

Multilayer Architecture for Heterogeneous Geospatial Data Analytics: Querying and Understanding EO Archives

Kevin Alonso, Daniela Espinoza-Molina, *Member, IEEE*, and Mihai Datcu, *Fellow Member, IEEE*
 DLR - German Aerospace Center, Wessling, 82234 Germany

Abstract—The constantly growing process of the Earth Observation (EO) data and their heterogeneity require new systems and tools for effectively querying and understanding the available data archives. In this paper we present a tool for heterogeneous geospatial data analytics. The system implements different web technologies in a multilayer server-client architecture allowing the user to visually analyze satellite images, maps and in-situ information. Specifically, the information managed is composed of EO multispectral and SAR products along with the multitemporal in-situ LUCAS surveys. The integration of these data provides a very useful information during the EO scene interpretation process. The system also offers interactive tools for the detection of optimal datasets for EO multitemporal image change detection, providing at the same time ground truth points for both, human and machine analysis. Furthermore, we show by means of visual analytic representations a way to analyze and understand the content and distribution of the EO databases.

Index Terms—Big Data, Data Integration, GIS, LUCAS, SAR.

I. INTRODUCTION

ARTH Observation (EO) data are in a constant growing process since the last decades. The existing missions generate a massive amount of heterogeneous data that already requires of Big Data approaches in order to exploit the information more efficiently. Thus, the research community faces a heterogeneous Big Data scenario where the main challenge is not only to provide better and more efficient algorithms, but also to design and implement tools that allow a greater exploitation of the available information.

Looking at the evolution of the EO mission goals and the analysis instruments, the data heterogeneity is remarkable. The well-known panchromatic and multispectral image sensors have been complemented by sensors capable of capturing hyperspectral images and a wide range of Synthetic Aperture Radar (SAR) images, generated with different techniques such as Polarimetric SAR (PolSAR) or Interferometric SAR (InSAR). But heterogeneity in EO is not only due to the nature of the imagery. EO products also comprise metadata providing useful additional information such as satellite orbit state vectors, geographical coordinates and data acquisition times. One

last source of heterogeneity comes from third party systems not directly related with satellite EO products but which are commonly used during EO analysis tasks. In this aspect, Geographic Information Systems (GIS) [1][2] play a key role. A GIS is defined in [3] as "a computer-based information systems that enables capture, modeling, storage, retrieval, sharing, manipulation, analysis, and presentation of geographically referenced data". From the initial standalone GIS architectures, the internet development has promoted the intercommunication among GIS specially via Web services [4].

In the same way, plenty of the scientific community work has been focused in the link of information sources by means of integration and fusion of the different information sources. We can find wide scope initiatives such as, Earth Observing System Data and Information System (EOSDIS) [5] or Global Earth Observation System of Systems (GEOSS) [6]. EOSDIS is NASA's system of systems that offers among others, metadata services and user tools which provide searching, visualizing and downloading functionalities for remote sensing data. GEOSS offers processing functionalities to discover, interact and access diverse EO information for a broad range of users. More specialized implementations have been presented for security and hazard decision makers like GEODEC [7] or the system introduced in [8] which aims to support the Earthquake research and disaster response. On the same subject, the work in [9] presents a geospatial data processing functionality to support collaborative and more efficient emergency response. This is achieved by integrating distributed in-situ data with very high resolution optical EO images providing: geospatial data queries, on demand image processing, and fast map visualizations. We also find technology projects like EOLib [10] or research projects like TELEIOS [11] where EO image metadata and linked data are used as query parameters in order to improve EO image retrieval results. The data fusion from third party sources have also been used in [12], where along with the information extracted from EO image analysis data, another layer extracted from OpenStreetMaps was used in the learning stage of the retrieval system.

The heterogeneous data integration brings new possibilities for data representation and visualization. In this sense, visual analytic techniques try to combine automatic analysis methodologies with interactive visualization tools

in order to improve the understanding and analysis on Big Data scale datasets [13]. In [14] visual analytic techniques were successfully used on large geospatial datasets with data mining purposes. Another remarkable example is LandEx GeoWeb tool [15] which provides a visual search engine to retrieve similar tiles based on pattern inputs and similarity maps.

In this heterogeneous Big Data scenario, this paper aims to extend the work presented in [16] where a web based GIS for land cover visual analytics was presented. The presented system architecture is able to easily integrate different types of EO products, existing digital maps and in-situ information. The actual system implementation integrates in-situ data from the European LUCAS survey [17], publicly available digital maps (e.g., OpenStreetMaps) and different EO products (e.g., multispectral and SAR images). This system is a tool that supports EO analysts and expert users by means of information integration or through analytical processes with the final goal of promoting the use and improving the understanding of the EO data.

The paper is structured as follows: Section II deeply describes the architecture of the system and the multiple layers that compose it. Section III shows the performance evaluation of the system in multiuser and multidevice environments. Section IV presents different functionalities of the system, such as capabilities for a better understanding of EO images (Section IV-A), tools for optimum dataset selection (Section IV-B), and statistical processes to generate interactive visual analytics (Section IV-C). Finally, Section V contains the general conclusions.

II. ARCHITECTURE

This system is designed to support EO analysts and expert users through analytical processes. Thus, it is meant to provide tools for data visualization, statistical analysis and data management, with the objective to improve the EO image understanding and help in the dataset selection for multitemporal change detection. Another technical requirement is the capability to handle multiple users accessing simultaneously the system from different devices running different Operative Systems (OS). Aiming to offer such an ubiquitous tool, the system is based on web technologies following a server-client philosophy. The architecture of the system has been designed using a multilayer approach, see Fig. 1. The main server side is composed by: the data source layer, data ingestion layer, database management system, and user oriented web functionality layer. The system is also specifically designed to rely all the computational complexity over the server, making the client side lightweight. In this way, the client is only composed by the Graphic User Interface (GUI) layer, which can be accessed from any electronic device capable of running a web browser with HTML-5 compatibility.

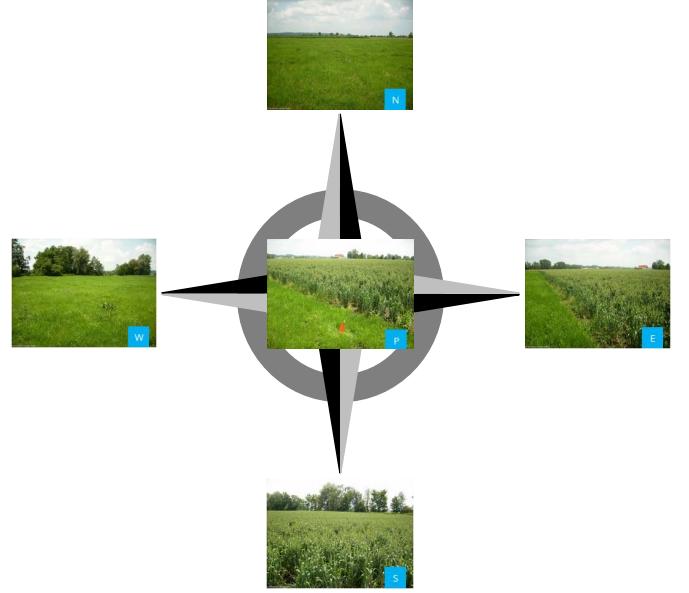


Fig. 2. LUCAS in-situ images of a survey point in Germany. The images are composed of one photo of the exact survey geographical point of the acquisition and four images pointing to the main cardinal directions.

