

Extensiones a DEHIA en el marco de una arquitectura distribuida con microservicios

Jose Arcidiacono ¹, Alejandra B. Lliteras ^{2,3}, Patricia Bazán ¹

¹ UNLP, Facultad de Informática, LINTI ² UNLP, Facultad de Informática, LIFIA, ³ CICPBA

jarcidiacono@linti.unlp.edu.ar, alejandra.lliteras@lifia.info.unlp.edu.ar,
pbaz@info.unlp.edu.ar

RESUMEN

DEHIA es una plataforma para la creación y ejecución de actividades de recolección de datos que requieren de la intervención humana para realizar este objetivo.

La plataforma está definida sobre una arquitectura distribuida, basada en microservicios y ya cuenta con revisiones en torno a la distribución por componentes en contenedores independientes que aportan a la interoperabilidad pero que aumenta la complejidad operacional y requiere automatización y monitoreo.

En este trabajo, se presentan dos líneas de trabajo que extienden la plataforma en torno a la recolección de datos: 1- realizar la recolección ex-situ y 2- definir un componente de visualización de los datos recolectados.

Palabras clave: Microservicios, Recolección de datos, Interoperabilidad

CONTEXTO

Este proyecto surge como extensión del proyecto “DEHIA: Una plataforma liviana para definir y ejecutar actividades con intervención humana basada en workflows” [Arcidiacono, 2020], en el cual se presenta una aplicación web que permite a usuarios finales definir actividades que requieran de

la intervención humana con el fin de realizar recolección y análisis de datos, así como una aplicación móvil para ejecutar dichas actividades.

Dado que esta plataforma pone el foco en el análisis de los datos, además de la recolección de los mismos, resulta de interés poder completar parte de la arquitectura de microservicios en la que se basa DEHIA con un visualizador que permita un análisis integral, eficaz y flexible, siguiendo la estructura y filosofía de DEHIA.

Por otra parte, DEHIA realiza recolección de datos mediante una aplicación móvil, que permite obtener los datos in situ y tanto de manera online como offline (para enviar los datos recolectados cuando se logre conectividad online).

Así, un cliente web que permitiera recolectar datos ex-situ (es decir fuera del lugar donde se recolectaron) para DEHIA, daría la posibilidad de resolver los workflows de forma online y sin necesidad de tener instalada la aplicación de forma local. Aunque limita a que los usuarios tomen datos en un medio alternativo (por ejemplo, planillas de papel u anotaciones libres) y luego los carguen ya no in-situ.

1- INTRODUCCIÓN

Las arquitecturas orientadas a servicios (SOA - Services Oriented Architecture) tienen como objetivo la reutilización de componentes funcionales autónomos que puedan interoperar, componerse, integrarse y escalar para resolver funcionalidades más complejas.

DEHIA adopta una arquitectura basada en *micro*-servicios, un estilo de arquitectura que simplifica los contratos entre dichos servicios favoreciendo su composición donde cada uno de ellos ofrece una interfaz que permite la integración de nuevos componentes.

Dentro de la arquitectura de DEHIA hay tres componentes que representan la extensión planteada en este trabajo y son: el proveedor de identidad y autorización, el recolector de datos y el gestor de workflows.

DEHIA permite recolectar datos desde su aplicación móvil, en la cual los usuarios ejecutan los workflows de actividades mientras recorren de manera in-situ un lugar. Sin embargo, es posible que ciertos dominios requieran que los datos no fueran recolectados desde una aplicación móvil, o bien que los usuarios que recolectan los datos no quieran dejar prácticas arraigadas y bien probadas como la recolección de datos en planillas de papel. En tal caso, el foco principal, es lograr que los datos sean recolectados de alguna manera y posteriormente cargados de forma homogénea para su posterior uso. Para este caso en particular lo propuesto hasta el momento en DEHIA es insuficiente.

Adicionalmente, una vez que los datos son recolectados y almacenados, surge la necesidad de que los mismos puedan ser presentados de manera comprensible para las personas. Desde esta perspectiva, la

visualización de datos mediante gráficos ayuda a la comunicación y permite revelar información que los modelos y simulaciones pueden perder [Unwin, 2020], por ejemplo, valores atípicos y patrones locales [Unwin, 2020]. La visualización propuesta en DEHIA no incluye actualmente visualización de los datos mediante gráficos: sólo propone la visualización de datos en forma de tabla.

