

SISGEN-HIV - Sistema Informático de Secuencias Genéticas de HIV

Paula Aulicino¹, Agustina Bazzano², Ignacio Saporiti², German Biagioli², Andrea Mangano¹, Luisa Sen¹, Susana Rodríguez³, Fernando Sanz⁴, Federico Raggi⁴, Rubén Aspres⁴, Gustavo Carolo⁴

¹ Laboratorio de Biología Celular y Retrovirus, Hospital de Pediatría J.P. Garrahan
{paulicino, amangano, lsen}@garrahan.gov.ar

² Flux IT

{agustina.bazzano, ignacio.saporiti,
german.biagioli}@fluxit.com.ar

³ Dirección Asociada de Docencia e Investigación, Hospital de Pediatría J.P. Garrahan
{sprodiguez}@garrahan.gov.ar

⁴ Gerencia de Sistemas, Hospital de Pediatría J.P. Garrahan

{fsanz, fraggi, raspres, gcarolo}@garrahan.gov.ar

Resumen. SISGEN-HIV es la primera herramienta bioinformática para el análisis conjunto de datos clínicos y genéticos aplicada al ámbito hospitalario. Fue construida por un equipo interdisciplinario de especialistas en biología molecular y en informática, que tomaron el desafío de unir distintos conocimientos y experiencias para mejorar la calidad del seguimiento de los niños con HIV.

Abstract. SISGEN-HIV is the first bioinformatics tool applied to the hospital setting for joint analysis of clinical and genetic data. It was built by an interdisciplinary team of specialists in molecular biology and computer science, who took the challenge of joining different knowledge and experiences to improve the quality of monitoring of children with HIV.

1 Introducción

La idea de SISGEN-HIV surgió en el Laboratorio de Biología Celular y Retrovirus del Hospital Garrahan para el estudio de la infección por HIV-1, como respuesta a la necesidad de registrar y almacenar datos genéticos y secuencias de ADN para su fácil interpretación y vinculación a parámetros clínicos de evolución y tratamiento de los pacientes en seguimiento infectológico en el Hospital Garrahan.

Es un ejemplo de construcción interdisciplinaria entre el lenguaje de la Biología Molecular y la Informática que ayuda a mejorar la calidad de la atención en salud.

En este trabajo se cuenta la experiencia del desarrollo de esta herramienta y los factores que llevaron al éxito del proyecto, haciendo foco en la dinámica de

colaboración entre una entidad pública con una empresa privada para generar una solución de vanguardia para el tratamiento de HIV en niños.

El desarrollo de la herramienta fue realizado en forma conjunta entre los investigadores del Laboratorio, los especialistas de Flux IT y el Servicio de Sistemas del Hospital, con el apoyo de la Dirección Asociada de Docencia e Investigación.

2 Problemática

En el Laboratorio de Biología Celular y Retrovirus se estudia la infección por el virus HIV en niños [1]. Una de las características del HIV es su alta tasa de mutación, lo que genera una gran diversidad genética. Aún dentro del mismo individuo existen miles de "cuasiespecies" virales que van evolucionando y acumulando mutaciones para evadir la respuesta inmune del individuo y la acción de las drogas antirretrovirales.

Para estudiar los cambios en el genoma viral, se obtienen las secuencias de ADN del HIV a partir de las muestras de sangre de los individuos infectados, que luego deben ser interpretadas con programas de análisis específicos que traducen la secuencia de ADN a secuencia de proteínas [2], o que aplican modelos evolutivos para predecir el impacto de los cambios en la progresión de la enfermedad y la respuesta a los tratamientos [3].

Dado que las secuencias de ADN del HIV están compuestas únicamente por la combinación de 4 letras (A, C, T, G) y pueden tener una longitud de hasta 9000 caracteres (este es el largo total del genoma del HIV), se requería un almacenamiento en un formato adecuado (archivos .fas) y en forma disociada de los datos del individuo que le dieron origen. Debido al número creciente de secuencias, a la necesidad de almacenarlas de forma tal de lograr la fácil interpretación y asociarlas con datos clínicos y bioquímicos disponibles en la historia clínica del paciente y con otros datos genéticos y de laboratorio que podrían influir en la evolución de la enfermedad, es que surge la necesidad de crear la herramienta bioinformática SISGEN-HIV.