A. Data Source

One of the system's main feature is the possibility to work with heterogeneous information sources. The heterogeneity on the data offers big possibilities to the researchers, allowing them to analyze specific scenarios from different perspectives. The presented system is able to manage and process information from the LUCAS survey and different satellite imagery.

Since 2006, EUROSTAT carries out survey campaigns every three years to monitor the state and change dynamics in land use and cover in the European Union (EU) called the LUCAS survey. The survey comprises on the ground observations that can be divided in three types: 1) micro data of the land cover, land use and environmental parameters associated to the single surveyed points; 2) in-situ photos of each point and landscape photos in the four cardinal directions, see Fig. 2; and 3) statistical tables with aggregated results by land cover/use at geographical level. LUCAS 2009 includes 234.561 points visited in-situ by 500 field surveyors on 23 countries, defining 77 different land cover classes. LUCAS 2012 survey includes 270.389 points visited in-situ by 594 field surveyors on 27 countries, defining 83 different land cover classes. In 2015 between March and October EUROSTAT is carrying out the LUCAS 2015 survey. Surveyors from 28 Member States will visit a total of 273.401 points.

Despite of the actual implementation, the designed system architecture, based on well-known standards, makes possible to easily integrate more data sources, e.g., hyperspectral images.

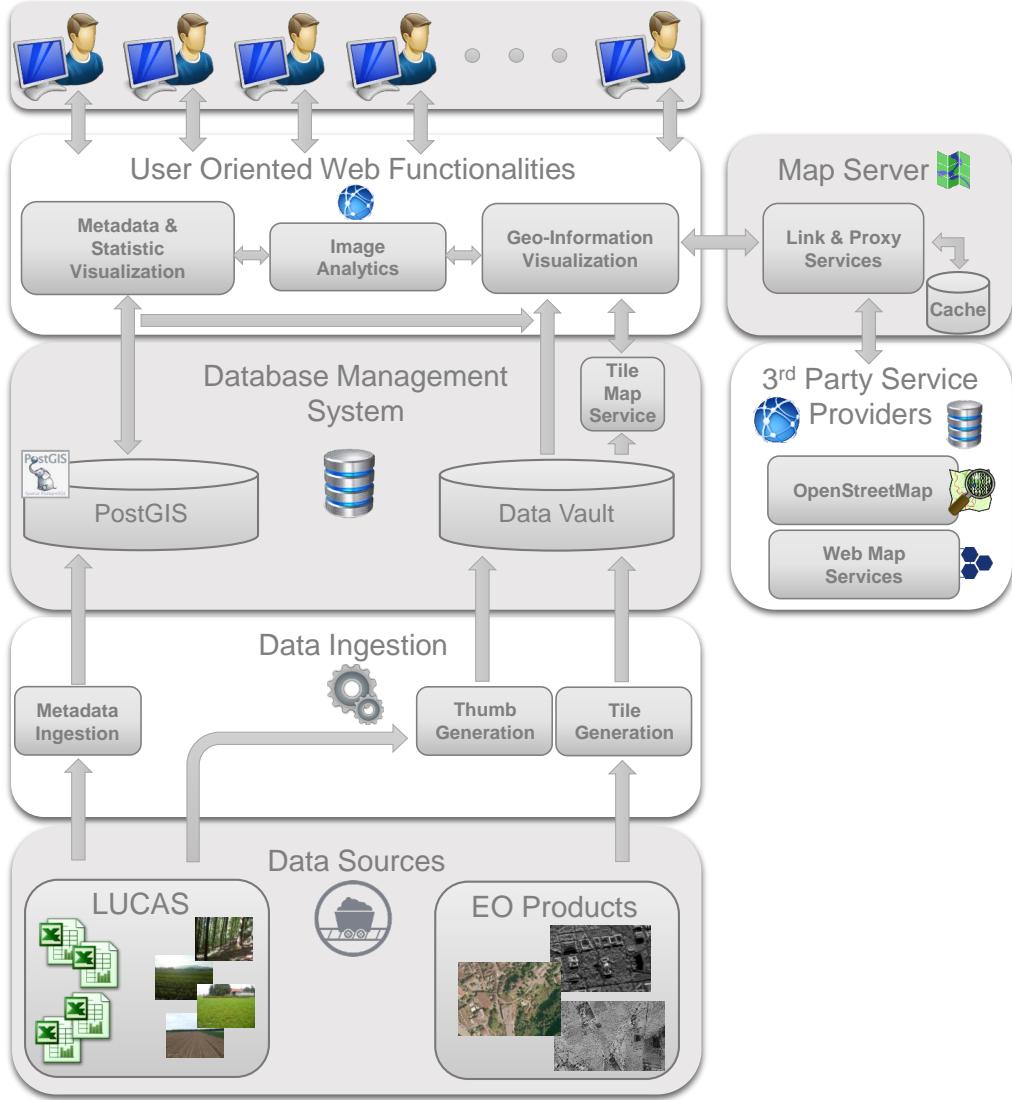


Fig. 1. System architecture. Following a server-client philosophy, the system is designed to rely all the computational complexity over the server making the client side lightweight.

B. Data Ingestion

The Data Ingestion layer performs one time non iterative processes in order to populate the system data repositories. Two main processes can be differentiate: metadata ingestion, and tile generation.

The first process extracts the LUCAS survey metadata, in CSV format, to be ingested into the system geographical database. The ingestion is made by parsing the extracted metadata to series of Standard Query Language (SQL) queries that insert and submit the information to the geographical database.

The second process of the ingestion module produces tiles of the EO products at different zoom levels in order to make more efficient the visualization. The system base projection is WGS-84. If the data source has a different projection, a reprojection process can be performed before the tile generation.

C. Database Management System

The Database Management System is composed by the main geographical database and a data repository in the system archive. The data repository stores the tiles generated from the EO products. The tiles are stored in an very specific directory structure in order to be accessed as a Tile Map Services (TMS) [18]. A TMS only provides access to the geographical map representation of the EO data, not to the data. The geographical database rests on PostgreSQL (object-relational database) technology [19][20] with the spatial database extension PostGIS. The extension adds support for geographic objects allowing location queries to be run in SQL. PostGIS also enables the creation of a database schema defining spatially the content of LUCAS survey of different years, and in consequence, allowing spatio-temporal queries.