Se sabe que en la actualidad contar con datos es de suma importancia para diferentes acciones, por ejemplo, compararlos, detectar tendencias y mostrar información [Faronius Hofmann & Håkansson, 2021].

Hasta el momento se cuenta con una herramienta de recolección de datos (por parte de personas) y un mecanismo de almacenamiento apropiado para el uso que se le da actualmente a la plataforma. Sin embargo, sería interesante considerar un uso más extendido que llevaría al desafío de capturar, almacenar, manipular y mostrar grandes volúmenes de datos.

2- LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La extensión a DEHIA tiene por objetivo profundizar algunas de las líneas de investigación que la plataforma abarca. En particular, la extensión presentada en este trabajo se enfoca en las áreas, como 1- Integración de Componentes Distribuidos, 2- Interacción Humano- Computador, 3 - Usabilidad y Accesibilidad tanto web como móvil, 4- Recolección de datos 5- Visualización.

A partir del proyecto de investigación presentado en [Lamoureux, Z et al., 2019], los autores proponen criterios/aspectos a tener en cuenta a la hora de analizar

plataformas de Ciencia Ciudadana para la recolección de datos y realizan un benchmarking de algunas aplicaciones en dicho dominio, como por ejemplo, Zooniverse [Simpson, R. et al., 2014] y [Cox, J. et al., 2015] así como EpiCollect [Aanensen, D. M. et al., 2009] y [Gohil, R. et al., 2020] se observa la necesidad de proponer este cliente web ex-situ, con la intención de extender el alcance a más usuarios al momento de recolectar datos en DEHIA.

Por otra parte, la visualización de los datos recolectados y el conocimiento generado por las personas, permite traducir los datos en gráficos y elementos visuales que facilitan a los usuarios finales identificar patrones, extraer conocimiento significativo y distinguir las perspectivas que dicha información presenta. Así es como, a través de una adecuada visualización, el usuario final encuentra una manera accesible de interactuar con los datos y generar un posterior análisis, característica muy importante para construir una comunicación eficiente y concreta de los resultados obtenidos y el valor de los mismos. Lograr dichas visualizaciones puede resultar una tarea difícil para aquellas personas sin conocimientos específicos de este proceso y, particularmente, de las herramientas tecnológicas disponibles, es por eso que este componente de visualización se propone acortar la brecha entre la representación de los datos y sus usuarios finales.

La Figura 1 muestra la integración entre la plataforma DEHIA actual y las diferentes componentes relativas a cada línea de investigación, diferenciando en la misma aquellas líneas de trabajo que se encuentran en progreso de otras posibles.

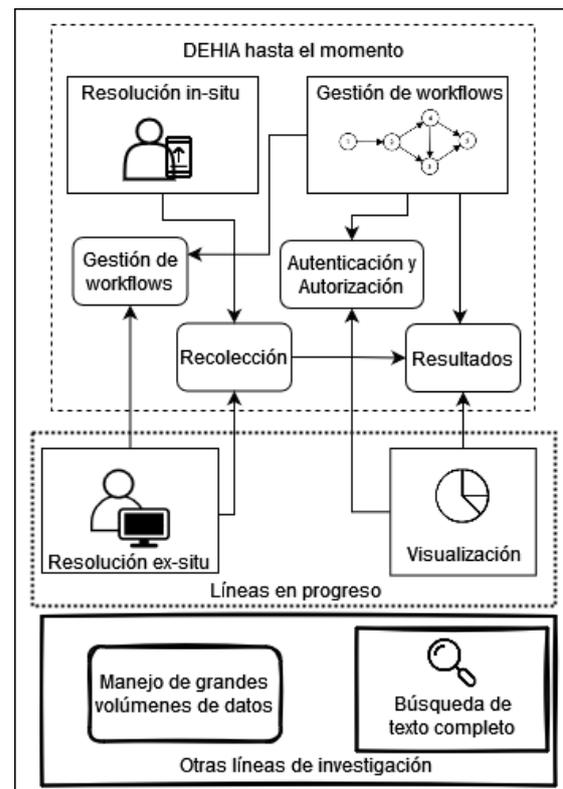


Figura 1 - Componentes de DEHIA a lo largo de las líneas de investigación

En la Figura 1 pueden observarse los componentes cliente – en este caso los que tienen interfaz de usuario – como rectángulos con línea llena, mientras que los servicios se encuadran en rectángulos con los bordes redondeados. Las flechas indican tanto la dependencia entre los componentes como la comunicación que sucede entre ellos a través de HTTP.