La generación de datos genéticos en forma de secuencias de ADN, polimorfismos de nucleótido simple (SNPs) o variaciones alélicas es un fenómeno cada vez más frecuente en la medicina. Sin embargo, no se encontraba una manera práctica y útil de almacenar estos datos en forma vinculada a los datos clínicos, de tratamiento y otros que son registrados como parte de la historia clínica del paciente.

Cabe aclarar que la información contenida en las secuencias de ADN es útil únicamente luego de utilizar procesos de análisis que facilitan su interpretación, es decir, se transforman los datos de las secuencias en información que luego con la interpretación se convierten en conocimiento.

Si bien existen diferentes herramientas en bases de datos genéticas internacionales para el análisis de secuencias de ADN del virus HIV-1 y de polimorfismos presentes

en genes humanos relacionados con la inmunopatogenia de la infección viral, las necesidades que se tenían en el laboratorio eran bien específicas. Se trataba no sólo de almacenar y analizar los datos genéticos en sí, sino también informatizar puntualmente el proceso de trabajo de las doctoras en el laboratorio, que hasta el momento gestionaban con planillas Excel y base Access (con sus limitaciones asociadas). Dadas estas condiciones, se decidió hacer el desarrollo de un sistema a medida, que permitiera además tener una fuerte comunicación con el sistema propio del hospital y poder aprovechar los datos de la historia clínica informatizada de los pacientes.

Uno de los mayores usos que se le da a los datos obtenidos de la secuencia de ADN del virus y de las consultas a bases de datos externas, es para realizar un estudio genotípico de resistencia a los antirretrovirales. A partir de la secuencia y de acuerdo a la información reportada en las bases de datos públicas [3], se determina si el HIV del paciente posee mutaciones de resistencia a las drogas antirretrovirales, con el fin de guiar la elección de los tratamientos en los pacientes infectados. Este informe es complejo, porque además de requerir la interpretación de las secuencias con herramientas disponibles on-line, el reporte de las mutaciones observadas debe hacerse manualmente en el sistema de resultados de laboratorios del Hospital Garrahan, tarea no sólo tediosa, sino con riesgo de cometer errores durante el pasaje manual.

3 Desafíos

Más allá de los desafíos técnicos, uno de los mayores logros fue alcanzar un entendimiento mutuo y generar la relación de confianza necesaria para lograr el éxito del proyecto. La inserción de una empresa privada en una institución pública con fuerte orientación científica no siempre es bien vista, posiblemente fundado en el desconocimiento o en experiencias pasadas.

Por el lado informático, la clave fue lograr entender, desde un punto de vista biológico, las necesidades y la información que maneja el personal del laboratorio en su trabajo diario y cómo funciona el hospital en torno al manejo de estos datos. Cómo interactúan las diferentes áreas, qué información es compartida, cuál es propia del laboratorio, así como también estudiar los conceptos de genética y biología molecular y clínicos referidos tanto al HIV como a los pacientes.

Por el lado del laboratorio, el desafío fue entender y aprender cómo se construye una herramienta de software. Aprender a describir requerimientos, imaginar una aplicación que no se conoce aún, tener en cuenta las diferencias que tiene un sistema a medida con respecto a las planillas que se utilizaban y cómo eso impacta en los procesos y trabajos diarios, cuáles son las cosas que se pueden mejorar del proceso actual, entender cómo priorizar.

Resumiendo, el gran desafío de este proyecto fue lograr una comunicación efectiva, porque solamente con una comunicación clara, transparente y bidireccional se puede

realizar la transferencia de conocimiento necesaria para encarar exitosamente un proyecto de esta índole.

4 Marco de Trabajo

En un contexto de proyecto donde el usuario a priori no tenía totalmente definido lo que necesitaba y donde el equipo de desarrollo desconocía el modelo de dominio sobre el cual debería trabajar, las metodologías ágiles [4] brindaron el marco adecuado para poder sobrellevar la incertidumbre inicialmente.

El puntapié inicial fue la necesidad concreta de almacenar datos genéticos obtenidos del virus de los pacientes. A partir de ahí se inició una primera etapa de “Análisis, conceptualización y prototipado inicial”.

