UPCEO, connecting statistics and people using R

Pau Fonseca i Casas, Raül Tormos, Josep Casanovas

***Abstract*—** A methodology and a tool that implements this methodology are developed using R to construct a web site that allows a lay user to consult statistical information owned by an institution and stored in a cloud database. This methodology was developed following the open-data philosophy and was implemented with open-source software using R as a key element. The proposed methodology was applied successfully to develop a tool to manage the data of the Centre d'Estudis d'Opinió, but it can be applied to another statistical center to enable open access to its data. The system is deployed on a cloud infrastructure that scales according to demand, implementing a 24/7 solution. A user (or a computer program) can access the information on the website using the R language as a communication channel or using a programming application interface. Additionally, in the R language, a common framework can be defined to structure the various processes involved in any statistical operation.

***Keywords*—** Web; Cloud; R language; R-Serve; API; Surveys

I. Introduction

THE primary goal of the project is to develop a methodology that leads to the implementation of a tool to analyze statistical information online. This research has various facets. First, a mechanism must be defined to manage the large amount of data generated by the surveys and the studies, ensuring that the information remains safe and that the analysts can work with it. Second, a mechanism is required to define what information can be published on the web and what information is not ready to be published (e.g., information that must be anonymized). Finally, a mechanism is required to allow mass media, other research institutions, and the general public to work with the data to obtain new information. To solve these problems, a methodology was defined with the aim of simplifying the interaction with the data of all the actors involved.

The tool that implements the proposed methodology (named UPCEO) addresses all of these various aspects; the last feature described in this paper allows the interaction of the users with the data.

This project pursues the idea of open data, i.e., certain data should be freely available to everyone who desires to use them and republish them, as they wish. The concept of data open to everyone is not new. It was established with the formation of the World Data Center system (WDC) during the International Geophysical Year in 1957 – 1958 [1]. In the beginning, the WDC had centers in the United States, Europe, the Soviet Union and Japan, now it includes 52 centers in 12 countries. The Science Ministers of the Organization for Economic Co-operation and Development (OECD) signed a declaration stating that all the information created or found by the public must be freely available [2]. Following this direction, certain legal tools, such as Open Data Commons [3] came into existence to simplify the use of Open Data over the Internet. In that sense, several tools exist that allow the final user to access information, such as the system in [4], a website devoted to the representation of information on a map, or the Socrata® system [5], a system that supports some interesting applications, such as Data.gov [6] that has the primary mission "*.. to improve access to Federal data and expand creative use of those data beyond the walls of government by encouraging innovative ideas (e.g., web applications).*"

There not only exist several websites and tools to access information but also several applications that allow the reuse and sharing of code related to the access of public information, such as [7] or [8]. The next step is to allow users without technical knowledge to access the information and perform easy tasks with it. To do this, the user must be able to execute tasks on a remote server that stores both remote information and certain statistical functions.

The possibility to allow end-users to execute certain statistical functions to obtain new information from the data were described by [9]. Several different tools exist to show information over the web and allow the execution of statistical functions by the end users, e.g., the NESSTAR system [10]. In parallel with these proprietary solutions, several efforts are focused to develop APIs to access statistical information. As an example, Data.org is preparing an API that allows users to interact with the system data to build their own applications and mash-ups; the [11] has also implemented an API to interact with its data. However, the question of how to develop and use these APIs remains. Every infrastructure that develops this type of solution implements a new API, and the developers must be able to address all of them.

Another problem is related to the data preparation; several alternatives exist to define the surveys, e.g., [12] or [13]. These tools allow the user to export the data to various formats to

Pau Fonseca i Casas, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, CA 80034 ESP (corresponding author, phone: (+34) 93 401 7732; fax: (+34) 93 401 5855; e-mail: pau@fib.upc.edu).

Raül Tormos, Centre d'Estudis d'Opinió, Barcelona, CA 08009 ESP (e-mail: rtormos.ceo@gencat.cat).

Josep Casanovas, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, CA 80034 ESP (e-mail: josepk@fib.upc.edu).

ISBN: 978-1-61804-292-7

perform posterior analyses (a well-known format is the Triple-S, an XML for survey software that enables the user to import and export surveys between different software). The main issue with this approach is that manual operations are required to process the data. In our proposed approximation, once the surveys are completed by the users, they can easily be uploaded in the system, and all of the answers can be related directly to the historical representation of each of the proposed questions.

