PREFIX keyword for handling the namespace issue, includes SELECT clause to specify what the query should return, optionally specifies FROM that provides the URI of the dataset to use, and defines the
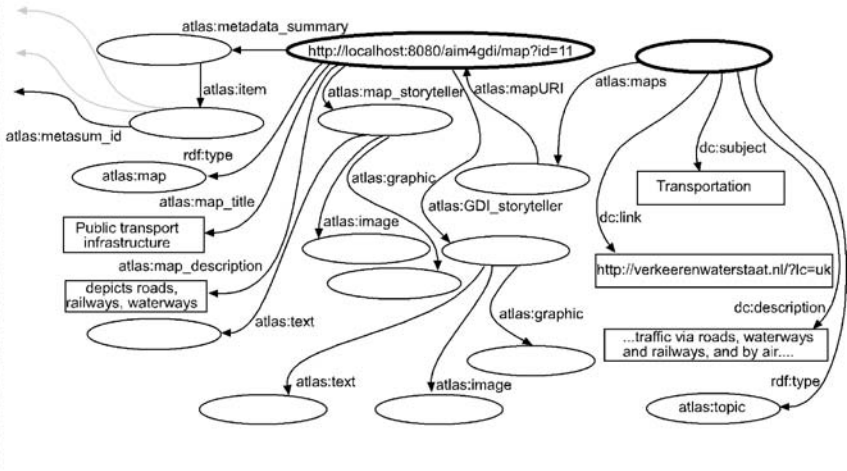

Fig. 2 The RDF graphs of the atlas directory

WHERE clause that defines a series of triple patterns. SPARQL also allows the application to matching multiple graph patterns.

To exemplify the SPARQL usage for querying metadata summaries and the directory, two tasks exemplified in Section 2 are recalled: a GIS technician, Danny, requires a dataset for his agriculture map-updating project and a transportation engineer, Lisa, who requires datasets for planning a traffic survey.

5.1 Searching resources through summaries

For Danny's case, the more explicit requirements he could have are: "provide me a dataset about vegetation types published by the GIS section of the Overijssel Province (GIS_OV), and offered as a shapefile dataset." This request, hypothetically, is best queried via searching interface. For this purpose, the search parameters that Danny defined through form widgets (textbox, drop down menus, and combo boxes) are sent to the server and handled by a query program once Danny presses the "search" button. The SELECT query syntax for this request in SPARQL is expressed as follows:

> **PREFIX** dcterms: <http://purl.org/dc/terms/>
> **PREFIX** ms: <http://metasum/elements/1.0/>
> **PREFIX** atlas: <http://kartoweb.itc.nl/atlas/elements#>
> **PREFIX** xsd: <http://www.w3.org/2001/XMLSchema>
> **PREFIX** dc: <http://purl.org/dc/elements/1.1/>
> **PREFIX** rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
> **SELECT** ?westlimit ?northlimit ?eastlimit ?southlimit ?item ?resource ?publisher ?title ?abstract ?topic ?format ?scale
> **FROM** <"+url_ms+">
> **WHERE** {
> ?item ?x [?y [dcterms:westlimit ?westlimit],[dcterms:northlimit ?northlimit],[dcterms:eastlimit ?eastlimit], [dcterms:southlimit ?southlimit]].
> **FILTER** (?westlimit > +qXNW+ && ?southlimit > +qYSE+ && ?eastlimit < +qXSE + && ?northlimit < +qYNW+).
> ?item rdf:type ?resource; dc:publisher ?publisher ; atlas:metasum_id ?meta_id ; dc:title ?title;
> ?item dc:description ?abstract ; ms:topicov [dc:subject ?topic] .
> ?item ms:usage [ms:representation ?datatype] , [ms:format ?format] , [ms:scale ?scale].
> **FILTER** (regex(?title, \+qkeyword+\) && regex(?abstract, \+qkeyword+\)|| regex(?topic, \+qtopic+\) && regex(?publisher, \+qpublisher+).};

PREFIX keywords are used to specify the namespaces used in metadata summaries. The query was made against metadata summaries file that reside in the server and filtered according to parameters that Danny specifies (using the FILTER keyword).

The "*qXNW*," "*qYNW*," "*qXSE*," and "*qYSE*" in the above lines are the corresponding westlimit, northlimit, eastlimit, and southlimit of the "Overijssel" dataset. These position values are gained with the help of "PlaceBoxes" data (an XML file). This file can be seen as a simple gazetteer data and used to provide footprints of the geographic place names of the country. The incoming queries with having place name or administrative unit parameters are first to be matched with this PlaceBoxes file to gain the corresponding westlimit, northlimit, eastlimit, and southlimit of the area of interest. Other search terms that Danny might have include the topic, type of data, the name of publisher, format and scale of the vegetation data that Danny is interested in. Those search terms are used as parameters

to filter the possible results. Since ARQ processor support all kinds of data types filtering (e.g., xsd:date, xsd:integer, xsd:string), hence the application provides flexibilities to Danny and other users to define many questions in relation to the data they look for.