Ya desde esta etapa se entendió que trabajar iterativamente era la manera fundamental para avanzar. Por lo cual se plantearon, desde el inicio, reuniones periódicas cada quince días que se utilizaron para realizar un relevamiento totalmente iterativo [Fig.1]

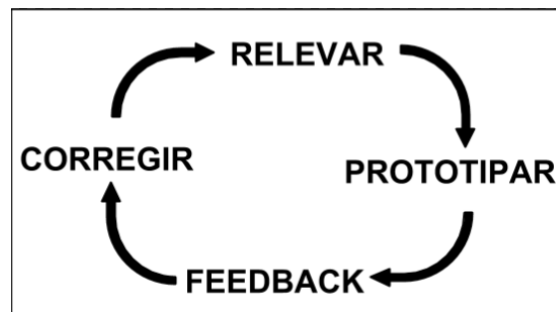


Fig. 1. Relevamiento iterativo

Esto permitió empezar a acortar las distancias propias de hablar distintos lenguajes. Mediante el prototipado [5] de algunas funcionalidades se facilitó, por un lado, el proceso de aprendizaje de la biología y genética del virus HIV y por el otro, se empezó a vislumbrar el alcance de lo que se quería construir y a tener entregables materializados, aunque sea en papel, para poder hablar sobre temas más concretos.

Esta etapa, que generó una activa ida y vuelta de parte de ambos grupos, duró unos 4 meses. En ella se realizaron distintas actividades de relevamiento, entendimiento de los perfiles de los usuarios [6] [Fig.2] (con la colaboración y participación fundamental de analistas de usabilidad y experiencia de usuario [7]), mocks de baja fidelidad [5], diagramas de entidad-relación, entre otros.

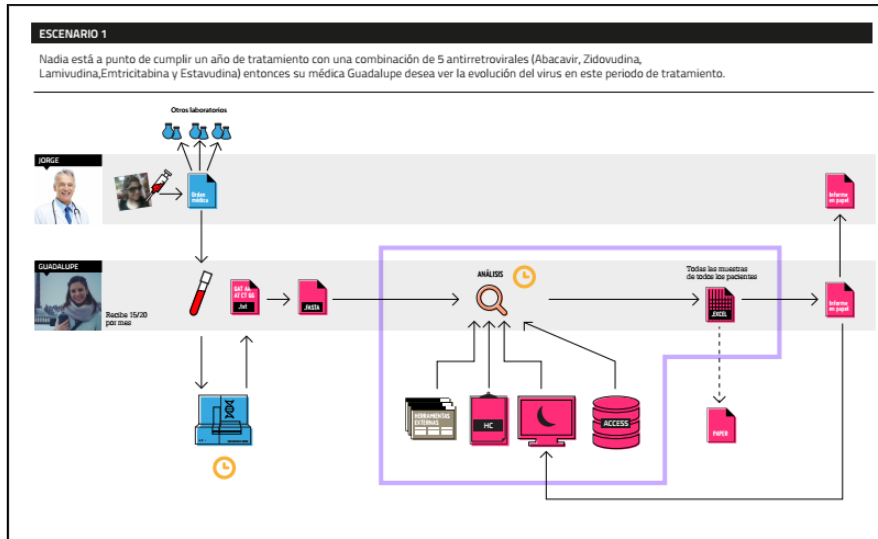


Fig.2. Análisis de escenarios según perfiles de usuarios.

A medida que se fue avanzando en el relevamiento, se fue armando un Visual Story Map [8] con los usuarios lo que permitió visualizar y acordar el alcance de la aplicación [Fig.3].



Fig. 3. Story Map a alto nivel

Y finalmente, se pudo definir claramente el objetivo del proyecto: “Construir una herramienta que permita vincular la información de datos clínicos con datos de

secuencias, datos de tratamientos y de SNPs que hoy día se encuentra dispersa en diferentes medios. Que también permita cargar y hacer uso de esa información mediante la realización búsquedas por distintos criterios y la visualización amigable de los resultados.”

Este análisis y conceptualización que generalmente en las metodologías ágiles se llama Sprint (o iteración) fue necesario para preparar el proyecto antes de comenzar el desarrollo: permitió ir aprendiendo la complejidad del dominio, generar confianza y sentido de equipo colaborativo, atributos necesarios para avanzar con el desarrollo.

Una vez finalizada la primera etapa, se comenzó a trabajar más formalmente con SCRUM [9], partiendo de un Product Backlog construido desde el Story Map y realizando las ceremonias propuestas por este framework.

Antes de cada Sprint (iteración) se planificaba con los usuarios cuáles funcionalidades se incluirían en la próxima entrega, siempre intentando priorizar aquellas que entregasen más valor.

Durante cada sprint, se mantenía el contacto para reportar avances, impedimentos o despejar dudas, dando una visibilidad total en todo momento del estado del proyecto.