II. The Proposed Solution

The statistical institutions that desire to publish complex studios often deal with complex and unstructured data. For this, we propose a methodology based on the R language [14] [15] that simplifies the CRUD (create, read, update and delete) operations that can be performed over the data. To be capable to interact with the data, it is necessary to define a flow for the statistical studies that a statistical institution wants to publish. To do so, it is first necessary to categorize the data that we own in the system. We have the surveys that are the elements that lead to obtaining information from the representative sample of the population of study. These surveys must also be managed by the system. In our proposal, they are represented by an initial matrix of data, containing the questions (and the answers to these questions). Because a survey can be related with other surveys (to obtain information over time), it is necessary to define a superstructure to relate the various initial matrixes between them at two levels: at the matrix level, and at the table-field level.

Additionally, often the data obtained from the survey cannot be published (maybe some information contained in the data are not anonymous), and hence some transformations to the data must be performed to assure the perfect anonymity of the data. After this is performed, several versions of a study can be published, for example, to correct errors detected in the data. The public must have access only to those matrixes of data that pass the necessary quality control, and the other matrixes are stored on the system as working matrixes but are not accessible to the general public. Every study has descriptors to identify the nature of the study and an identification number. For each one of the studies, at least one matrix representing the survey exists. All of the versions obtained from this work are stored in the study structure. Usually, this implies modifying the matrix structures or adding new information. For that, a working matrix exists, representing the last up-to-date matrix related to the studies. The definitive matrix is the matrix that the users can operate using R operations.

Because various matrixes exist, different roles must be defined. Table 1 presents the minimum roles we propose to achieve with this approach. Each one of these roles has different privileges in the final application. For example, an *analyst* can add new studies, add new matrixes to the system, and modify *working* matrixes, whereas an *external* user can only perform the statistical operations allowed by the system with the *definitive* matrix.

Table 1. System roles.

Role	Description
Administrator:	Controls access to the system and defines the roles of the other users.
Analyst:	Manages the information related to the studies (matrix, documentation, etc.)
External:	Can access the system to perform specific operations.

To manage the matrices of data and allow a modification of these data over a cloud infrastructure, worldwide organizations are developing approaches to share statistical information over the Web using an API. From our point of view, this is not enough to address statistical information and data because of the inherent complexity of its nature, and this approach requires continuous modifications of the API functions to accommodate them to the new requirements of the users and institutions that use these data. In our approach, a statistical language is used, to provide a common mechanism to access all the information. The data contained in the proposed platform can be published over the internet using the statistical language itself. The result is that the user can interact with the system using the full power of the selected language, and there is no need to define new functions through the API to interact with the data.

1.1 Beyond the API, using the R language

In our approach, we select the R language [14] due to is power and because is a widely accepted language in the statistical community. R is a free software environment for statistical computing and graphics; see [16] [17] or the web site http://r-project.org. R software can be executed on a wide variety of UNIX platforms, on Windows, on Linux and on MacOS.

This approach is opposite to the approach followed by API development. In this approach, the system allows an authorized user, or program, to access the data and obtain, using R syntax, all the data and information desired. The concern is related not with the implementation of new APIs or protocols to allow access to specific statistical information or data but with limiting the amount of information that can be obtained over the web. This implies limiting the R operations that can be implemented on the server. Fortunately, this configuration can be accomplished through the RServe package [18], which allows the user to define what instructions can be used over the web.

The power of R does not rely only on strong statistical and graphical facilities but also on versatility. Any element of the research community can improve the system by adding new modules to perform statistical operations. One of the packages we need for our approach is RServe. R usually works in standalone applications, and to connect the different services to R, the R-Serve package must be used. R-Serve can be executed from a command. RServe is a TCP/IP server that allows other programs to use the R facilities from various languages without the need to initialize R or link to the R library [19]. Each

connection has a separate workspace and working directory, which is an essential feature for this project.

The sequences to start using the service are (i) start the R console, (ii) on the console, load the RServe library, and (iii) start the RServe server.

For most users, the default configuration is satisfactory; however, for this project, RServe must be configured to coordinate the different elements that comprise the system. RServe usually works with several default parameters that can be modified in the *config* file. The configuration file is located at /etc/Rserv.conf (on a Linux server, this location can be changed during compilation, specifying the option -DCONFIG_FILE=<new path>). New configuration files can be added with the command --RS-conf (this is an argument in the command line). The complete documentation of the package can be found in [18].