5.2 Browsing resources through the directory

The task that Lisa needs to accomplish is to produce a spatial analysis to support a traffic survey campaign that aims to measure traffic noise level as well as noise perception among population in the southern part of Overijssel province. Using the Aim4GDI interface, she could browse relevant maps within the "environment" topic to check whether relevant information that might be beneficial to her task can be found. With browsing maps through the directory, she issues a query to populate the map container. Through SPARQL, a request like "provide me a list of summaries of GDI resources related to the map of environmental quality" is built as follows:

> **SELECT** ?item ?title
> **FROM** <"+*url_ms*+">
> **FROM NAMED** <"+*dir*+">
> **WHERE** {
> **GRAPH** <"+*dir*+"> {
> ?b ?x [atlas:mapURI ?map].
> ?map rdf:type atlas:map .
> ?map atlas:map_title **?maptitle**; atlas:metadata_summary ?gr.
> ?gr atlas:item [atlas:metasum_id ?item_id]
> **FILTER** (regex (**?maptitle**, \+*qmaptitle*+\)).
> }.
> ?item atlas:metasum_id ?item_id .
> ?item dc:title ?title .
> ?item dc:description ?abstract.
> };

These query lines make use of two graphs to retrieve the requested result. In essence, the graph of metadata summary (represented with "*url_ms*") is matched against the graph of directory ("*dir*") in which the matching is limited into the pattern that meets string constraint ("*qmaptitle*") that reflected the map title selected.

The use of the directory linked to summaries offers possibilities to broaden the search scope (involving atlas resources and GDI resources). As an example, consider the following request as Lisa expands her query: "show me summaries categorized as feature services dealing with the environment topic and offered by Ministry of Housing & Environment (i.e., "VROM") and those summaries are within the maps that have documentation and images with keyword 'sound pollution'." This request can be expressed as query string as follows:

> **SELECT** ?item ?title
> **FROM** <"+url_ms+">
> **FROM NAMED** <"+dir+">
> **WHERE** {
> **GRAPH** <"+dir+"> {
> ?b ?x [atlas:mapURI ?map].
> ?map rdf:type atlas:map.

?map atlas:map_title **?maptitle**.
?map atlas:map_storyteller ?gr.
?gr atlas:text [atlas:textTitle **?texttitle**], [atlas:textURL ?texturl].
?map atlas:metadata_summary ?ms.
?ms atlas:item [atlas:metasum_id ?item_id]
FILTER (regex (**?maptitle**, \+*qmaptitle*+\) && regex (**?texttitle**, \+*qtext*+\)). }.
?item atlas:metasum_id ?item_id.
?item rdf:type **?itemtype**
?item ms:topicov [dc:subject **?topic**].
?item dc:publisher **?publisher**.
?item dc:description ?abstract.
FILTER (regex (**?topic**, \+*qtopic*+\) && regex (**?publisher**, \+*qpub*+\) && regex(**?
itemtype**, \+*qrestype*+\)).
};

Such composite query is intended to get specific items out of a data graph (in this case is the atlas directory) using keywords FROM NAMED, WHERE, and GRAPH. The items returned in that part of query, are then matched against another data graph (in this case is the RDF data of metadata summaries). In Lisa's case, the submitted values of "*qtopic*," "*qpub*," "*qrestype*" to find pattern in the metadata summary correspond to "Environment," "VROM," "WFS." The "*qtext*" in the directory meanwhile, corresponds to "sound pollution." These values are submitted through the widgets form offered within the user interface. As seen in the above code, these values are used to filter the query (using keyword: FILTER).

