

# Workflows and service discovery: A mobile device approach

Ricardo Holthausen, Sergio Díaz-Del-Pino, Esteban Pérez-Wohlfeil, Pablo Rodríguez-Brazzarola, and Oswaldo Trelles

Computer Architecture Department, University of Málaga  
{ricardoholthausen, sergiodiaazdp, estebanpw, pabrod, ortrelles}@uma.es

**Abstract.** Bioinformatics has moved from command-line standalone programs to web-service based environments. Such trend has resulted in an enormous amount of online resources which can be hard to find and identify, let alone execute and exploit. Furthermore, these resources are aimed -in general- to solve specific tasks. Usually, these tasks need to be combined in order to achieve the desired results. In this line, finding the appropriate set of tools to build up a workflow to solve a problem with the services available in a repository is itself a complex exercise. Issues such as services discovering, composition and representation appear. On the technological side, mobile devices have experienced an incredible growth in the number of users and technical capabilities. Starting from this reality, in the present paper, we propose a solution for service discovering and workflow generation while distinct approaches of representing workflows in a mobile environment are reviewed and discussed. As a proof of concept, a specific use case has been developed: we have embedded an expanded version of our Magallanes search engine into mORCA, our mobile client for bioinformatics. Such composition delivers a powerful and ubiquitous solution that provides the user with a handy tool for not only generate and represent workflows, but also services, data types, operations and service types discovery.

**Keywords:** mORCA, Magallanes, search engine, workflow generation, mobile devices

## 1 Introduction

In the recent years, a vast amount of Life Science tools have been made accessible through the Internet [1]. Without the need of profound computational knowledge, genome browsers, sequence comparison software, public databases, among others, can now be used by both graphic user interfaces and web-services. Nowadays, a large deal of online resources are available to researchers. Nonetheless, they can face the difficulty to find, identify and execute them. Furthermore, and in order to enhance the usability of these tools, the prevailing trend is the combination of them into scientifically repeatable workflows; In this sense, a complete experiment can usually be seen as a work-flow [2] in which different

operations are linked together based on their input and output files, providing a procedure that can be reused by other researchers, thus avoiding a tedious and repetitive task, that is also prone to human error.

For the past decades, repositories have been a common framework regarding Bioinformatics (as exemplified by the existence and growth of EBI[3], NCBI[4] and INB[5]). This type of service metadata storage has become highly relevant in this field as one of the main actors of Internet resources used to find valuable Bioinformatics software [6]. In this line, MAPI [7] was born as a modular software framework for their standardization and use through different clients by a common access layer. On the other hand, Magallanes [8] was developed as an intelligent search engine built for its use with MAPI, using different repositories in order to discover service types, operations, data types and services, along with automatic workflow composition (given two data types, workflows source and target).

Regarding workflow management, several tools have been developed over the years. Galaxy [9], for instance, allows the user to compose workflows by a friendly interface using drag and drop functionality, while Kepler [10] allows to design, execute, reuse, evolve, archive, and share scientific workflows. Moreover, Taverna [11] provides a workflow tool suite in which the user can use both Web Services and local tools. Another distinct project regarding workflows management is myExperiment [12], a sort of social network where users can submit and share their procedures in different formats (Taverna, RapidMiner, Galaxy, KNIME, Kepler, ...).

Recently, the community is starting to be aware of the significant evolution in the use of mobile devices, and the possibilities they provide by reason of their ubiquity and usability. This can be seen with the great amount of tools that are being developed in Bioinformatics, in order to take advantage of the capabilities of this relatively new environment. In this line, different apps have been developed. Galaxy Portal App [13] was born as an interface to the Galaxy system [9] for its use through tablets and smartphones, whereas mORCA [14] was developed to ease the integration of Web-Services using metadata available in different catalogues and repositories such as the INB repository [5].

The proposed work stands out from the previously mentioned workflows management systems by proposing a way of automatically generating workflows from a mobile device. Workflow generation is a difficult task that has been addressed through several different approaches but less commonly in mobile devices. Therefore, the main contribution of this work is to join the concepts discussed previously (web services, repositories, workflow management and mobile environment) by extending mORCA, combining it with Magallanes search engine and workflow generation. This provides the latter with a new interface, and the former with two new features. Besides this, those web-services currently available in

mORCA are linked with the coincident search results obtained from the Magallanes search. Thus, mORCA's functionalities are increased, facilitating finding relevant information, and hence providing biologists an easy-to-use mobile solution for discovering a wide range of resources manually and automatically, taking into account the intrinsic limitations of mobile devices, but empowering its strengths.