1.1.1 Using R on the statistical study lifecycle

Three main areas must be covered: the management of a questionnaire (starting a new study), the management of the matrixes related to the study, and the management of the operations that can be applied to the public matrixes of the study. In each one of these three areas, we propose to use R language as a basic element to simplify the interaction. This leads to a simplification in the maintainability and further expansion of the system.

To prepare a new questionnaire, first and foremost, the questions must be defined. This is not an easy task because of the diversity of questions that can appear in a single questionnaire and also because the various surveys must consistently be related to each other to make it possible to obtain accurate conclusions over time. Various alternatives exists to prepare surveys, e.g., [12], or [13]. Using these alternatives, the questions can be defined, and they can be sorted on questionnaires that the respondents must answer. Often, these alternatives can export the data to various formats for posterior analysis (such as Triple-S). In our proposal, the relations between the various questions that compose the questionnaires must also be defined; this information (which can be stored in the database for its posterior use) helps us in the review of the complete history of the questions. The answers to the various questionnaires and the history of changes are also available. For example, if we include a question such as, "What party would you vote for in the next election?" and in a new version of a questionnaire, it changes to "If elections were to be held tomorrow, what party or coalition would you vote for?" we must keep the relation between both questions, indicating that they represent the same underlying concept. This simplifies the statistical use in the operations tool, merging the information to construct, for example, a time series.

In that sense, the present approach simplifies the ulterior data management; however, this implies that the uploading process is not easy because it is necessary to create the relationships of the questions, surveys and answers in the database. Additionally, the matrix files can be large and represented in various formats. In our approach, all the information is transformed to a specific XML file that always has the same structure. This enables the user to work with surveys that have the answers in several formats, such as Excel, SPSS, Minitab or R, among many others.

Thanks to the use of an XML base representation for the uploading and management of the data matrixes, it is possible to incorporate tools that access the questions. These questions can be presented to the user in various ways, i.e., *editions*. All of the editions of a question can be related, simplifying the operation of merging surveys. The users can build a new questionnaire, and after the questionnaires are defined in the system, they can be related in a matrix that contains the data obtained from the respondents. The key element of our proposed approach is to always retain the relation between the questions, the questionnaires and the answers.

Finally, and because we propose to use the R language, the users can execute the operations written in R (from a subset of the allowed operations) with the data loaded on the system.In this approach, the relation between all of the various questions is preserved. Additionally, the R language will be used as an API to obtain information from the system instead of defining an API.

III. THE UPCEO APPLICATION

Three institutions are involved in this real project, the *Centre d'Estudis d'Opinió* (CEO), the InLab FIB and the Centre de *Telecomunicacions i Tecnologies de la Informació* (CTTI). The CEO is the official survey institute of the Generalitat de Catalunya. It handles the government's political surveys, barometers, election studies, and other public opinion polls in Catalonia. As defined in their institutional functions, "It is a tool (the CEO) of the Catalan government aimed at providing a rigorous and quality service to those institutions and individuals interested in the evolution of Catalan public opinion." One of its commitments is to make the information readily accessible to the public.

InLab FIB is an innovation and research lab based in the Barcelona School of Informatics, Universitat Politècnica de Catalunya - Barcelona Tech (UPC) that integrates academic personnel from various UPC departments and its own technical staff to provide solutions to a wide range of demands that involve several areas of expertise. InLab FIB, formerly LCFIB, has more than three decades of experience in developing applications using the latest ICT technologies, collaborating in various research and innovation projects and creating customized solutions for public administrations, industry, large companies and SMEs using agile methodologies.

The *Centre de Telecomunicacions i Tecnologies de la Informació* (CTTI) [20] is an infrastructure that can host all of the services that the various organizations that belong to the *Generalitat de Catalunya* requires. This infrastructure is maintained by a licensed private enterprise (now T-Systems). This is convenient for the project because, when the CEO publishes a new study, the quantity of resources required to supply the punctual demand can be bigger than the resources required in a usual day. Additionally, because CTTI ensures that the system is working 24/7, it can be convenient for the

daily work to provide the infrastructure for the CEO database to store all of the information regarding the studies. The CEO primarily manages surveys related to political public opinion. The studies derived from these surveys are published on the CEO website to ensure that the public has knowledge about the studies.

