


La gestión de los datos, un proceso esencial para el desarrollo del Biobanco

Data management, an essential process for the development of Biobanks

María L. Luna-González², Elizabeth Guío-Mahecha¹, Silvia Becerra-Bayona², Norma Serrano-Díaz^{1,2}

Forma de citar: Luna-González ML, Guío-Mahecha E, Becerra-Bayona S, Serrano-Díaz N. La gestión de datos, un proceso esencial para el desarrollo del Biobanco. Rev Univ Ind Santander Salud. 2016; 48(4): 548-553. DOI: <http://dx.doi.org/10.18273/revsal.v48n4-2016013> 

RESUMEN

El proceso de recolección de material biológico humano, necesita cada vez más la organización de la información para la gestión de las muestras y la integración de los datos relacionados con el participante o donante, provenientes de diferentes fuentes. Con el avance de la computación, esto permitirá identificar posibles interacciones sociodemográficas, genéticas, ambientales, entre otras con determinada enfermedad. Las estructuras de datos, los sistemas de codificación y los sistemas de metadatos, se han convertido en un desafío para la organización de los biobancos. La gestión, integración, seguridad, privacidad y análisis de los datos, son retos importantes para los investigadores y la informática. La normalización de los datos, la armonización e interoperabilidad de sistemas informáticos de biobancos permitirán el óptimo uso del material biológico, convirtiéndose en no solo un gran recurso para estudios epidemiológicos y clínicos a gran escala, sino también en bases para nuevas pruebas de diagnóstico e intervenciones terapéuticas personalizadas.

Palabras clave: Biobanco, sistemas de información, material biológico, gestión de datos.

ABSTRACT

The collection, processing and storage of biological samples need a system for not only organizing and managing the patient samples but also integrating data records from different sources related to these patients. Along with computer advancement, these integration processes will allow to identify possible relationships between sociodemographic, genetic and environmental factors with specific diseases. Therefore, data structures, coding and metadata systems, have become essential elements for controlling biobanks. In fact, management, integration, security, privacy and data analysis are current challenges for scientists and computer administrators. The standardization of data, harmonization and interoperability of biobank computer systems will help to have an optimum use of biological material. As a result, these advances will turn into a great resource for large-scale epidemiological and clinical studies as well as the basis for new diagnostic tests and personalized therapies.

Keywords: Biobank, data systems, biological samples, data management.

1. Fundación Cardiovascular de Colombia FCV. Bucaramanga, Colombia

2. Universidad Autónoma de Bucaramanga UNAB. Bucaramanga, Colombia

Correspondencia: Norma C. Serrano Díaz. Dirección: Calle 155A No. 23-58. Correo electrónico: normaserrano@fcv.org. Teléfono: (7) 6396767 Ext: 214, Móvil: 310-2740099.

INTRODUCCIÓN

Los biobancos con fines de investigación biomédica, como entidades públicas o privadas, sin ánimo de lucro, recopilan muestras biológicas humanas con fines de investigación, bajo estándares de calidad y principios éticos y legales que pretenden garantizar la calidad del material almacenado y los derechos de los donantes¹⁻⁴.

Los biobancos constan de dos componentes: el almacenamiento y gestión de material biológico (recolección de la información de soporte que acompaña a las muestras) y la información asociada a los donantes; el garantizar la buena trazabilidad de esta información, acredita el material biológico⁵.

El auge y la necesidad de los biobancos es tal, que diferentes regiones del mundo vienen realizando esfuerzos para organizar infraestructuras en red, que permitan mayores oportunidades para la investigación biomédica⁶⁻⁸. Las tecnologías de la información y comunicación (TIC) se constituyen entonces en la herramienta clave para soportar los procedimientos operativos estandarizados que puedan certificar la calidad del material biológico humano (MBH), y permitan además la interacción de grupos de investigación y países aliados, con autonomía del fin con la que se conserve el MBH recopilado⁶⁻⁹.

La armonización de muestras, datos almacenados y el compartir de los mismos, ha sido uno de los puntos más importantes para la interoperabilidad, sustentabilidad y calidad de las redes de biobanco, con proyección a sistemas bioinformáticos centralizados para su operación, incluyendo fundamentos éticos y legales para el intercambio de datos^{6,9,10}.

Armonización y Estandarización de Datos

La normalización, en su definición misma, permite ajustar elementos semejantes a un tipo común¹⁰, tal como lo pretende la *International Organization for Standardization* (ISO), mediante normas que garanticen productos o servicios fiables y de buena calidad¹².

En el área de los sistemas informáticos, desde el origen de las bases de datos (BD), inicialmente bajo un modelo de relación entre tablas de datos, útil para aplicaciones de gestión, la normalización de las BD es regla indispensable, para dar la seguridad que no existirán elementos repetidos, dependencias entre campos identificadores (denominados llaves) e independencia entre los demás atributos¹³. Esto conlleva a establecer requisitos para los sistemas de información, necesidad

aplicable por tanto, para el tema en mención de biobancos.

No existen requisitos para infraestructura de TIC para biobancos, algunos cuentan con plataformas en las cuales, a través de intranet, se dispone de un soporte informático¹⁴ y otras aplicaciones de software han sido implementadas, como sistemas para la gestión de MBH y datos asociados para biobancos^{7,16}.