Finally, as this work brings together two elements such as mobile environment and workflow representation, the way of representing such pieces of information in this new environment is discussed. To do so, several ways of representing workflows have been analyzed in order to determine which one would be more preferable.

## 2 Methods

In order to provide mORCA with Magallanes functionalities, a Web-Service implementation of the latter has been employed, using its Web Services Description Language (WSDL) interface [15]. Although mORCA's architecture specifications are thoroughly explained in its paper, it is necessary to briefly remark its functioning, in order to gain a better grasp of the current work implementation. mORCA was developed by using a multi-layer architecture, in which components related to the actual tools, services and repositories are provided by MAPI, whereas a second layer works as a middleware for different elements, such as user authentication, repository browsing, service discovery, service parameter composition, service invocation, file management and service execution monitoring.

The present work takes advantage of the first mentioned layer by making use of Magallanes web-service version, whose functionality is based on the client-server model, and that is developed with an established technology such as Java, running in a Java Servlet Container (Tomcat Server).

The Magallanes search engine integration in mORCA has been carried out using the main components of mORCA (developed using web technologies as jQuery and jQuery mobile). mORCA is a modular system, which facilitates the inclusion of new features (repositories, tools, types of visualizations). The main goal was to include Magallanes functionalities while maintaining the original idea of using Ajax requests transparently to the user. More information about this work's methods, along with a sequence diagram can be found in the supplementary material that is provided.

The automatic workflows generation uses the algorithm described in Magallanes [8]. This algorithm proceeds by finding the shortest sequence of non-redundant services that matches the outputs with the inputs, linking the source and target data types. In summary, it identifies all the services and operations

that produce a certain output (target). All those input data types are then used as targets in the next step.

Regarding the search engine interface, the Magallanes original interface has been carefully tailored in order to fit mORCA applications environment and to maintain its responsive characteristics, while preserving all the functionalities available in the web version of Magallanes search engine. As can be seen in Figure 1 (A), the user selects the repository in which to search resources. The search terms are entered by the user in a search-box, and different query options can be set in three different sets of checkboxes:

- In the first one, a certain policy can be selected regarding the treatment of the terms entered (i.e.: or, and, or the use of regular expressions).
- Regarding the second set, the user can be select the type/s of resources he/she wants to discover.
- Finally, in the third one, two search options are added: follow links and case sensitive.

Results, composed by Datatypes, Services, ServiceTypes and Operations are displayed in a dynamic list that distinguish them by their type. Each result is presented in a box, whose components are:

- Result type (Data type, Service, Service type or Operation)
- Name
- A short description
- Links (Optional)
- Execution link (Optional, in case it is an available Service in mORCA)
- Set as workflow source/target buttons (Figure 1 (C))

For reasons of simplicity, and in order to avoid an overwhelming amount of outcomes they are displayed using lazy loading methods. Thus, firstly only 10 search results are displayed, and the list can be enlarged by tapping the More results button (Figure 1 (B)).

One characteristic added to this version of Magallanes is the direct linking of those services in the search results that are available in mORCA. Thus, the user can access to the mORCA services by searching them instead of searching for it in the mORCAs service list for the current repository. Another functionality available at Magallanes that has also been included in its integration in mORCA is the did you mean module, which suggests alternative search terms if it detects a likely typo, so the user can click on the suggested search instead of correcting

the current one. [8].

For the workflow visualization, the source and target data types are displayed. Right below the source data type, a select drop-down menu appears, containing all the operations that receive the source data type as input. When one of them is selected, the corresponding output data type appears below, together with another select drop-down menu with the following available operations, until the target data type is reached. An example of this workflow displaying can be seen in Figure 1 (D).

### 3 Results

In this section, two use cases of Magallanes in mORCA are demonstrated. Firstly, an example of searching and invoking a service by using the search engine will be described. Secondly, a use case of a workflow generation will be depicted. In order to avoid redundancy, the term clicking will be used in reference to clicking or tapping from this on.