We implement a system to simplify the management and use of statistical information over a web. The specific implementation is represented in Figure 1. The system is composed of different layers, each one of which is related to the various services that the system must provide. The web server is based on a WebLogic Oracle® application [21], using Apache Struts [22] [23] and Java as the infrastructure to define the interface of the system and to establish communication with the R system. The main purpose of using R is to implement various operations that deal with data (see 1.1.1). As an example, we use R to obtain the data from the matrix and the surveys that usually are in the original form of Excel spreadsheets, SPSS files or SAS files; here, R is used as the bridge between all of the various file formats. The R language can be used by users and other applications as an API to communicate with the system to obtain statistical data. In Figure 1, the structure of the system is shown. The entire system is on the CTTI cloud infrastructure. The various files related to the application are stored on an NAS system. The studies are stored in an Oracle database to manage the various files of the system. The R application is installed on the system with the RServe package, defining a set of operations (as an API) and publishing them on the internet using the WebLogic platform.

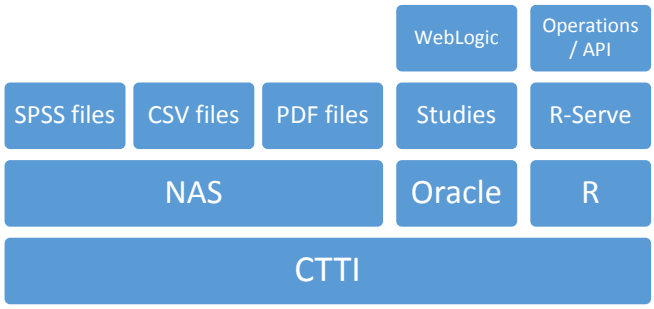

Figure 1. System structure.

From an operations point of view, when a user requests a specific study, he obtains its related documents, mainly *.pdf* files and links to other data related to the survey. With these data, the user can perform various operations (with R), obtaining new data and information. These results can then be exported in CSV file format that can be analyzed in more detail using any statistical package. As shown in Figure 1, the matrix is stored in its original form on the NAS, implying that various formats must be stored in the system. This way, the information generation process can be reproduced exactly as it was by the analyst.

The main file formats that can be used by the CEO analyst are Excel spreadsheets, SPSS *.sav* files and *.csv* files. R is a key element to manage this diversity of formats. Because the application uses R, the information can be read and operated. R can also store or export the new matrix of data in a new format that can be stored again in the database or managed by an external user.

The various functionalities in the system are:

Questionnaire manager manages the questions related to each one of the different questionnaires of the system; see Figure 2 and Figure 3. In our approach, all of the questions must be related to allow a temporal analysis of the data stored on the database.

Matrix manager manages the information related to the matrix generated by the surveys; see Figure 4.

Operation shows the information to the users and other applications (websites) through the R language.

The application can be accessed at http://ceo.gencat.cat/ceop/AppJava/pages. The website is in the Catalan language, and the option that gives access to the operations is "*Banc de dades del BOP*," located at the bottom of the page. This option leads users to the page where a specific study, http://ceo.gencat.cat/ceoa/AppJava/OperacionsExtern.do, is found. This initial listing shows the latest studies performed by the CEO analysts.

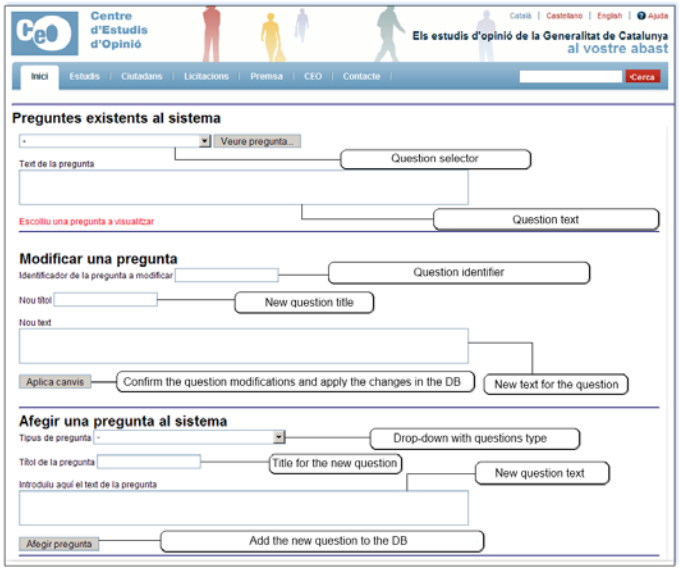

Figure 2. The process of creating a new question is integrated into the application, simplifying the process of reuse and relating the questions of all the questionnaires that exist in the system, as is proposed by our approach.