En la reunión de revisión (Sprint Review) realizada siempre en el Hospital con todos los interesados, los usuarios podían interactuar y probar las nuevas funcionalidades agregadas para darlas como finalizadas o sugerir cambios o correcciones, identificar los puntos débiles o que no habían quedado claros según sucedía. Y con todo este feedback, planificar la siguiente iteración.

5 Solución

El diseño y arquitectura de la información [10][Fig.4] permitió que a partir de 5 módulos aparentemente independientes entre sí (Secuencias, SNPs, Tratamientos, Laboratorio y Pacientes), cada usuario del laboratorio pueda realizar sus investigaciones y seguimientos según su enfoque, pero ahora con la posibilidad de poder cruzar cualquier tipo de información mediante las búsquedas combinadas que SISGEN provee para crear nuevas líneas de investigación y empezar a responder preguntas que antes era impensable encarar, por ejemplo recuperar datos genéticos pero basados en criterios epidemiológicos, de tratamiento y/o de evolución clínica. De esta manera también se puede realizar un seguimiento completo del paciente combinando distintos enfoques.

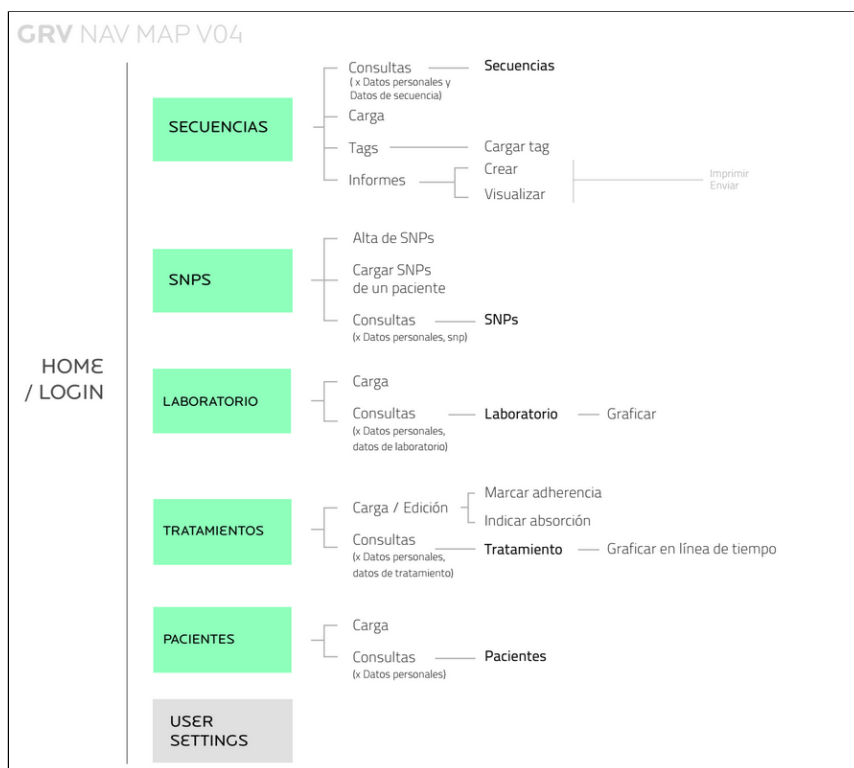


Fig.4. Arquitectura de la información

Entonces, SISGEN-HIV no sólo soluciona el almacenamiento de los datos genéticos en un formato de fácil interpretación sino que en realidad se trata de una base de datos unificada donde conviven:

- secuencias genéticas del virus,
- datos asociados a las muestras,
- información “complementaria” proveniente de BBDD públicas,
- datos genéticos propios del paciente (Polimorfismos de Nucleótido Simple, estudios cuantitativos, haplotipos y estudios sobre gen HLA),
- datos filiatorios (del Sistema del Hospital),
- datos de historia clínica:
 - diagnóstico,
 - perinatólogicos,

- tratamientos,
- resultados de laboratorio.

Permite el acceso a toda la información, utilizando todos los datos disponibles y provenientes de distintas fuentes para generar información útil y obtener resultados más fácil y rápidamente.

Otra ventaja lograda con esta solución radica en la simplicidad ganada para realizar los Informes de Resistencia, ya que al cargar una secuencia de nucleótidos, SISGEN-HIV automáticamente recupera los datos de mutaciones e inhibidores [3], arma el template del informe de resistencia con esta información, y con un click se puede enviar directamente al Sistema del Hospital o descargar como documento. Esto reduce los tiempos de operatoria prácticamente a la mitad y minimiza los errores asociados al pasaje manual de datos.