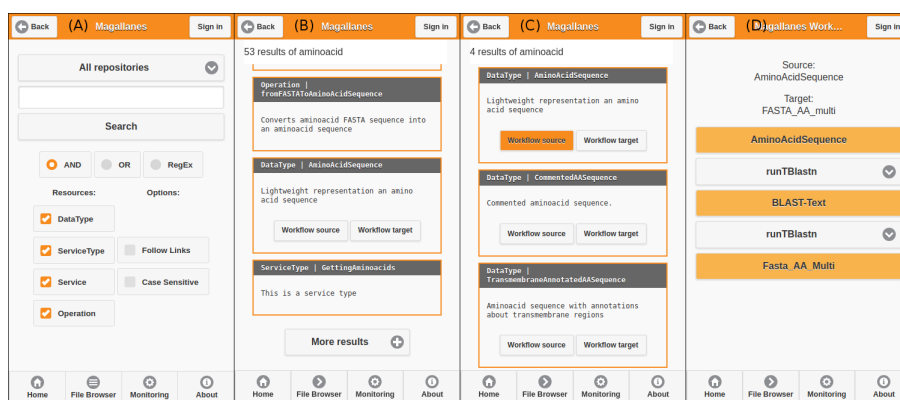


Fig. 1: Screenshots from Magallanes in mORCA. (A) The Magallanes main screen with the search-box and the different options. (B) An example of the search results, and the More results button for the lazy loading characteristic. (C) Search results of data types containing aminoacid. The first data type shown is set as the workflow source. (D) Workflow generated (represented in one of the formats that will be analyzed in the discussion). The data types are shown in orange, and the select-boxes contain the available operations from the data type above.

#### 3.1 Searching and invoking a service

The first action required from the end-user is to (1) go to Magallanes index page by clicking the button in mORCA's main page. Then, (2) the search options

can be set. The user can select a repository to discover resources from, a way of treating the different search terms entered (And/Or/Regular expression), restrict the type of resources to be discovered (service types, operations, services or data types), and set two auxiliary options (case sensitivity and include links in the results). Once the search is done, the user can (3) navigate through the results obtained. Specifically, just ten results are displayed at a first moment, and the user can enlarge this list by clicking in the More results button available at the bottom of the page. Finally, (4) having located the desired service, the user can invoke it right from the results page by clicking the Open resource button, and make use of it in the same way as described in the mORCA documentation [14].

### 3.2 Generating a workflow

The first steps (1, 2 and 3) are similar to the previous use-case. Regarding the search options (step 2), it is advisable to restrict the resources type to just data types, as they are the input and output of the workflow. Once the desired results are found, (4), the source and target data type are selected by clicking in the corresponding button available at the bottom of the result box. When both ends have been selected, a button will appear at the top of the results window. By clicking it (5) the user will have access to the generated workflow, which can be explored by using the select drop-down menus and buttons in the workflow page.

## 4 Discussion

When it comes to Bioinformatics workflow representation, the main manner consists on different boxes for each operation or data-type, which are interconnected by arrows indicating the order. This is a clear and concise way of depicting such a procedure, but when we switch from the WIMP paradigm to the mobile environment, whose relevance is constantly increasing, other ways have to be explored, as mobile devices have intrinsic limitations and differences with respect to the WIMP environment (v.g.: screen size, tapping instead of clicking, etc.) [16]. Besides the usual way of workflow representation (Figure 2 (B), two other ideas have been analyzed).

Firstly, and trying to take advantage of the mobile devices characteristics, we can see a workflow composed by Magallanes in Figure 1.D. The idea is to provide the user with a fully-responsive workflow that can be explored and edited to some extent (i.e.: the workflow input and output do not change, but the different intermediate operations can be selected) and that is adaptable to the screen size changes (i.e.: the workflow will have the same shape in both mobile devices and regular web-browsers). The main drawback inherent to this kind of representation lies in its limitations regarding the workflow complexity. The workflows

provided by Magallanes web service are, to some extent, linear, and with just one input and one output. This can be a problem when trying to represent workflows from other sources (v.g.: Taverna, Galaxy).