A compact scheme of the geographical database for the LUCAS data is shown in Fig. 3. The *Point* table contains

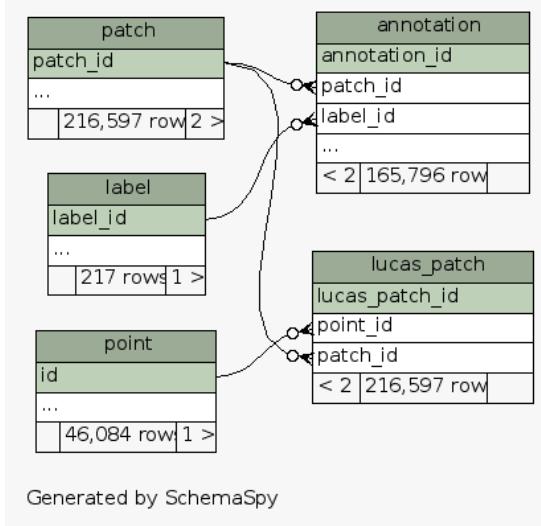


Fig. 3. Compact relationship of the LUCAS database scheme

all the geographical locations of the points where the survey was performed. In addition, this table stores information about the date of survey, latitude and longitude coordinates, and geometry. The *Patch* table stores the information of the photos taken on the observed points. This table provides an URL to each picture in the data repository, thus linking the survey metadata with the multimedia images. The table *Lucas-Patch* consists of the relation between the point and path tables. *Label* table comprises the semantic labels, which describe the land use and land cover categories. The labels are stored following a hierarchy, which is specified in the level field of the table. The *Annotation* table stores the annotation of one patch with several semantic labels, that is the relation between patch and semantic labels.

D. User Oriented Web Functionalities

The User Oriented Web Functionalities module processes every user request. This layer implements all the server logic procedures required. It is divided in three main logic blocks: 1) Geo-Information Visualization, 2) Metadata and Statistic Visualization, and 3) Image analytics.

The first logic block, the Geo-Information Visualization block, performs the communication protocols with third party service providers and/or the Database Management System in order to retrieve the required visual information. Going into detail in the communication with the third party service providers, the system implements in parallel an instance of MapServer [21]. The system centralizes all the communication with the third party providers through the MapServer, who works as a proxy. In this way the user and the system main logic remain isolated and avoid cross-domain communications. Furthermore, MapServer already implements most of the standardised protocols for communication and publication of spatial data on the web, e.g., Web Map Service (WMS) [22], allowing an easy connection between the main server, acting as a client,

and the third party service providers which serve the data. WMS is the most spread geographical map producer standard on the web. WMS providers generate on-live map representations in a pictorial format, e.g., PNG, of the geographic information to every request. In this sense, one last useful functionality of MapServer is the data caching. Mapserver implements tile caching capabilities through the MapCache project which can improve the system performance in an multi-user environment by reducing the data request number to WMS providers. All the obtained visual content can be used in the analytical block or just be directly presented to the user.

The Metadata and Statistic Visualization block is composed by a set of functions and procedures to collect the data from the database required to generate the required data visualization. It is important to point out that the visualization is done in the user machine and this block functionality is limited to the data acquisition, processing and parsing. There are two types of data to handle: 1) raw data, and 2) statistical data. The raw data comprises the information of a single LUCAS point using direct metadata retrieval from the database without any processing step. On the other hand, the statistical data generally involves the retrieval and statistical processing of a group of LUCAS points in a geographical region. Both data are finally parsed to the structure required by the GUI.

The products generated by the previous processes are presented together to the user by the Image Analytic block. This process collects the interaction of the user with the data and sends the required instructions to Geo-Information and statistical visualization processes in case an update is required.

E. Graphic User Interface

The computational complexity relies on the server making the client lightweight and operational in any electronic device capable of running a web browser with HTML-5. Through the user interface, shown in Fig. 4, the user interacts with the system. The main map canvas, upper-center, is developed using WebGLEarth library [23] [24], which takes advantage of the Web Graphic Library (WebGL) [25] technology to render a 3D Earth globe on the browser without any external plug-in requirement. The canvas supports several WMS or TMS layers simultaneously, which can be enable/disable at will. Thus, the user is able to visualize simultaneously or to switch from OpenStreetMap to the EO SAR layer just by clicking a button. The communications required to obtain the WMS layers from third party services are transparent to the user, who receives all the information from the Image Analytics module.

Another feature of the map canvas is the capability to use a polygon based region selection tool, which allows to focus the analysis in an specific area of interest. Once a region is selected, it is possible to: 1) query all the points inside the region, 2) get the survey points with land cover changes among the surveys or even 3) ask for the points