ISBN: 978-1-61804-292-7

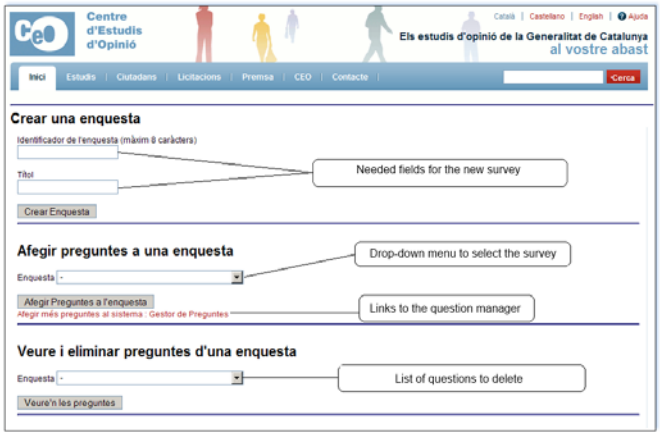

Figure 3. The process of defining a new survey can be performed entirely in the application, simplifying the survey management, as well as its posterior use.

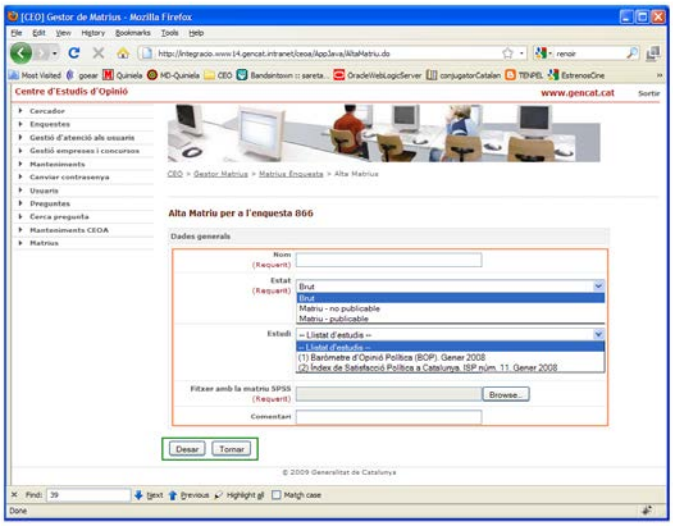

Figure 4. Uploading a new matrix containing the data of a survey to the system.

IV. UPCEO Implementation and Calibration

The entire application resides as a cloud solution supported by the *Generalitat de Catalunya*, hosted by the *Centre de Telecomunicacions i Tecnologies de la Informació* (CTTI). In this cloud solution, the options to work and to modify the upload code are limited, as is explained in section A. Because of the complexity of the structure and the required security concerns, a test infrastructure was implemented to test and implement the R operations. The test infrastructure is composed of a server and a client. On the server side, a machine acts as a Web server (using IBM WebLogic), hosting the MySQL database, storing the data on the NAS (Network Attached Storage) and executing R-Serve. On the client side, a java program (implemented on NetBeans and named JGUIforR; see Figure 5) is used to define the GUI and the R code needed to execute the operations and manage the matrixes.

The client application must first be connected with the server side. The IP of the R server instance we want to use is defined. In this case, the application is connecting with a server that is executed on the same machine as the JGUIforR.

Once this is completed, the connection with the server is established using the File menu. Two options are available. **RComand** implies that the user is working with a local instance of R. In that case, it is not necessary to define the IP. **RComandTCP** implies that the user is working with a remote instance of R; in that case, the IP of the remote server must be defined.

If the connection is established without error, a message appears in the **R Comands** window showing the version of the R engine used on the server side.

To start working, a dataset must be selected, in this case, an SPSS® dataset. Opening a new dataset is as easy as going to the File menu and selecting a new **Matrix** of data.