6 Visualizing the GDI resources

The previous sections discuss the organization of metadata and the handling of queries. This section discusses the visualization of metadata summaries and the GDI resources. As our aim is to facilitate discovery and synthesis of the GDI resources, the mapping functionalities play a crucial role to represent the thematic maps of the atlas, the storyteller content, and search results including the visualization of metadata summaries as well as the WFS and WMS accessed.
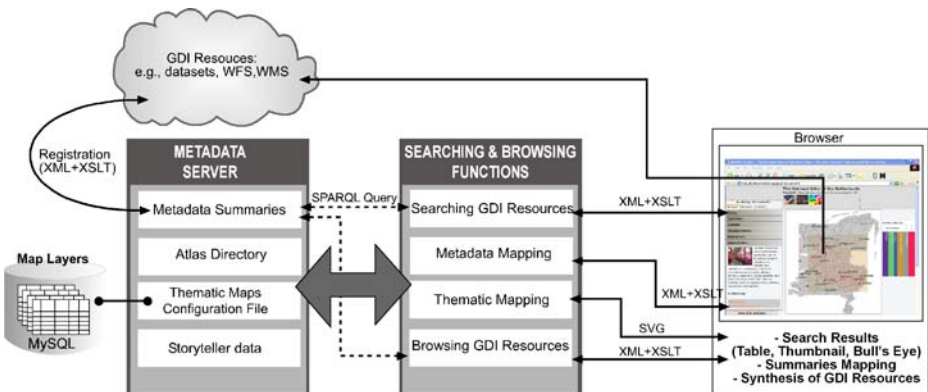
Fig. 3 RDF Query (with SPARQL) and content transformation (with XSLT) play an important role for presenting the search results and the synthesis of metadata summaries of GDI resources

Figure 3 depicts a simplified process of the query and mapping functionalities of the system. Through the browser, users are expected to send a request for a specific information retrieval or presentation, the browser passes this request to the server. At the server side, the corresponding query against the directory and (or) the metadata summary is processed. The result of the query is encoded as "SPARQL query results XML format" [26]. As shown in the following excerpted listing, this response is transformed first into a simple XML data before it is sent into the stream.

```
<results>
  <item meta_id="http://overijssel.nl/gis/transport/dataxml/buszone.xml">
    <westlimit>182000</westlimit>
    <northlimit>540000</northlimit>
    <eastlimit>260000</eastlimit>
    <southlimit>450000</southlimit>
    <publisher>Opleiding van GIS Province Overijssel</publisher>
    <title>Overijssel bus zones</title>
    <topic>Transportation</topic>
  </item>
</results>
```

At the browser side, a corresponding XSLT template is used to transform that simple XML data in the stream into a specific requested presentation (see Table 1). The requested presentation, either as Scalable Vector Graphic (SVG) or HTML page, is then inserted into the HTML document or the embedded SVG graphics. As the use of XMLHttpRequest object (AJAX—Asynchronous JavaScript and XML approach) in the browser to support server-client communication is proven to give benefits in terms of performance and usability [27]–[29], this application implements the AJAX approach in handling requests and responses.

The following will only focus on the presentation of search results and the visualization of metadata summaries and the external feature services.

6.1 Search results presentation

Search results can be represented in forms of: table and thumbnails' views. In conjunction with table and thumbnail views, search results can also be presented as and linked to graphical representations. Metadata mapping and the bull's eye metaphor are the graphical representations that the Aim4GDI provides.

Table 1 The types of the GDI resources' representations and their corresponding data processing on the server and browser

Display	Server		Browser + JavaScript		
	RDF data		XML data	XSLT template	CSS/XHTML/SVG
Table	SPARQL Query	→	XML Results →	Table.xslt	→ Table view (XHTML)
Thumbnails	SPARQL Query	→	XML Results →	Thumbnails.xslt	→ Thumbnails view (XHTML)
Bulls eye	SPARQL Query	→	XML Results →	DrawBE.xslt	→ Bulls eye view (SVG)
Footprints	SPARQL Query	→	XML Results →	DrawMS.xslt	→ Metadata boxes (SVG)
WFS	Proxy access	→	GML →	WFS2SVG.xslt	→ Features (SVG)
WMS	Proxy access	→	image →		Image (SVG)

The table displays each item of search results as a row. In the thumbnails' view, sample images (map pictures) of data or map/feature services with its corresponding title are represented for each result. Metadata mapping concerns with the depiction of metadata footprints (taken from metadata bounding box) into the map with stylized attributes are also plotted on top of the map. The bull's eye metaphor represents the degree of relevance of search results against the query (see Fig. 4). The relevance can be the degree of closeness of spatial or thematic distance of the search results to the query. In this paper, we only concentrate on the spatial distance.