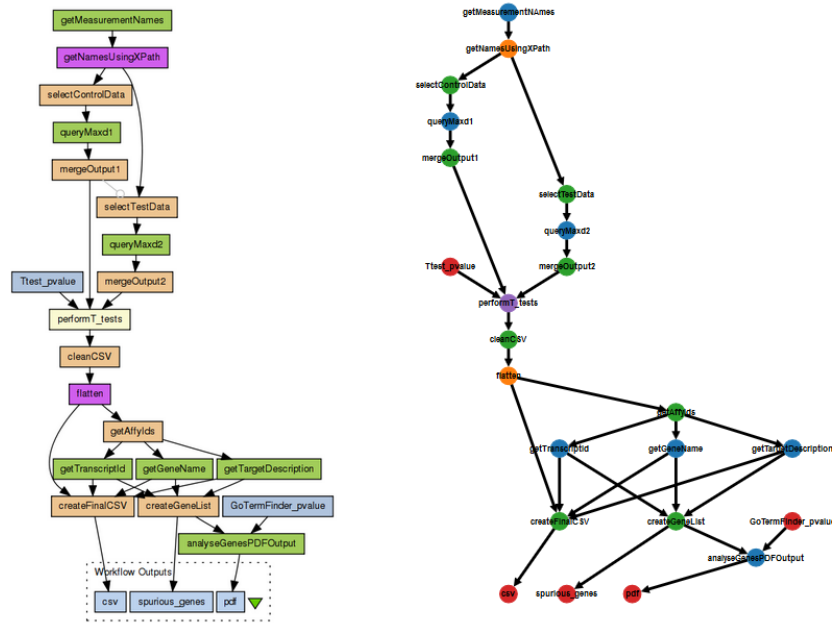


Fig. 2: On the left, (1) a workflow for the identification of differential genes generated as an image using Taverna. [17] On the right, (2) Same workflow represented as an interactive graph using web technologies in a mobile device.

Secondly, an adaptation from the classic way of representing a workflow (i.e.: boxes and arrows) is provided in Figure 2 (A). As can be seen, the boxes have been replaced by smaller bubbles. The idea behind this is to try to increase its adaptability for smaller-sized screens such as mobile devices ones, as well as to add the possibility to zoom in and out, and show/hide the elements names when zooming out (as the workflow complexity would increase). This interactivity allows the user not only to preview the workflow but to modify and adapt it before the execution. This way, and because the dynamically nature of mORCA's interfaces, customizable workflows for each user needs could be easily reproduced.

After considering the two ways mentioned, the first one was selected for the integration of Magallanes in mORCA, due to the linear nature of the workflows generated by Magallanes, as well as its adaptability to mORCA's current inter-

face. Moreover, this is the alternative in which mobile devices characteristics can be better exploited. Profound studies about acceptance are candidates for future works, in order to extend this debate. What cannot be denied is mobile devices have come to stay, and the scientific community has to reap their benefits.

## 5 Conclusions

Adapting services to mobile devices is a step that should be taken carefully due to the limitations of these tools in terms of diversity of screen sizes (smartphones, phablets, tablets), performance (significantly lower than a PC), and the actual change of interface paradigm (from WIMP - Windows, Icons, Mouse, Pointer to a touch screen). In this sense, and also given the complexity inherent to discovering not only services, but workflows and data types (among other resources) in a repositories system, it has been implemented and adapted an intelligent search engine that includes all the functionalities available in the original version, and also the linking of those services available at mORCA to the search results.

Thus, the functionalities of mORCA have been increased, including a way of discovering several types of resources, and also providing an easy-to-use mobile solution for workflows generation, an environment that has not been properly explored yet, despite mobile devices are all pervasive nowadays.

Three different methods of mobile representation have been studied to identify its strengths and limitations, in order to provide an optimal solution to the workflow representation problem.

Regarding future work, more research has to be carried out, as it has been said previously, in order to obtain information about which is the most-accepted way of representing workflows. Furthermore, as mobile environment is a development area under constant evolution, we must pay attention to technological advancements and improvements, in order to take advantage of them. Besides, due to the modular characteristics of mORCA, more functionalities can be added in the future, expanding its capabilities and providing the community with a richer tool.

## Acknowledgements

This work has been partially supported by the European project ELIXIR-EXCELERATE (grant no. 676559), the Spanish national projects Plataforma de Recursos Biomoleculares y Bioinformáticos (ISCIII-PT13.0001.0012) and RIRAAF (ISCIII-RD12/0013/0006) and the University of Málaga.