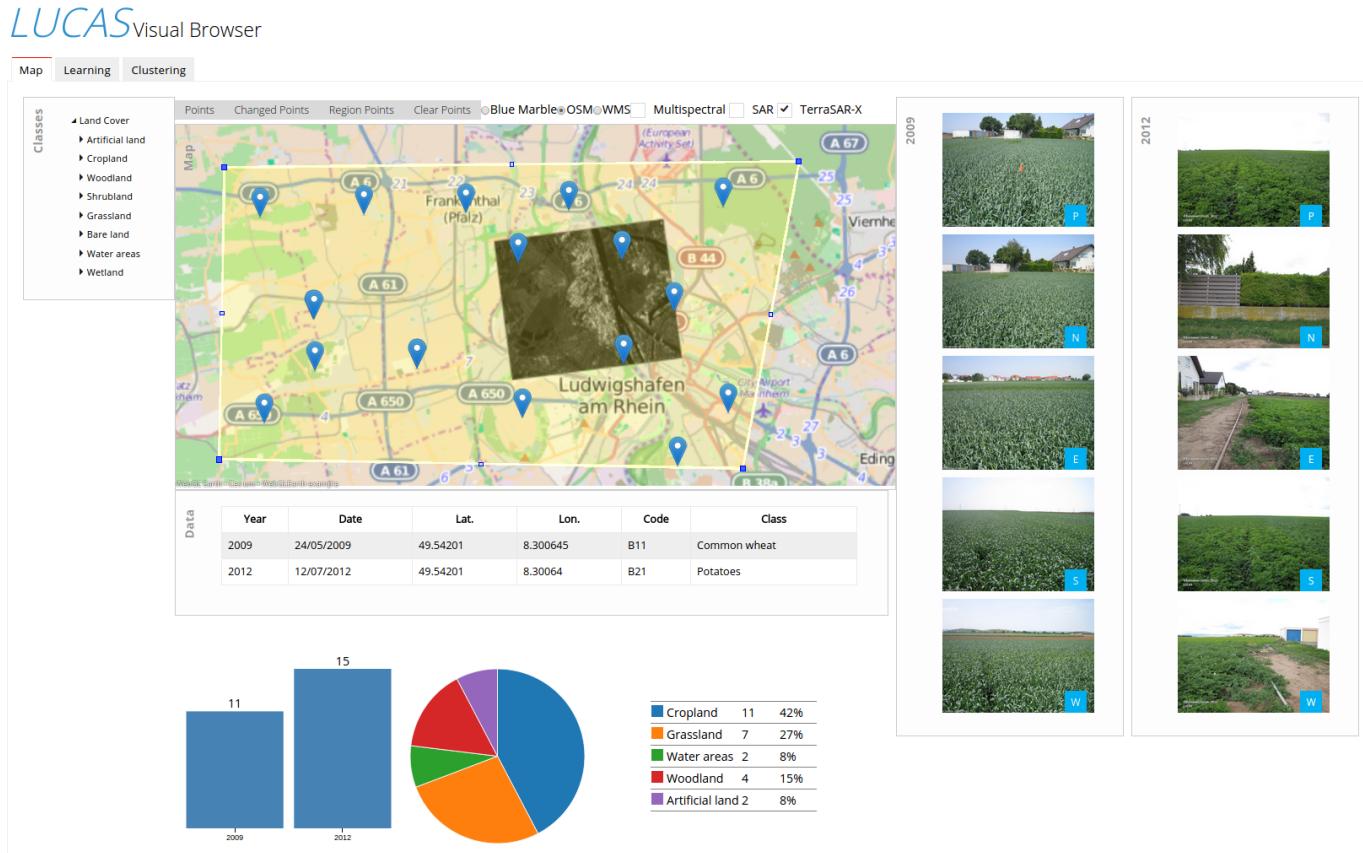


Fig. 4. Graphic user interface of the system.

with a specific land cover. While the first two options are done by pressing the buttons over the map canvas, the specific land cover query is done using the hierarchical land cover tree in the upper-left of the GUI. Besides the preconstructed queries, the users can produce more tailored queries by selecting specific combination of land cover changes, survey dates or even maximum distances between points.

The points are displayed on the map by using markers. Marker generation is a build-in feature of WebGLEarth library, which we customized in order to introduce Scalable Vector Graphics (SVG) markers with the capability to include a variety of color codes to represent different information of the data.

Along with the region points, interactive statistical charts with the point information are presented, lower-center in Fig. 4. The charts are generated using D3JS [26] library which provides a wide variety of interactive visualization components. It is also possible to click over each point for checking its specific information. This information is shown under the map canvas in a table form, in which the user can check specific details about the location, land cover or acquisition times. While the table is generated, the in-situ images of the different surveys are loaded for comparison and analysis in the right side of the GUI.

III. PERFORMANCE EVALUATION

The performance of the system is evaluated on a development workstation with an Intel Core i7-2760QM CPU, 8 Gigabyte of RAM memory, Gigabit Ethernet network adapter and Ubuntu 12.04 as operative system. We present two different performance evaluations: 1) multiuser evaluation, and 2) multidevice evaluation.

The multiuser tests have been performed on an Ethernet Local Area Network (LAN) with a total of ten simultaneous users running Ubuntu or Windows7 as OS with Firefox or Chrome as Browser. During the tests the loading of main page of the system turned out to be the mayor bottle neck in the system performance. The initial HTML file links all the required libraries and has a size of 4.15 Megabyte. This initial loading is unique per user session but it is the most data intensive request the server must handle. The rest of the requests consist mainly of tiles and specific data requests which are numerous but small in size. Typical loading time for the main page is 4 seconds, but it goes up to 12 seconds when al the users start a session synchronized. The multiuser tests also show that once this initial data transmission peak is over the system can handle the user interactions seamlessly. The users are able to use the system with a mean latency around 250 ms with spikes of 700 ms when requesting high amount of tiles.

The multidevice test is carried out on a Wireless Lo-

cal Area Network (WLAN) using the following devices: laptops running Windows 7, Windows 10 and Ubuntu; a smartphone running Android; and two different tablet devices running Android and iOS. As mentioned in Section II by relying on web technologies the system is independent of the OS and it only requires a web browser supporting HTML-5 and WebGL. Thus the system has been successfully tested in the most used browsers in the market: Chrome, Firefox, Microsoft Edge, Safari and Opera.

Nevertheless, before the system enters into production state, a precise study of the user community and server requirements will be performed.

IV. CASE STUDIES

This section aims to show the potential of the system. We present three different use cases. First, we show how the system helps the user to understand the image content. Second, we present a case of study where the system is used for optimum dataset recognition and selection. Both cases exploit the integration of LUCAS survey data in order to enhance or improve the analysis and work performed with EO data. The third case presents visual representations of the database which are used to explore and analyze the content.

A. Image Understanding

Image understanding refers to the capability of the users to interpret the content of the image they are studying. For the purpose of introducing the image understanding capabilities of the system we present three different scenarios in different cities.

1) *Munich*: This initial case of study is placed in the city of Munich, Germany. The system is ingested with the LUCAS information of Germany, linked with an OpenStreetMap layer; and two EO products of Munich: a multispectral image from WorldView-2 and a SAR image from TerraSAR-X. Both EO images have pixel spacing of 1.25 meter, covering an area of 24 km². The size of the total scene is 4890x3202 pixels.

In this case we present a scenario where the availability of heterogeneous data sources from a same location allow a better understanding of the EO scene by expert and non-expert users. The data used on the experiment are shown in Fig. 6. Analysing just the SAR image, Fig. 6a, it is possible for both users to deduce that the main vertical structures of the image correspond to bridges. Moreover, an expert user most probably would interpret, due the intensity of the surrounding pixels, that the bridges are over several lanes of railways.

For the second step of the experiment, the users have also available a multispectral image, shown in Fig. 6b, for the scene interpretation. In this case the railway assumption would be clear. With the multispectral image the initial assumption about the bridges can be modified. The right structure of the image corresponds clearly to a bridge, but a question rises concerning the element on the



Fig. 7. Image understanding case scenario of Karlsruhe, Germany.

left. Due to its width the left structure can not be a bridge where the cars can transit, but it could still be a gangway for pedestrians.

In the next step of this experiment we add one more data source to the scene interpretation process, the map layer with the OpenStreetMap information, Fig. 6c. The map clearly identifies the railways and the big bridge, providing at the same time more detailed information about the surrounding buildings, street names, etc. On the other hand, it does not help with the interpretation of the unidentified structure clearly visible in the SAR image.

The last step uses the remaining source integrated by our system, the LUCAS surveys. Going back to Fig. 6c it is visible a blue marker pointing out the availability of information from the LUCAS survey. Retrieving this information and adding it to the scene interpretation, the users would know that the survey point is classified as *non built-up area* inside the *artificial land* land cover category. Moreover, analysing one of the available photos, see Fig. 6d, the users can finally get an interpretation of the unidentified structure. The structure correspond to a main overhead line supporting infrastructure for the trains.

2) *Karlsruhe*: The second case of study is located in the city of Karlsruhe, Germany. The users have available the LUCAS data, OpenStreetMap information and a TerraSAR-X image. The EO image, Fig. 7, is a spotlight image with 1.25 meter pixel space and a size of 4343x5741 pixels covering an area of 60 km².