Once the matrix is loaded, a message is shown to the user in the **R Comands** area, as shown in Figure 5. At this point, all the operations are active, and the user can start working with the matrix.

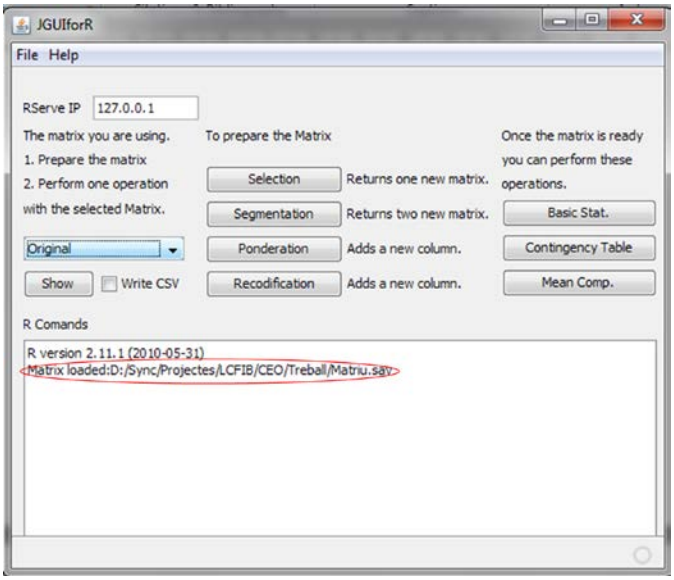

Figure 5. Matrix successfully loaded. All of the options are now activated, and the user can start working with the matrix. The source code of JGUIforR can be downloaded for free at https://svn.java.net/svn/jguiforr~jguifor/.

CEO analysts use this software to understand the operations that the system publishes and to understand the behavior desired in the final implementation of the client, using, in that case, Apache struts [22] to build the website.

As shown in Figure 5, the operations are divided into two main groups. The first includes the preparation of the matrix, selection of a portion of the data of the entire matrix, segmentation of the matrix, weighting of some of the columns of the matrix and recodification. The other operations that can be executed operate over this matrix (calculating the mean, the max, the min values, compiling a contingency table or performing a mean comparison between two variables, etc.).

A. Deploying the system

After the operations perform as expected on the Java platform, the system can be deployed on the CTTI infrastructure. This project represents the first deployment of

ISBN: 978-1-61804-292-7

RServe on the CTTI infrastructure, which implies the need to define roles and protocols to ensure 24/7 support. The system also has high security concerns. First, the application is deployed on the working server, a machine accessible only to the computers located at the InLab FIB laboratory. Once the application passes the tests on this machine, it is deployed at the integration level of the CTTI infrastructure. Here, the application is tested in an environment that is not equal to the production environment but has similar security levels and the same software. After the application performs well there, it can be deployed to a preproduction level. Here, the application runs on an exact replica of the final infrastructure, on the same hardware and executing the same software that the application will find in the production environment. At this level, a set of tests are performed, and the application must pass all of them to be deployed to the production level.

At the production level, the application is available for public use. This is the last step of the deployment, and the current state of the case presented here. Once the system is deployed, the operations performed by the user must never modify the information stored in the server. The system must also be able to store information regarding the various activities that each of the users performs.

When an operation is selected, the R syntax is stored in the database. This syntax is not executed immediately on the system; it is only executed when the user requests results (for example, executes the operations of basic statistics, a contingency table or a mean comparison). This is because the time required to perform an operation bottlenecks at the transference of the data and establishing the connections between the client and RServe. After the connections are established (less than a minute), R performs well and returns the new data very fast.

V. Concluding remarks

This study develops a novel approach to present statistical information over the web following the open-data philosophy. In this approach, the R statistical package is a key element to manage and display the information, allowing the user to perform a number of statistical operations with the data.

From the point of view of data management, the structure of the surveys, the structure that relates the questionnaires and the questions and the related matrix that contains the data, often follow different formats in a real environment. This is true even if a single team manages the information because technology changes and the tools used can be diverse, depending on the objectives of the specific work. This ecosystem of data formats often makes working with the data more difficult. Thus, mechanisms are necessary to translate the information from one format to another. Often, these mechanisms are prone to errors and require the use of tools that are often not well-known by all of the members of the team. In this approach, R is the bridge between the various formats that are stored in the database and is also the language used to recover and work with the information contained in the system. Thus, the CEO analysts store the information in the system using the format they use and understand, and the system is able, using R, to work with the data and to formulate new matrixes of data that can be used again by the experts using their common statistical tools.