## References

1. A. Teufel, M. Krupp, A. Weinmann, and P.R. Galle. Current bioinformatics tools in genomic biomedical research (review). *International Journal of Molecular Medicine*, 17, pages 967–973, 2006.
2. Alexander Garcia Castro, Samuel Thoraval, Leyla J Garcia, and Mark A Ragan. Workflows in bioinformatics: meta-analysis and prototype implementation of a workflow generator. *BMC Bioinformatics*, 6(87), 2005.
3. Ebi: The european bioinformatics institute. <https://www.ebi.ac.uk/>. Last accessed Jan 2018.
4. Ncbi: National center for biotechnology information. <https://www.ncbi.nlm.nih.gov/>. Last accessed Jan 2018.
5. Inb: The spanish institute for bioinformatics. <http://www.inab.org/>. Last accessed Jan 2018.
6. Don Gilbert. Bioinformatics software resources. *Briefings in Bioinformatics*, 5(3):300–304, 2004.
7. Johan Karlsson and Oswaldo Trelles. Mapi: a software framework for distributed biomedical applications. *Journal of Biomedical Semantics*, 4(4), 2013.
8. Javier Ros, Johan Karlsson, and Oswaldo Trelles. Magallanes: a web services discovery and automatic workflow composition tool. *BMC Bioinformatics*, 10(334), 2009.
9. Enis Afgan, Dannon Baker, Marius van den Beek, Daniel Blankenberg, Dave Bouvier, Martin ech, John Chilton, Dave Clements, Nate Coraor, Carl Eberhard, Bjrn Grning, Aysam Guerler, Jennifer Hillman-Jackson, Greg Von Kuster, Eric Rasche, Nicola Soranzo, Nitesh Turaga, James Taylor, Anton Nekrutenko, and Jeremy Goecks. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(W1):W3–W10, 2016.
10. Derik Barseghian, Ilkay Altintas, Matthew B. Jones, Daniel Crawl, Nathan Potter, James Gallagher, Peter Cornillon, Mark Schildhauer, Elizabeth T. Borer, Eric W. Seabloom, and Parvies R. Hosseini. Workflows and extensions to the kepler scientific workflow system to support environmental sensor data access and analysis. *Ecological Informatics*, 5:42–50, 2010.
11. Katherine Wolstencroft, Robert Haines, Donal Fellows, Alan Williams, David Withers, Stuart Owen, Stian Soiland-Reyes, Ian Dunlop, Aleksandra Nenadic, Paul Fisher, Jiten Bhagat, Khalid Belhajjame, Finn Bacall, Alex Hardisty, Abraham Nieva de la Hidalga, Maria P. Balcazar Vargas, Shoaib Sufi, and Carole Goble. The taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud. *Nucleic Acids Research*, 41(W1):W557–W561, 2013.
12. Carole A. Goble, Jiten Bhagat, Sergejs Aleksejevs, Don Cruickshank, Danius Michaelides, David Newman, Mark Borkum, Sean Bechhofer, Marco Roos, Peter Li, and David De Roure. myexperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research*, 38:W677–W682, 2010.
13. Claus Brnich, Ivar Grytten, Eivind Hovig, Jonas Paulsen, Martin ech, and Geir Kjetil Sandve. Galaxy portal: interacting with the galaxy platform through mobile devices. *Bioinformatics*, 32(11):1743–1745, 2016.
14. Sergio Díaz del Pino, Juan Falgueras, Esteban Pérez-Wohlfeil, and Oswaldo Trelles. morca: Sailing bioinformatics world with mobile devices. *Bioinformatics*, 2017.

15. Erik Christensen, Francisco Curbera, Greg Meredith, Sanjiva Weerawarana, et al. Web services description language (wsdl) 1.1, 2001.
16. Victor Cheung, Jens Heydekorn, Stacey Scott, and Raimund Dachsel. Revisiting hovering: Interaction guides for interactive surfaces. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, ITS '12*, pages 355–358, New York, NY, USA, 2012. ACM.
17. P. Li, J. I. Castrillo, G. Velarde, I. Wassink, S. Soiland-Reyes, S. Owen, D. Withers, T. Oinn, M.R. Pocock, C.A. Goble, S.G. Oliver, and D.B. Kell. Performing statistical analyses on quantitative data in taverna workflows: an example using r and maxdbrowse to identify differentially expressed genes from microarray data. *BMC Bioinformatics*, 9(334):1743–1745, 2008.