SISGEN-HIV además preserva la confidencialidad de los datos de los pacientes que reciben atención en el Garrahan, ya que se encuentra instalado en el propio servidor del hospital.

La innovación estuvo basada en la interconexión de datos clínicos, genéticos y moleculares para su análisis conjunto en el estudio de una enfermedad particular. Esto será de utilidad en la toma de decisiones médicas y en el asesoramiento familiar, como así también en la búsqueda de factores inmunogenéticos, o fármaco-genéticos relacionados con diferencias inter-individuales en el pronóstico clínico y tratamiento de la enfermedad.

La primera versión de SISGEN-HIV es completamente funcional para la carga de secuencias y datos de pacientes. Tiene una comunicación bidireccional con la base de datos del hospital, y agilizó los procesos reduciendo el tiempo de los mismos a menos de la mitad. El sistema se utiliza en el trabajo cotidiano del laboratorio, y está reemplazando las bases de datos como excel y access anteriores, permitiendo cargar los resultados de los estudios realizado directamente en el sistema.

Se espera que la aplicación comience a ser utilizada por otros laboratorios, siempre dentro del ámbito mismo del hospital, debido que es una implementación hecha a medida con las particularidades propias de la institución. No obstante, bien se puede replicar la experiencia y forma de trabajo lograda y aprendida en otros centros de salud.

6 Conclusiones

El trabajo colaborativo derribó mitos que existen acerca del trabajo en conjunto entre lo público y privado. Se evidencia, a partir de estas experiencias, un claro ejemplo de articulación exitosa entre un hospital público, instituciones del sistema científico nacional y la industria. Es un camino a seguir que permite realizar una óptima transferencia de conocimiento y tecnología a través de un proyecto conjunto de innovación, formando equipos interdisciplinarios que potencien las actividades entre la industria y el sector científico-tecnológico, y de esta forma poder generar un

impacto real en la sociedad. La colaboración permitió la posibilidad de construcción interdisciplinaria entre el lenguaje de la biología molecular y la informática, para mejorar la calidad de atención en salud.

SISGEN-HIV es un producto de alto nivel innovador, que genera valor agregado ayudando a los investigadores a mejorar y optimizar su trabajo. Es una herramienta que posibilita el almacenamiento de datos en forma segura y eficiente, recurso imprescindible para la generación de información válida para la construcción de hipótesis de investigación y nuevos conocimientos.

Proyectos como éste contribuyen a reafirmar el posicionamiento del Hospital como referente a nivel nacional, generando un impacto positivo en la sociedad a través de la mejora en la calidad de atención por la aplicación de soluciones en el área de salud.

Referencias

- [1] Carballal, Guadalupe y Oubiña, José. *Virología médica*. Buenos Aires: El Ateneo, 1998. Print.
- [2] “HIV Sequence Locator.” *Sequence Locator Tool*. Web. 17 Jun. 2015.
<<http://www.hiv.lanl.gov/content/sequence/locate/locate.html>>
- [3] Shafer, Robert W. “Rationale And Uses of a Public HIV Drug-Resistance Database.” *The Journal of Infectious Diseases J INFECT DIS* 194.s1 (2006): n. pag. Web.
- [4] Rasmusson, Jonathan. *The Agile Samurai: How Agile Masters Deliver Great Software*. Raleigh, NC: The Pragmatic Bookshelf, 2010. Print.
- [5] Buxton, William, Sheelagh Carpendale, and Saul Greenberg. *Sketching User Experiences: The Workbook*. San Francisco, CA: Morgan Kaufmann Publishers In, 2011. Print.
- [6] Cooper, Alan. *About Face: the Essentials of User Interface Design*. Foster City, CA: IDG Books Worldwide, 1995. Print.
- [7] Garrett, Jesse James. *The Elements of User Experience: User-Centered Design for the Web*. Indianapolis, IN: New Riders, 2002. Print.
- [8] Patton, Jeff et al. *User Story Mapping: Discover the Whole Story, Build the Right Product*. Print.
- [9] Kniberg, Henrik. *Scrum And Xp from the Trenches: How We Do Scrum*. S.l.: C4Media Inc., 2007. Print.
- [10] Rosenfeld, Louis, and Peter Morville. *Information Architecture for the World Wide Web*. Beijing: O'Reilly, 2002. Print.