To generate the table view, an XSLT template is applied against each item of the returned search results into rows and columns presentation. Meanwhile to produce a thumbnails' view, the associated stylesheet is used to plot the submitted preview image along with the title of the summaries. Although the table and the thumbnail presentations are familiar displays to web users, however, they still need functions to conveniently filter the results. Hence, sorting functionality could be helpful to improve the efficiency of searching. In most today's geoportals (e.g., [9]–[11]), this function unfortunately is hardly offered to users. Concerning our previous example, Danny could prioritize the scale attribute during his assessment for the fitness for use by sorting out items by its scale or title, for instance. To enable users sorting out the items displayed, a sorting template (an XSLT file) against the XML data is prepared. Suppose a user sorts the results, the parameter value of the selected field to be sorted is passed to the template via a JavaScript function. This value needs to be added into the xsl:sort attribute to enable the sorting data is processed.

```
<xsl:param name="sortKey" />
<xsl:param name="sortOrder" />
<xsl:template match="results">
  <table>
    <tr>
      <th href="#" onclick="sortTable('title','ascending');">Title</th>
      <th>Abstract</th>
      <th href="#" onclick="sortTable('format','ascending');">Format</th>
      ...
    </tr>
<xsl:apply-templates select="item">
  <xsl:sort select="*[name( ) = *$sortKey] order={*$sortOrder}"/>
</xsl:apply-templates>
<xsl:template match="item">
    <tr>
      <td><xsl:value-of select="title"/></td>
      <td><xsl:value-of select="abstract"/></td>
      <td><xsl:value-of select="format"/></td>
      ...
    </tr>
  </table>
</xsl:template>
```

Fig. 4 (a). Searching (Danny's case): available road datasets and feature services are presented as a table, ▶ thumbnails and a bull's eye view (b). Browsing (Lisa's case): Metadata regarding road datasets are superimposed on top of point features of WFS concerning the environmental quality and the population density map of the Overijssel province. Next to it is the display of the metadata legend

a

Title	Abstract	Format	Scale	Publisher	Type of resource	Semantic Relevance	Geographic Relevance	Map It
Enschede Roads	This data depicts the class roads in the municipality of Enschede	GML	1:250000	ITC	WFSitem	1.3	274612.068547615	☐
train station from nwb-spoor	the roads determined by the municipalities which are rather to be used for transporting of dangerous goods.	coverage	1:10000	GisOV	item	1.3	301230.4432905147	☐
ANWB Car Route	Overview of the ANWB car routes in the province of Overijssel.	shapefile	1:10000	GisOV	item	1.3	287502.1389137827	☐
Cark park Areas	Cark park Areas – Een bestand met de in de provincie Overijssel aanwezige carpoolpleinen met daarvan weergegeven het gebruik	shapefile	1:10000	GisOV	item	1.3	287502.1389137827	☐
Roads for transporting dangerous goods	the roads determined by the municipalities which are rather to be used for transporting of dangerous goods.	coverage	1:10000	GisOV	item	1.3	301230.4432905147	☐
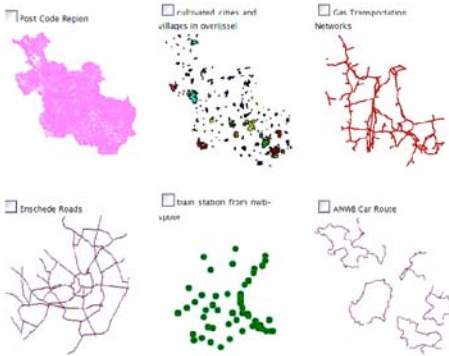
b
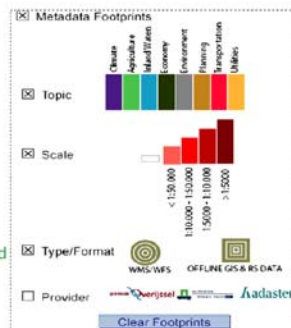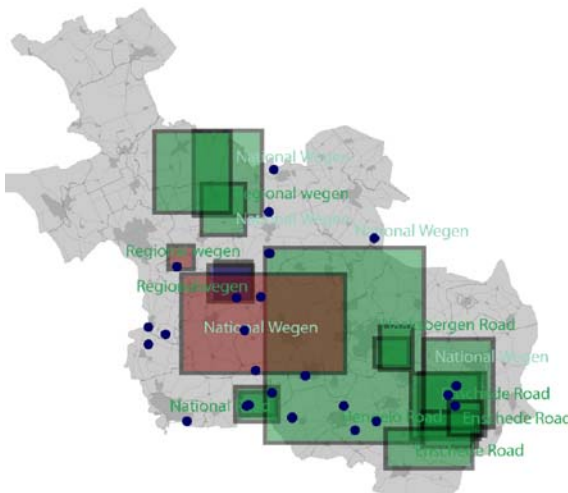
As mentioned previously, graphical representation of search results are: point symbols on top of bull's eye metaphor and metadata mapping on top of the main map / thematic layers. The bull's eye metaphor has been used often in the field of information retrieval to signify the assessment of search results [30], [31]. In case of projecting search results into the bull's eye pane (Fig. 4a), the transformation stylesheet is applied against the XML data. The items of search results are represented as the point symbols, which are plotted according to their geographic relevance (in terms of area distance) to the area of interest of the query.