Because the system must be able to work at all times, a cloud solution must be implemented to simplify the management of the infrastructure. The amount of access of the external users depends on several factors, e.g., when a new study is offered to the public. This implies that, at times, the traffic to the site is heavy, an aspect that can become a problem for the servers and site management. The cloud solution proposed stores all the information obtained from the CEO studies, allowing 24/7 access to all the information by all the users, and allowing, depending on the user role, the manipulation of the data and the creation of new information and matrixes. Working with the data is accomplished using R as a statistical engine; a user can execute queries and obtain new information regarding the matrixes of data related to a survey. Additionally, because all the operations implemented use R syntax, adding new operations is easy and only requires the addition of a new R code and the definition of a new interface. Thus, the systems implemented based on this approach are extremely scalable and expandable.

Since all of the access to the statistical information is based on the R language, new websites or applications (such as JGUIforR) can be developed that access the data through the use of R statements. This implies that the application goes further than the definition of an API because it uses a statistical language. The power and extensibility of R ensures that we can obtain all the information needed, and the user must only define the subset (if it is needed) of the R instructions that an external user (application or website) can execute. Currently, researchers from various Catalonian institutions are building their own mash-ups using the application. In the future, more capabilities will be added to the application by adding new R language instructions open to public use. There is an additional goal of open access to the institutions, allowing them to access all the information from the CEO servers and define the queries they need for each application (in the broad sense that an application can be a simple query that can reside in a spreadsheet, or a complete web application with various mash-ups).

Last but not least, a set of operations can be defined as an R script. This definition implies that repetitive operations can be performed with fewer errors and in less time.

VI. Acknowledgements

This paper is the result of hard work done by InLab FIB and CEO for three years. We wish to thank the different personnel that are involved with different stages of the project, especially Marta Cuatrecases, Joan Giralt Duran, Sara Royuela Alcazar, José Francisco Crespo Sanjusto, Albert Carrera Mateu and Xavier Canal Masjuan, as members of the InLab FIB that actively developed the application, and Rosa Maria Capo as a member of the CEO that helped us in the development of the tool.

VII. References

[1] World Data Center, "World Data System of International Council for Science," 2010. [Online]. Available: http://www.icsu-wds.org/. [Accessed 11 11 2011].

[2] Organisation For Economic Co-Operation And Development, "OEOECD Principles and Guidelines for Access to Research Data from Public Funding," 2007.

[3] Open Knowledge Foundation, "Legal tools for Open Data," 2011. [Online]. Available: http://opendatacommons.org/. [Accessed 11 11 2011].

[4] open3, "DataMaps.eu," 2011. [Online]. Available: http://www.datamaps.eu/. [Accessed 11 11 2011].

[5] Socrata, Inc, "Socrata, The Open Data Company," 2011. [Online]. Available: http://www.socrata.com/. [Accessed 11 11 2011].

[6] Federal Government, "Data.gov Empowering People," 2011. [Online]. Available: http://www.data.gov/. [Accessed 11 11 2011].

[7] Code for America Labs, Inc , "Code for America," 2011. [Online]. Available: http://codeforamerica.org/. [Accessed 14 11 2011].

[8] Leipziger Agenda 21, "API.LEIPZIG," 2011. [Online]. Available: http://www.apileipzig.de/. [Accessed 14 11 2011].

[9] B. Sundgren, "Making Statistical Data More Available," in *Workshop on R&D Opportunities in Federal Information Services.*, Virginia, USA., 1997.

[10] T. Assini, "NESSTAR: A Semantic Web Application for Statistical Data and Metadata.," in *WWW2002 Conference.*, Hawai, 2002.

[11] New York State Senate, "NYSenate.gov Application Protocol Interface (API)," 2011. [Online]. Available: http://www.nysenate.gov/developers/api. [Accessed 14 11 2011].

[12] Snap Surveys Ltd, "Online surveys," 2012. [Online]. Available: http://www.snapsurveys.com/. [Accessed 20 10 2012].

[13] University of Ottawa, "Snap Surveys," 2012. [Online]. Available: http://www.ccs.uottawa.ca/webmaster/survey/. [Accessed 20 10 2012].