Starting again from the SAR image, Fig. 8a, the users are able to identify a slightly brighter striped element in the middle of the street that diagonally crosses the image. Generally streets present a low and homogeneous backscatter coefficient in SAR images, since the flat surface of such targets favours specular reflection. Moving now to the OpenStreetMap layer, Fig. 8b, it is possible to identify clearly two separate roads on the street probably one for each traffic direction. However there is nothing that give

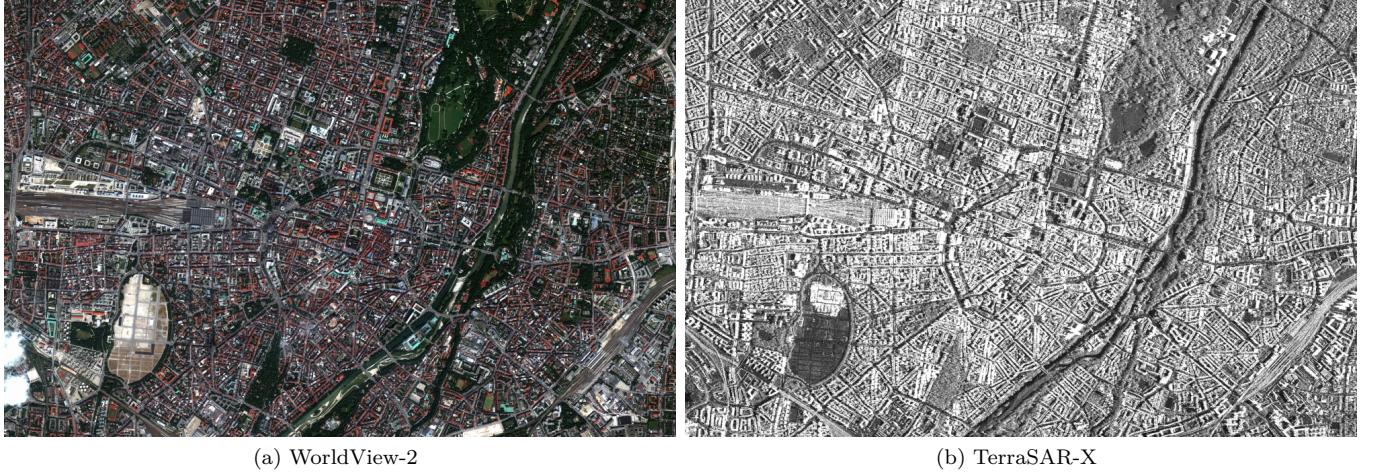


Fig. 5. Image understanding case scenario of Munich, Germany.

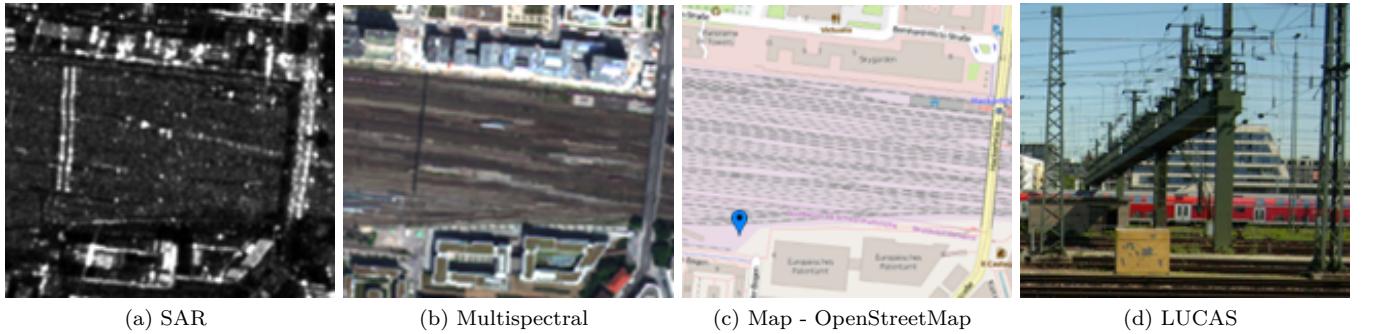


Fig. 6. Munich EO scene understanding. From the (a) SAR image an user can recognize two different bridge structures. When adding (b) Multispectral image to the scene interpretation it is clear that the left structure, because of the small width, can be at most a gangway for pedestrians but the resolution of the image does not allow a correct identification. Adding the (c) Map layer the user realises that the structure is not appearing what practically discard the gangway. Including the in-situ information from (d) LUCAS surveys the user can finally recognise the unidentified structure as a main overhead line supporting infrastructure for the trains.

us an explanation to the stripped element in the middle of the street. As in the previous case there is available in-situ information on that street, Fig. 8c, where you can clearly see a tram line. The more intense back-scatter then is explained by the track ballast that are used to level and hold the rails in place. The ballast is made of crushed stone with sharp edges which increase the back-scatter coefficient in comparison with the road explaining the increase of brightness and, therefore, the unidentified stripped element.

3) *Stuttgart*: The third image understanding scenario is based on Stuttgart, Germany. The available data are composed of the LUCAS survey, OpenStreetMap information and a TerraSAR-X image. The EO image, Fig. 9, is a spotlight image with 1.25 meter pixel space and a size of 6472x3617 pixels covering an area of 56 km².

Focusing our analysis in Fig. 10a, the SAR image shows several low back-scattering rectangular elements, some of them surrounded by high back-scattering elements. The low back-scattering, as explained in the previous example, is normally due to a flat surfaces, e.g., streets, water bodies or sport courts. In this case such amount of small water bodies can only be explained due to agri/aqua-



Fig. 9. Image understanding case scenario of Stuttgart, Germany.



Fig. 8. Karlsruhe EO scene understanding. From the (a) SAR image an user can recognize a slightly brighter striped element in the middle of the street that diagonally crosses the image. When adding (b) OpenStreetMap image to the scene interpretation it is possible to identify clearly two separate roads on the street probably one for each traffic direction. Nevertheless the striped element still remains unknown. Adding the in-situ information from (d) LUCAS surveys the user can finally recognise the unidentified structure as the ballast made of sharp stones and used to keep the rails in place.

cultural exploitation. Aquaculture exploitation can be discarded because these type of installations are usually along river/lake borders. The agricultural use matching this kind of SAR pattern is related with flooded crops like rice. In this case the probability of the image to contain rice fields is very low but nevertheless the use of the map information can help discarding totally this hypothesis. Analysing the map, Fig. 10b, we can see that the element corresponds to the legend of *sportpitch* in OpenStreetMap. Finally analysing the LUCAS in-situ information we can know that the small pitches correspond to tennis courts and the surrounding high back-scattering is probably due to the metallic fences and light poles.

B. Optimum Dataset Selection

In this case of study we intend to show the possibilities of the presented tool for the selection of optimal datasets and ground truth information. This capabilities can be proved useful for the EO analyst in different contexts such as change detection on EO image time series or during new EO analyst training activities.

With this purpose we define a region that encloses the whole Germany and make use of the in-situ information provided by LUCAS surveys of 2009 and 2012. Moreover, taking advantage of the query capabilities described in Section II-E we present different hypothetical scenarios where the presented system can be useful.