As applied in the Alexandria Digital Library [32], the geographic relevance in this work is calculated using Hausdorff distance measure; meanwhile orientation of each item is defined based upon the center of metadata footprints towards the center of area of interest. Hausdorff distance measure is commonly applied to detect shape similarity (shape, size, and location) in the fields of image matching and pattern recognition [33]. The algorithm calculates the maximum distance between any points of one bounding box to the other bounding box. Hausdorff is chosen as it is simple to be implemented; yet it gives a reasonable outcome due to its sensitivity to position [34].

The stylesheet template that is used to project items of search results on top the bull's eye is excerpted in the following. The key parameters to plot the items of results are the angle and the scaled-Hausdorff distance. Different styles of representations are generated based upon the resources' characteristics; thus, the filtering process is applied in the transformation template. This gives wide options to style the display of the GDI resources. As exemplified here, when the type of a resource is WFS, the symbol is a circle and a specific class is assigned, otherwise a rectangle is drawn and another class is used. The style class is assigned based on the value of its attributes, for example its topic, data representation, and scale. This style class corresponds to the CSS (Cascading Stylesheet) fragments that rules the visual appearance of the item displayed.

```
<xsl:choose>
<xsl:when test="$stringres = 'http://metasum/elements/1.0/WFSitem'">
   <circle cx="{$dhsc * $cosX}" cy="{$dhsc * $cosX}">
      <xsl:attribute name="class"><xsl:value-of select="resource"/></xsl:attribute>
   </circle>
   <text x="{($dhsc $cosX) + 1.5}" y="{$dhsc $cosX}"/>
      <xsl:value-of select="title"/>
   </text>
</xsl:when>
<xsl:otherwise>
   <rect x="{$dhsc $cosX} " y="{$dhsc $cosX}">
      <xsl:attribute name="class"><xsl:value-of select="resource"/></xsl:attribute>
   </rect>
   <text x="{$dhsc $cosX} " y="{$dhsc $cosX} "/>
      <xsl:value-of select="title"/>
   </text>
</xsl:otherwise>
</xsl:choose>
```

Results presented in the interface (either in the table, thumbnail, or in the bull's eye view) can be superimposed on top of the default map and or selected thematic layers. The area of coverage is symbolized by the bounding box values of the metadata. The topic is differentiated by different color hues. Other attributes of the summaries are represented as

textual and point symbols within its bounding box. Specifically, symbols referring to scale categories, types of GDI resources as well as providers' logo can be displayed or hided on request (the layer control is offered in the metadata legend).

The stylesheet template used for projecting summaries on top of the map has two processing principles: (1) assigning values of the style class of the summary's elements and (2) plotting the bounding box and its correspondent center. The following exemplifies the syntax used to assign a corresponding style class for the "scale" element of metadata summaries.

```
<xsl:variable name="plot_scale">
    <xsl:choose>
        <xsl:when test="scale &gt; 50000">
            <xsl:value-of select="'scale1'"/>
        </xsl:when>
        <xsl:when test="(scale &lt; 50000) and (scale &gt; 10000)">
            <xsl:value-of select="'scale2'"/>
        </xsl:when>
        <xsl:when test="(scale &lt; 10000) and (scale &gt; 5000)">
             <xsl:value-of select="'scale3'"/>
        </xsl:when>
            <xsl:when test="scale &lt; 5000">
        <xsl:value-of select="'scale4'"/>
            </xsl:when>
        <xsl:otherwise>
            <xsl:value-of select="'scale0'"/>
        </xsl:otherwise>
    </xsl:choose>
</xsl:variable>
```

Meanwhile the following fragment shows the syntax used to convert the values of an item's bounding box into a rectangle element in SVG with its associated style.

```
<g>
  <rect x="{$xnw} " y="-{$ynw}" height="{$diverY}">
      <xsl:attribute name="class"><xsl:value-of select="$res"/></xsl:attribute>
  </rect>
      <text x="{$xnw + ($diverX div 2)}" y="-{$ynw -($diverY div 2)}">
  <xsl:value-of select="title"/>
  </text>
</g>
```

With the style class has been defined earlier, such as in the case of the "scale" element shown earlier, the corresponding SVG symbol for representing a specific value of element of metadata summaries can be plotted using this specific processing:

```
<g class="plot_scale" visibility="visible">
    <use>
        <xsl:attribute name="xlink:href">
            <xsl:text>#</xsl:text><xsl:value-of select="$plot_scale"/>
        </xsl:attribute>
```

```
<xsl:attribute name="transform">
        <xsl:text>translate(</xsl:text>
        <xsl:value-of select="$xnw + ($diverX div 2)"/>
        <xsl:text>,-</xsl:text>
        <xsl:value-of select="$ynw - ($diverY div 2)"/>
        <xsl:text>) </xsl:text>
        <xsl:text>scale(250)</xsl:text>
</xsl:attribute>
    </use>
</g>
```

6.2 Superimposition of summaries and feature services

This section will elaborate on the mapping scheme of summaries and feature services superimposed over the thematic map while browsing the GDI storyteller.