[14] J. Adler, R in a Nutshell: A Desktop Quick Reference, O'Reilly Media, 2009.

[15] P. Teetor, R Cookbook, O'Reilly Media, Inc., 2011.

[16] F. Murtagh, Correspondence analysis and data coding with Java and R, C. S. a. D. A. Chapman and Hall, Ed., 2008.

[17] M. W. Trosset, An introduction to statistical inference and its applications with R, vol. 81, Chapman and Hall., 2010.

[18] Rforge.net, "Rserve - Binary R server," 2011. [Online]. Available: http://www.rforge.net/Rserve/doc.html. [Accessed 14 11 2011].

[19] S. Urbanek, "RServe," 2010. [Online]. Available: http://www.rforge.net/Rserve/. [Accessed 05 July 2010].

[20] Generalitat de Catalunya, "DOGC núm. 5359 - 15/04/2009," 2009. [Online]. Available: http://www.gencat.cat/diari/5359/09082146.htm. [Accessed 9 9 2010].

[21] Oracle, "Oracle Weblogic Server," 2010. [Online]. Available: http://www.oracle.com/technetwork/middleware/weblogic/overview/index.html. [Accessed 11 11 2010].

[22] Apache Software Foundation, "Apache Struts," 2010. [Online]. Available: http://struts.apache.org/. [Accessed 11 11 2010].

[23] C. Cavaness, Programming Jakarta Struts,, O'Reilly Media, 2004.

[24] D. Moore, "SQL Loader," 2003. [Online]. Available: http://www.oracleutilities.com/OSUtil/sqlldr.html. [Accessed 22 10 2012].

[25] A. Billington, "external tables in oracle 9i," 6 2007. [Online]. Available: http://www.oracle-developer.net/display.php?id=204. [Accessed 20 10 2012].

[26] ORACLE-BASE.com, "XMLType Datatype In Oracle9i," 2012. [Online]. Available: http://www.oracle-base.com/articles/9i/xmltype-datatype.php. [Accessed 20 10 2012].

Pau Fonseca i Casas is an associate professor of the Department of Statistics and Operational research of the Technical University of Catalonia, teaching in Statistics and Simulation areas. He owns a Ph.D. in Computer Science on from Technical University of Catalonia.
He works in the InLab FIB (http://inlab.fib.upc.edu/) as a head of the Environmental Simulation area, developing Simulation projects since 1998. He has been involved in more than 20 competitive projects and has published more than 80 papers on journals, conferences and books. He is also a lecturer on Universitat Politècnica de Catalunya – BarcelonaTech, and collaborates with the Universitat Oberta de Catalunya, teaching in Simulation and Statistics area at degree and master levels. His research interests are discrete simulation applied to industrial, environmental and social models, and the formal representation of such models. His website is http://www-eio.upc.es/~pau/.

Raül Tormos is senior survey researcher at the Centre d'Estudis d'Opinió, the official institute for public opinion studies of the Government of Catalonia (Spain). He is also lecturer at the Autonomous University of Barcelona, the University of Barcelona, and the School of Public Administration of Catalonia. He teaches quantitative methods, comparative analysis, official statistics and survey methodology, both at graduate and undergraduate levels. He obtained his PhD (European Doctor) in political science at the Universitat Autònoma de Barcelona. Earlier, he was awarded a full-year stipend by the European Commission (under the TMR funding scheme) as pre-doctoral research fellow at the Mannheim Center for European Social Research (MZES), University of Mannheim. His research interests involve the study of values, attitudes and political behavior, age-period-cohort effects, quantitative research methods and survey methodology. He has done specialized training at the University of Essex, Universidad de Salamanca, Research and Expertise Centre for Survey Methodology, University of Oslo, and University of California at Berkeley. His research has been published in journals such as European Political Science Review or Revista Española de Investigaciones Sociológicas.

Josep Casanovas is a full professor in Operations Research, specializing in Simulation Systems. He is one of the founders of the Barcelona School of Informatics (FIB), of which he was Dean from 1998 to 2004. He is also the director of inLab FIB, a research lab that has been particularly active in technology transfer to business. Among his recent projects is the cooperation in the creation of simulation environments for people and vehicle flow in the new Barcelona airport terminal. He has led several EU-funded projects in the area of simulation and operations research and is a strong advocate of the knowledge and technology transfer function between university and society.