In the initial scenario EO analysts need to define a location for the acquisition of EO products for sunflower crop analysis and change detection in time series. Operating the system, it is possible to easily retrieve the existing points with *Sunflower* crop. The procedure is as follows: 1) define a region, 2) deploy the tree showing the land cover classes in order to find the *Sunflower* crop, and 3) double-click to perform the request and visualize the results. The obtained results presented in Fig. 11 show the distribution of sunflower fields in Germany. Additionally, it is clearly visible a bigger concentration of sunflower fields in the north-east, in the region surrounding the capital, Berlin.



Fig. 11. Optimum dataset selection example for sunflower fields. If EO analysts would like to look for EO images containing sunflower crops in Germany, the system will show them the location of this type of field. The EO analysts would notice the proliferation of sunflower fields in the north-east of Germany, surrounding the capital, Berlin.

Taking into account this results, the EO analysts should delimit their analysis region around Berlin and order EO products of this zone. They could also take advantage of the information from the LUCAS points to classify the fields to which the GPS coordinates correspond to a sunflower field.

In the second scenario two EO analysts need to generate a data set for vineyard crop detection. One of them has available our system and the other will make use of traditional web search tools.

The EO analyst using the system follows a procedure similar to the one explained in the previous example, but now selecting the land cover *Vineyard*. The results, see Fig. 12a, show a concentration of vineyards in the west and south-west, with a bigger concentration around the west. The bigger density area corresponds to the Baden-Württemberg and Rhineland-Palatinate region and the fields appear to be along the Rhine river. Hence, the EO analyst should focus the efforts on these regions.

The procedure of the second EO analyst is also very

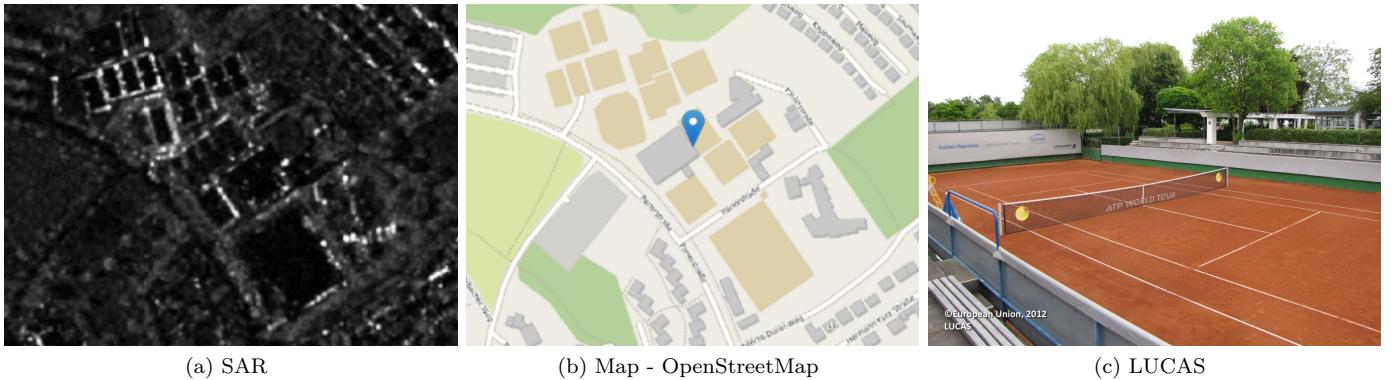


Fig. 10. Stuttgart EO image understanding. The (a) SAR image shows several low back-scattering rectangular elements, some of them surrounded by high back-scattering elements. By means of (b) OpenStreetMap image it is possible to discard agri/aqua-cultural exploitation and to know the actual use of unidentified elements as sport pitches. Finally, analysing the in-situ information from (d) LUCAS surveys, it is possible to know that at least some of them are tennis courts.

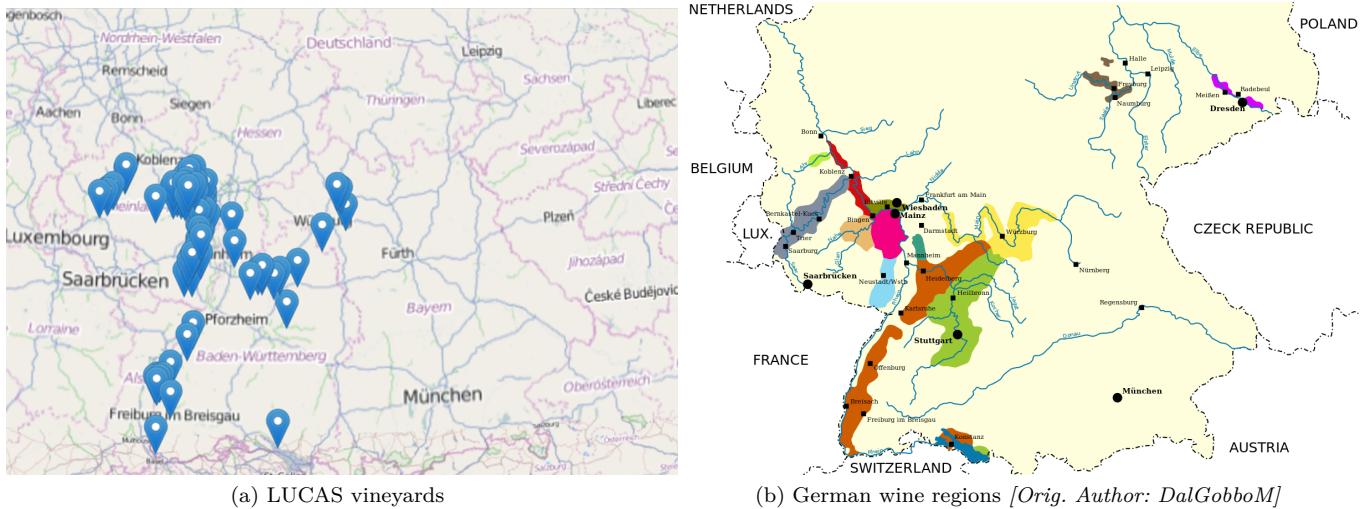


Fig. 12. Optimum dataset selection, vineyard crop detection. Two EO analyst look for vineyards using the system and in parallel, for comparison purposes, a search engine. The first EO analyst obtains (a) using the system feed with the LUCAS information. The second EO analyst obtains (b), a map from wikipedia.org which identifies the main German wine regions. It is visible the resemblance with the result shown in (a) with the main regions shown in (b). The use of the system supports the users with ground truth data and offers results in accordance to the information available on the internet.

simple. Open a web browser and using any of the available search engines type *German wine regions*. In most of the cases, the three first search results include the Wikipedia entry, where you can get access to the map shown in Fig. 12b. This map shows the 13 most important German wine regions, providing the answer to the second EO analyst.