To superimpose symbols of summaries over the map selected, the query by its linkage to the map selected is processed against metadata summaries (as shown in the first query example in the Section 5.2.). The step to process the output of the query into the SVG format involves the same stylesheet applied to project a single item of search results into the map.

Although the plotting of metadata footprints is not new (e.g., [35], [36]), however, the approach is commonly applied as the results of search request (as seen in [36]) or to give indication of the data coverage (as seen in Alexandria Digital Library [35]). The superimposition scheme is designed to offer possibilities for users to quickly review the area and the topic coverage of the GDI resources. Additionally, it is intended to enable users to compare and check density and pattern of the data and feature services offered within the GDI, either they are in the same topic of interest or cross-topic. As an example, we recall the Lisa's task. She could compare the availability of datasets related to the map of environmental quality and datasets related to the map of transportation networks. This can be done by loading the summaries that are related to the map of "transportation networks" on top of the displayed map of "environmental qualities" and its related summaries, or vice versa, depending on the focus of analysis. Such interaction scheme can provide a concise indication about the data and her area and topic of interest, thus she could quickly overview the data availability and hypothesize which datasets or map/feature services could be important to support her job. To help Lisa (and Danny and other users) assess the data suitability, the metadata legend is offered. Through the metadata legend, the visibility of the metadata components can be controlled. A line connecting the values of metadata elements signifies the individual summary inspected (Fig. 4b).

Each summary displayed over the map is accessible. Depending on the type of summaries selected for access, users can get different presentations. In case of the summary represents datasets, a new window that displays original metadata descriptions from the data provider web site is shown. If the summary is representing a feature service, the user can get the feature overlaid. For the later case, a transformation template *wfs-gml2svg.xsl* is used against the WFS response resulted. The correspond CSS style class is applied against the feature mapped. The subsequent lines show the principle processes that are applied to

convert GML returned from a feature service (polygon features) into the SVG elements to be inserted into the main map.

```
<!– draw base polygon detail –>
<xsl:template match="gml:featureMember" mode="baseDetail">
     <!– get key properties –>
     <xsl:variable name="fid"><xsl:value-of select=".//@fid"/></xsl:variable>
     <!– write SVG path element –>
     <g>
        <xsl:attribute name="class"><xsl:text>defArea</xsl:text></xsl:attribute>
        <xsl:apply-templates select=".//gml:Polygon">
           <xsl:with-param name="ID"><xsl:value-of select="$fid"/></xsl:with-
           param>
        </xsl:apply-templates>
     </g>
</xsl:template>
<xsl:template match=".//gml:Polygon" mode="class">
...
</xsl:template>
```

The organization of layers and styles of thematic maps as the background information for the superimposition scheme is handled using RIMapper architecture [37]. The approach that we use is to embed a static main map (an SVG map file) into the main HTML page. The thematic layers, symbols of metadata summaries, and corresponding legend are dynamically inserted into this SVG file on request with help of AJAX approach. With respect to thematic layers, they are stored using MySQL as OGC's Simple Feature in the background and proprietary RIM XML file is used to configure map elements and styles issues.

Such application design provides advantages for the atlas editor or GDI administrator to only need to modify the XML file definition and CSS fragments in case of editing the content and styles of the thematic layers and metadata display. In the Legend Box, users can modify the colour values and opacity of the of the layers' display. This could give benefits to users, since users might superimpose the thematic layers and GDI resources as many as they can scrutinize the display for their specific data discovery and analysis purposes.

7 Conclusions and further work

With summaries and the atlas directory expressed as RDF triples, and with the use of SPARQL for querying RDF data, traversing subject, predicate, and object of the GDI resources in support of metadata retrieval is becoming straightforward and well suited for a dynamic web portal such as the Aim4GDI discussed in this paper. Additionally, the ease to match triple pattern and to combine multiple graphs opens up more possibilities to process complex search terms for discovery purposes and to enrich the visualization and integration of the GDI resources. This includes the simplicity to visually cascade metadata items on top of a relevant thematic map for a data discovery purpose.

Users can perform searching of the GDI resources through the linked multi-views in forms of table, thumbnail, and bulls-eye metaphor. Using these linked views, users are capable to project a specific item on the map display. This visualization can give a new and unexpected

insight during the searching and browsing activities. In addition, to advance the browsing strategies, the synthesis of GDI resources through the storyteller view is offered to improve the use of GDI for decision support.