3- RESULTADOS ESPERADOS/OBTENIDOS

Dentro de los resultados esperados se encuentra el desarrollo e integración a la arquitectura actual de microservicios de DEHIA de un componente de visualización que permita el enriquecimiento de los datos recolectados mediante la plataforma y el análisis de los mismos a través de una variedad de gráficos y representaciones. Uno de los objetivos de este desarrollo en

particular es facilitar a usuarios finales el acceso a la información producto del proceso de recolección de datos con intervención humana, por lo cual se dio importancia a la facilidad de uso de la aplicación.

Finalmente, se espera lograr también otra extensión en forma de sitio web para la resolución ex-situ de las actividades creadas desde DEHIA, que esté integrado a la arquitectura y permita el envío de los datos cargados desde el mismo hacia la plataforma.

4- FORMACIÓN DE RECURSOS HUMANOS

La diversidad de líneas de investigación enunciadas, así como la gran variedad de aplicación de los resultados esperados, propone un foco de investigación claro y concreto para la formación de recursos humanos, que encuentran en esta propuesta diversas aristas de trabajo.

La propuesta aborda temas de arquitecturas distribuidas, ingeniería de software, HCI, accesibilidad y usabilidad web, así como aspectos relacionados a herramientas de autor, test cases, workflow con intervención humana, recolección, y visualización de datos.

Adicionalmente, la extensión vinculada con la visualización de datos será aplicada a un caso de estudio con perspectiva de género a partir del trabajo conjunto con la Prosecretaría de Derechos Humanos de la Universidad Nacional de La Plata.¹

El presente trabajo describe la línea de investigación que ha dado lugar a la formulación de propuestas de tesinas de grado por cada una de las extensiones, así como otra propuesta de tesina futura en

torno a la integración de un nuevo componente que incorpora un motor de búsqueda de texto completo.

5- BIBLIOGRAFÍA

[Aanensen, D. M. et al., 2009] Aanensen, D. M., Huntley, D. M., Feil, E. J., al-Own, F. A., & Spratt, B. G. (2009). EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. *PloS one*, 4(9), e6968.

[Arcidiacono, 2020] Arcidiacono, J. (2020). DEHIA: una plataforma liviana para definir y ejecutar actividades con intervención humana basadas en workflows. Tesina de Grado. Facultad de Informática, UNLP.

[Cox, J. et al., 2015] Cox, J., Oh, E. Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A., ... & Holmes, K. (2015). Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering*, 17(4), 28-41.

[Faronius Hofmann & Håkansson, 2021] Therese Faronius Hofmann and Linda Håkansson. (2021). Visualization Design Effects on Credibility and Data Perception, and the Importance of Digital Interaction. Uppsala universitet

[Gohil, R. et al., 2020] Gohil, R., Sharma, S., Sachdeva, S., Gupta, S., & Dhillon, M. S. (2020). EpiCollect 5: A Free, Fully Customizable Mobile-Based Application for Data Collection in Clinical Research.

[Lamoureux, Z et al., 2019] Lamoureux, Z., & Fast, V. (2019). The tools of citizen science: An evaluation of map-based crowdsourcing platforms. *Spatial Knowledge and Information Canada*, 7(4), 1.

¹ <https://unlp.edu.ar/ddhh>

[Simpson, R. et al., 2014] Simpson, R., Page, K. R., & De Roure, D. (2014, April). Zooniverse: observing the world's largest citizen science platform. In Proceedings of the 23rd international conference on world wide web (pp. 1049-1054).

[Unwin, 2020] Unwin, A. (2020). Why is Data Visualization Important? What is Important in Data Visualization?. Harvard Data Science Review, 2(1).