Comparing the obtained results, we can see how both of them are consistent. Our system, making use of the in-situ information provided by LUCAS survey, shows points belonging to the most of the wine regions. It is clearly visible a similarity in the form of the map with the distribution of the LUCAS points. Nevertheless, our system does not provide any result from the smaller regions located in the upper top of the map. This is probably related with the 2 km^2 grid between survey points, which make it difficult to register smaller wine regions.

Ending this case scenario we can sentence that the use of our system is reliable and mostly in accordance

with the information obtained from the web. Moreover, as mentioned previously the use of our system also provides with the additional information that can be used as ground truth in posterior analysis.

C. Visual representation of the archive

In the final case of study we present some of the visualization capabilities offered by the system using the query capabilities and the interactive statistical graphic representations.

The first example presents a visualization of LUCAS data in order to analyze some of the database content. Hence, in Fig. 13 we list the 16 German states and show the percentage of each LUCAS main class using the 2009 data. We can observe how *Cropland*, *Woodland* and *Grassland* are the more common land covers in Germany. Exceptions to this are the city states of Berlin, Bremen

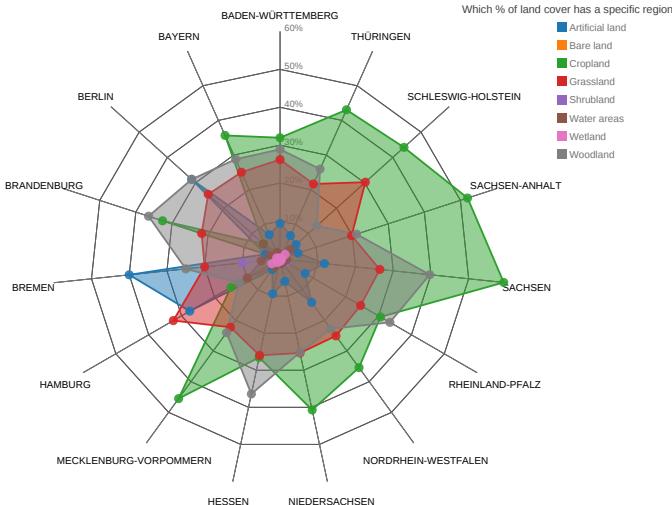


Fig. 13. LUCAS Data analytics. Each axis corresponds to one of the 16 German states and show the percentage of each LUCAS main class using the 2009 data. We can observe how *Cropland*, *Woodland* and *Grassland* are the more common land covers in Germany. The city states of Berlin, Bremen and Hamburg appear as an exception due to their limited extension of land, where the most common land cover is *Artificial land*.

and Hamburg, where due to the limited extension of land the biggest percentage is assigned to *Artificial land*.

The second example presents the available geospatial information from different cities, linking the information of LUCAS with a third EO database, not represented in Fig. 1, which contains semantic annotations of TerraSAR-X products. For the generation of this visualization the semantic annotations from both databases are parsed to a common land cover semantic, allowing the data integration and comparison. The resulting visualization is presented in Fig. 14. There, the inner circle colours correspond to different German cities: Stuttgart, Karlsruhe, Berlin, Bremen, and Cologne. The middle circle colors correspond to the data source of the information, as explained above, LUCAS or TerraSAR-X databases. Finally, the outer circle colours represents the general Land cover classes defined for this visualization: Artificial Land, Agriculture, Forest, Bareground, and Water.

Analysing the visualization is clear that the elements belonging to the LUCAS database are more common in general, with a predominance of elements corresponding to *Agriculture*, *Forest* and in a lesser amount to *Artificial Land*. As in the previous example, it can be easily detected the differences among the city states of Berlin and Bremen with the others. The formers have available a very big amount of TerraSAR-X patches, most of them belonging to *Artificial land*. The other cities with a larger extension of terrain have available much more LUCAS data in comparison to TerraSAR-X.

V. CONCLUSIONS

We have presented the architecture and a prototype of a multilayer system for heterogeneous geospatial data

analytics. The system implements a server-client architecture, which integrates several web technologies. One of the benefits related to the server-client approach is the simplicity of the client. The server is responsible of the most complex processing tasks making possible to offer lightweight clients for different devices. The presented architecture manages the data from the source. The initial layers read the original data and perform transformations to make viable the data integration. These heterogeneous data are linked and stored in a geographical database or in a system repository. The link among the data allows the User Oriented Web Functionality layer to exploit the database capabilities in order to perform geographical queries over the stored data. This layer also implements all the communication protocols to the linked third party services and the server logic that interacts with the user via the GUI.

The presented case studies show the system capabilities managing heterogeneous EO and in-situ data sources. In the first case of study the system proves its utility helping to get a better understanding of EO images for expert and non-expert users. The second case of study use the presented system as a tool for the selection of optimal datasets. Exploiting the in-situ information of LUCAS survey it is possible to use the surveyed point data as ground truth information for change detection on EO image time series. The final case of study presents different interactive visualizations in order to improve the knowledge of the content of the archive.

As future work, before the system enters into production state, we will perform a precise study of the user community and server requirements in order to assure a correct performance and scale of the system. Regarding the functionality upgrade roadmap, we plan to implement different data mining learning tools based on LUCAS and EO imagery fusion. This works will integrate and extent our research presented in [12]. In this way the system will be able to integrate and fuse the information obtained from in-situ sources with the one obtained form the maps and the EO products for machine learning purposes.

VI. ACKNOWLEDGMENT

The authors would like to thank European Space Imaging (EUSI) for providing the WorldView-2 images, the TerraSAR-X Science Service System (Project TELEIOS) for providing the TerraSAR-X images and the European Commission for providing the metadata and images of the LUCAS surveys.

REFERENCES

- [1] K. C. Clarke, "Advances in Geographic Information Systems," *Computers, Environment and Urban Systems*, vol. 10, no. 3-4, pp. 175–184, 1986.
- [2] R. L. Church, "Geographical Information Systems and Location Science," *Computers & Operations Research*, vol. 29, pp. 541–562, 2002.
- [3] M. Worboys and M. Duckham, *GIS: A Computing Perspective, 2nd Edition*, CRC Press, Inc., Boca Raton, FL, USA, 2004.
- [4] N. Alameh, "Chaining Geographic Information Web Services," *IEEE Internet Computing*, , no. October, 2003.