The storyteller view provides a tractable navigation scheme that enables users to sequentially look in detail the thematic-related resources and to interrelate them for comparison purposes. The navigation scheme in the storyteller view is designed to keep users in control of the display while browsing the GDI resources. The idea of combining thematic maps and the storyteller view in support of the browsing strategy of the atlas content is motivated by the facts that browsing geospatial resources in a GDI organization can be overcrowded and difficult when a systematic and coherent navigation scheme is not available. The development of effective interaction mechanism to develop such storyteller is still ongoing. The research activities in that direction as well as the improvement of visualization and interaction of the synthesis of the GDI resources will be our next focus.

Although the integration of WMS into this project has not been the focus, arguably the resulting maps can straightforwardly be incorporated into the SVG interface as shown e.g., in [38]. Another research agenda item in this project is to test the application effectiveness and efficiency in assisting users needs and tasks. For this purpose, we plan to conduct a usability evaluation using scenario-based development [39] involving potential users as test participants. Certain scenarios of uses, similar to Danny's and Lisa's scenarios mentioned in this paper, will be used as a framework to evaluate the prototype's abilities to fulfill the required activity, information, and interaction that users might need during their interactions with the Aim4GDI.

References

1. D.J. Maguire and P.A. Longley. "The emergence of geoportals and their role in spatial data infrastructures," *Computers, Environment, and Urban Systems*, Vol. 29:3–14, 2005.
2. OGC, *Geospatial Portal Reference Architecture*, vol. 2006, https://portal.opengeospatial.org/files/index. php?artifact_id=6669, 1 ed: Open Geospatial Consortium (OGC) Inc., 2004. May 31, 2006.
3. I. Masser. "The Future of Spatial Data Infrastructures," presented at ISPRS Workshop on Service and Application of Spatial Data Infrastructure, Hangzhou, China, 2005.
4. J. Crompvoets, A. Bregt, A. Rajabifard, and I. Williamson. "Assessing the worldwide developments of national spatial data clearinghouses," *International Journal of Geographical Information Science*, Vol. 18:665–689, 2004.
5. A. MacEachren, G. Cai, I. Rauschert, I. Brewer, I. Bolelli, B. Shaparenko, S. Fuhrmann, and H. Wang. "Enabling collaborative geoinformation access and decision-making through a natural, multimodal interface," *International Journal of Gegraphical Information Science*, Vol. 19:293–317, 2005.
6. L. Bernard, I. Kanellopoulos, A. Annoni, and P. Smits. "The European geoportal—one step towards the establishment of a European spatial data infrastructure," *Computer Environment and Urban Systems*, Vol. 29:15–31, 2005.
7. M.G. Tait. "Implementing geoportals: applications of distributed GIS," *Computer Environment and Urban Systems*, Vol. 29:33–47, 2005.
8. T. Aditya and M.-J. Kraak. "Geospatial data infrastructure portals: using national atlases as a metaphor," *Cartographica*, Vol. 41–2:115–134, 2006.
9. USGS, *Geospatial One Stop*, vol. 2006, http://gos2.geodata.gov/wps/portal/gos. Reston, VA: USGS, 2006. May 31, 2006.
10. Natural Resources Canada, *GeoConnections Discovery Portal*, vol. 2006, http://geodiscover.cgdi.ca, 2006. May 31, 2006.

11. INSPIRE, *INSPIRE: European Geo-Portal*, vol. 2006, http://eu-geoportal.jrc.it/, 2006. May 31, 2006.
12. GeoNetwork, *GeoNetwork Opensource*, vol. 2006, http://geonetwork-opensource.org/, 2006. May 31, 2006.
13. Great Britain History GIS Project, *A Vision of Britain Through Time*, http://www.visionofbritain.org.uk/index.jsp, 2006. May 31, 2006.
14. N. R. C. NRC, *The Atlas of Canada*, vol. 2006, http://atlas.nrcan.gc.ca/site/index.html, 2006. May 31, 2006.
15. U. S. Department of the Interior, *U.S. National Atlas*, http://nationalatlas.gov/, 2006.
16. B. Shneiderman. "Designing information-abundant web sites: issues and recommendations," *International Journal Human-Computer Studies*, Vol. 47:5–29, 1997.
17. W3C, *RDF/XML Syntax Specification (Revised)*, http://www.w3.org/TR/rdf-syntax-grammar/, W3C Recommendation 10 February 2004 ed, 2004.
18. S. Decker, F. v. Harmelen, J. Broekstra, M. Erdmann, D. Fensel, I. Horrocks, M. Klein, and S. Melnik. "The semantic web: the roles of XML and RDF," *Internet Computing IEEE*, Vol. 4:63–73, 2000.
19. M. Klein, "XML, RDF, and relatives," *IEEE Intelligent Systems*, Vol. 16:26–28, 2001.
20. S. Powers. Practical RDF. O'Reilley: Sebastopol, CA, 2003.
21. T. Berners-Lee, *Semantic Web Road Map*, vol. 2006, http://www.w3.org/DesignIssues/Semantic.html, 1998. November 11, 2006.
22. The Dublin Core Metadata Initiative, *Dublin Core*, http://dublincore.org/index.shtml, 2006. November 11, 2006.
23. J.J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson. "Jena: Implementing the Semantic Web Recommendations," presented at The 13th International World Wide Web conference, New York, NY, USA, 2004.
24. HP Labs Semantic Web Programme, *ARQ—A SPARQL Processor for Jena*, http://jena.sourceforge.net/ARQ, 2006. May 31, 2006.
25. W3C, *SPARQL Query Language for RDF*, http://www.w3.org/TR/rdf-sparql-query/, W3C Candidate Recommendation 6 April 2006 ed, 2006.
26. W3C, *SPARQL Query Results XML Format*, http://www.w3.org/TR/rdf-sparql-XMLres/, W3C Candidate Recommendation 6 April 2006 ed, 2006. May 31, 2006.
27. A. Charland, *Top Ten Reasons AJAX is Here To Stay*, http://www.developer.com/design/article.php/3567706: developer.com (Gamelan), 2006. May 31, 2006.
28. A. White, *Measuring the Benefits of AJAX*, http://www.developer.com/java/other/article.php/10936_3554271_1: developer.com (Gamelan), 2006. May 31, 2006.
29. L.D. Paulson, "Building rich web applications with Ajax," *Computer*, Vol. 38:14–17, 2005.
30. M. Rorvig and S. Fitzpatrick. "Shape recovery: a visual method for evaluation of information retrieval experiments," *Journal of the American Society for Information Science*, Vol. 51:1205–1210, 2000.
31. A.G. Sutcliffe, M. Ennis, and J. Hu. "Evaluating the effectiveness of visual interfaces for information retrieval," *International Journal Human-Computer Studies*, Vol. 53:741–763, 2000.
32. G. Janee and J. Frew. "Spatial Search, Ranking, and Interoperability," presented at Workshop on Geographic Information Retrieval, SIGIR 2004, Sheffield, UK, 2004.
33. Veltkamp. "Shape Matching: similirity measures and algorithms," presented at Shape Modelling and Applications SMI 2001, Genova, Italy, 2001.
34. N. Grégoire and M. Bouillot. *Hausdorff distance between convex polygons*, http://cgm.cs.mcgill.ca/~godfried/teaching/cg-projects/98/normand/main.html: McGill University, 1998. May 26, 2006.
35. G. Janee and J. Frew. "The ADEPT Digital Library Architecture," presented at International Conference on Digital Libraries, Portland, Oregon, 2002.
36. G. Cay. "GeoVSM: An Integrated Retrieval Model for Geographic Information," presented at The 2nd International Conference, GIScience 2002, Boulder, CO, 2002.
37. B. Kobben. "RIMapper—A Test Bed for Online Risk Indicator Maps Using Data-Driven SVG Visualisation," presented at Location Based Services and TeleCartography, Vienna, 2004.
38. J. Williams and A. Neumann, *Integrating OGC WMS Services with SVG mapping applications*, http://www.carto.net/papers/svg/ogc_wms_integration/: carto: net, 2006. May 31, 2006.
39. J. M. Carroll and M. B. Rosson. "Scenario-based design," in J. Jacko and A. Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Lawrence Erlbaum Associates, 1032–1050, 2002.

Trias Aditya graduated in Geodetic Engineering from Gadjah Mada University in 1998 and M.Sc. in Geoinformatics from ITC in 2003. His Ph.D. topic at ITC (in collaboration with Utrecht University) is "The National Atlas as a Metaphor for Accessing the Geospatial Data Infrastructure". He is based at the Department of Geomatics Engineering, Faculty of Engineering, Gadjah Mada University, Indonesia.

Menno-Jan Kraak holds a Ph.D. in Cartography from Delft University of Technology in 1988. In 1981, he graduated in Cartography from the Faculty of Geographical Sciences, Utrecht University (cum laude). After his compulsory army service (military geography unit), he started to work at the Faculty of Geodesy, Delft University of Technology as a senior lecturer in Cartography in 1983. In 1996, he started at ITC as professor in Geoinformatics, Cartography& Visualization, now titled Geovisualization. In 1998, an additional appoinment followed as professor in new visualization techniques at the Department of Geographical Sciences (since 2003 Faculty of Geosciences) at the Utrecht University. Currently, he is head of ITC's Geo-Information Processing Department.