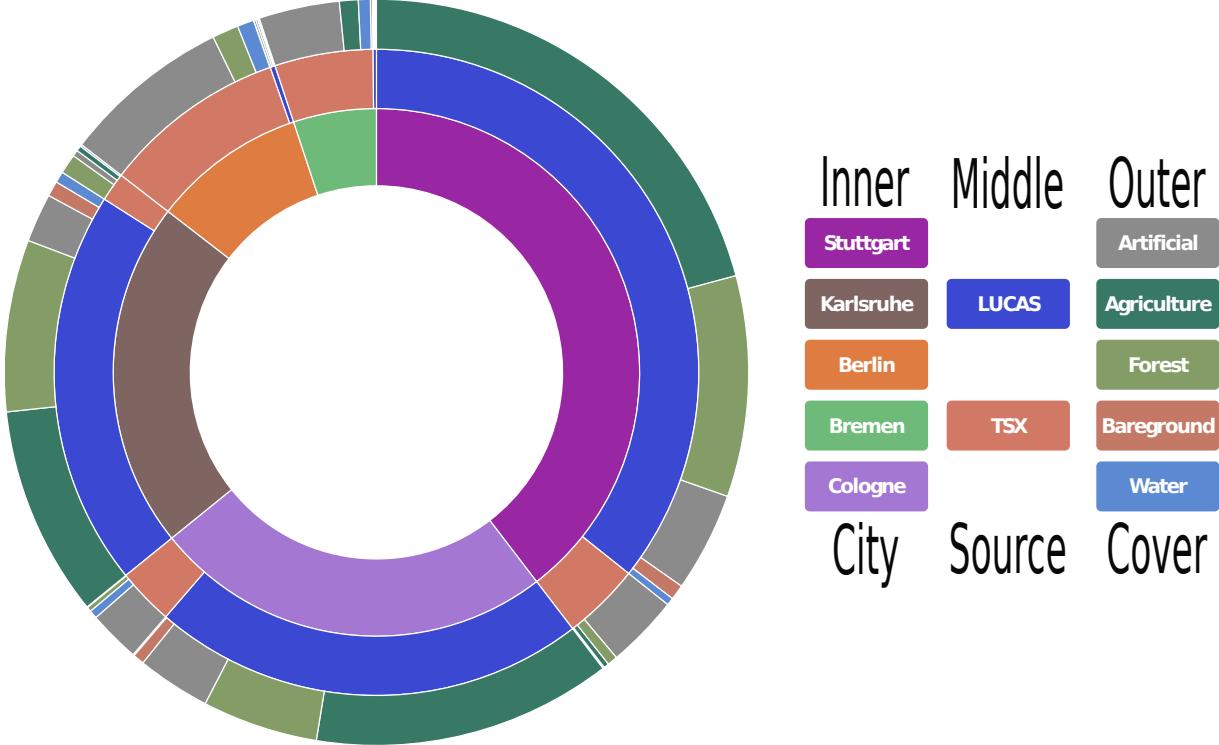


Fig. 14. LUCAS and TerraSAR-X Data analytics. Available geospatial information of different cities joining the information from LUCAS database and a database containing TerraSAR-X EO product semantic annotations. The inner circle elements correspond to different German cities. The middle circle elements correspond to the different data sources. The outer circle elements represent the different Land cover classes.

- [5] B Kobler, J Berbert, P Caulk, and P C Hariharan, "Architecture and design of storage and data management for the NASA Earth observing system Data and Information System (EOSDIS)," in *Mass Storage Systems, 1995. 'Storage - At the Forefront of Information Infrastructures'*, Proceedings of the Fourteenth IEEE Symposium on, sep 1995, pp. 65–76.
- [6] Group on Earth Observations (GEO), "GEOSS 10-Year Implementation Plan," Feb. 2005, [Online: <http://www.earthobservations.org/documents/10-YearImplementationPlan.pdf>].
- [7] C. Shahabi, F. Banaei-Kashani, A. Khoshgozaran, L. Nocera, and S. Xing, "GeoDec: A Framework to Effectively Visualize and Query Geospatial Data for Decision-Making," *IEEE Multimedia*, vol. 17, no. 3, pp. 14–23, 2010.
- [8] J. Wang, M. Pierce, Y. Ma, G. Fox, A. Donnellan, J. Parker, and M. Glasscoe, "Using service-based gis to support earthquake research and disaster response," *Computing in Science Engineering*, vol. 14, no. 5, pp. 21–30, Sept. 2012.
- [9] D. Brunner, Student Member, G. Lemoine, Senior Member, F. Thoorens, and L. Bruzzone, "Distributed Geospatial Data Processing Functionality to Support Collaborative and Rapid Emergency Response," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 33–46, 2009.
- [10] "EOLib. Earth Observation Librarian.," Dec. 2011, [Online: <http://deepenandlearn.esa.int/tiki-index.php?page=EOLIB+Project>].
- [11] D. Espinoza-Molina and M. Datcu, "Earth-Observation Image Retrieval Based on Content, Semantics, and Metadata," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 11, pp. 5145–5159, Nov. 2013.
- [12] K. Alonso and M. Datcu, "Accelerated Probabilistic Learning Concept for Mining Heterogeneous Earth Observation Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015.
- [13] J. Kohlhammer, D. Keim, M. Pohl, G. Santucci, and G. Andrienko, *Solving Problems with Visual Analytics*, vol. 7, 2011.
- [14] D. A. Keim, C. Panse, M. Sips, and S. C. North, "Visual Data Mining in Large Geospatial Point Sets," *Computer Graphics and Applications, IEEE*, vol. 24, no. 5, pp. 36–44, 2004.
- [15] T. F. Stepinski, P. Netzel, and J. Jasiewicz, "LandEx - A geoweb tool for query and retrieval of spatial patterns in land cover datasets," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 257–266, Jan. 2014.
- [16] K Alonso, D Espinoza-Molina, and M Datcu, "LUCAS Visual Browser: A Tool for Land Cover Visual Analytics," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1484–1487.
- [17] M. Kotzeva, T. Brandmüller, and Å. Önnerfors, Eds., *Eurostat regional yearbook 2014*, European Commission - Eurostat, 2014.
- [18] OSGeo, "Tile Map Service," [Online: http://wiki.osgeo.org/wiki/Tile_Map_Service_Specification].
- [19] M. Stonebraker and L. A. Rowe, "The Design of POSTGRES," in *Proceedings of the 1986 ACM SIGMOD International Conference on Management of Data*, Washington, D.C., USA, 1986, vol. 15, pp. 340–355, ACM.
- [20] M. Stonebraker, L. A. Rowe, and M. Hirohama, "The Implementation of POSTGRES," *IEEE Transactions on Knowledge and Data Engineering*, vol. 2, no. 1, pp. 125–142, 1990.
- [21] R. Vatsavai, S. Shekhar, T. E. Burk, and S. Lime, "UMN-MapServer: A High-performance, Interoperable, and Open Source Web Mapping and Geo-spatial Analysis System," in *Proceedings of the 4th International Conference on Geographic Information Science*. 2006, GIScience'06, pp. 400–417, Springer-Verlag.
- [22] OGC, "Web Map Service," Mar. 2006, [Online: <http://www.opengeospatial.org/standards/wms>].
- [23] P. Sloup, "WebGL Earth," Bachelor thesis, Masaryk University, Faculty of Informatics, Spring 2011.
- [24] Klokan Technologies GmbH, "WebGL Earth," 2011, [Online: <http://www.webglearth.org/>].
- [25] Mozilla Foundation, "Web Graphics Library," Mar. 2011, [Online: <https://www.khronos.org/registry/webgl/specs/1.0/>].
- [26] Mike Bostock, "D3 Data-Driven Documents," Feb. 2011, [Online: <https://d3js.org/>].