Existen datos básicos mínimos de los participantes o donantes del MBH que se deben contemplar para hacer parte de redes de biobancos: edad, género y patología, y otros referentes al material como tipo de muestra, temperaturas de procesos y almacenamiento. Sin embargo, a nivel de la Unión Europea, se estableció el conjunto de datos MIABIS (*Minimum Information About Biobank data Sharing*) con 52 atributos, que pretenden facilitar la armonización de datos para biobancos y que es utilizado por BBMRI (*Biobanking and Biomolecular Resources Research Infrastructure*)¹⁷.

Asociar los datos de los pacientes, con la información de los biobancos, requiere de un conjunto de tecnologías de información, que administre los flujos y transferencia automática de información, entre las diferentes fuentes, con anonimización de los datos de los participantes, de tal manera que no sea posible la vinculación directa entre unos y otros¹⁸.

Por otra parte, la estandarización y nomenclatura en la terminología clínica y de laboratorio, deben incluirse en el desarrollo de aplicativos de software para biobancos, como contribución para la conexión a bases de datos, la interoperabilidad y la asignación de datos clínicos, biológicos y genéticos desde distintas fuentes, en las diferentes plataformas que hacen uso de ellos^{19,20}.

Para facilitar la búsqueda de MBH, datos asociados y contribuir con la armonización, fue desarrollada Sample avAILability system (SAIL), aplicación web, con licencia open source (<https://github.com/sail>), que administra dos conjuntos de datos: uno para el vocabulario o metadato (términos médicos) y otro para la indexación de muestras. SAIL, evalúa la disponibilidad de datos, sin tener en cuenta su contenido, esto permite la localización e incorporación de MBH a estudios, y tan solo en la fase de análisis, la solicitud de acceso a datos específicos; además contempla la protección de datos, como parte de la privacidad de los participantes²¹.

SAIL ha sido aplicado en grandes colecciones de muestras: ENGAGE, *European Network for Genetic and Genomic Epidemiology*, el Karolinska Institutet

y el SUMMIT consortium, proporcionando buenos resultados, y continúa en la promoción de la construcción de ontologías con contribución de toda la comunidad de biobancos y el uso de mejores prácticas en pro de la interoperabilidad de los biobancos²².

Es importante desde el inicio de un biobanco prestar atención a la gestión de los datos, los flujos de trabajo automatizado, la armonización y la interoperabilidad, durante el desarrollo e implementación de sistemas informáticos, para no disminuir la utilidad de las muestras, contribuir con el desarrollo de marcadores predictivos, y estar preparados para la integración a biobancos de talla mundial, entre otros^{23,24}.

La evolución de las llamadas ciencias *ómicas*, demanda el acceso a MBH y sus datos, como recurso fundamental para mediciones y el desarrollo de la medicina personalizada; los resultados de estas nuevas mediciones (datos científicos), se deben retroalimentar a los biobancos, los cuales se potencian, además con la incorporación de datos clínicos de los participantes. Así, estos volúmenes de datos heterogéneos, desde su procesamiento, integración, fusión, protección y seguridad, con la utilidad de técnicas de minería de datos, y en la actualidad del big data, son un desafío para el tratamiento de los datos en los biobancos^{24,25}.

Como ya ha sido insistente, para que los datos compartidos sean comparables, necesitan en primer lugar ser armonizados. Diversas redes de biobanco, se encuentran en la tarea de la armonización: BBMRI-ERIC en Europa, BBRB y CAP *Biorepository Accreditation Program* en Estados Unidos, *Canadian Tumour Repository Network* en Canadá, *Human Tissue*

Authority y *NCRI's Confederation of Cancer Biobanks* (CCB) en Reino Unido, entre otros, para avanzar hacia una infraestructura de información universal para los bancos biológicos²³. Derivado de este trabajo en red y luego del trabajo de varios años (2012-2016), Europa cuenta actualmente con MIABIS 2.0 Core, con 22 atributos que describen los biobancos y propende por un uso más amplio y eficiente de los recursos biológicos y de esta manera lograr un estándar de información para biobancos²⁶.

Las Aplicaciones de Software y La Web

La recolección del MBH puede no representar un problema, en tanto la gestión de los datos asociados sí. La capacidad de gestión del MBH almacenado en un biobanco, que incluye el seguimiento o trazabilidad desde su ingreso en la toma de muestra hasta su salida para investigación (ubicación, procesos de descongelamiento, el consentimiento informado), requieren de un sistema de información de apoyo a los investigadores para consolidar la integridad de las muestras⁵.

Los desarrolladores de Sistemas de Gestión de Laboratorios (LIMS), utilizados en los laboratorios clínicos, han evolucionado para ofrecer servicios a repositorios de colecciones de muestras o biobancos. Estas aplicaciones de software, soportan los datos de las muestras desde su recopilación, almacenamiento, administración, búsqueda y recuperación de material con seguridad, y un rendimiento de análisis en tiempo real, confiable. Algunas aplicaciones soportan además consultas e informes de datos relacionados al MBH gestionado (**Tabla 1**)²⁷⁻³⁵.

Tabla 1. Soluciones de Software para Biobancos

Aplicación	Proveedor	Licencia	URL
NorayBanks	NorayBio	Comercial	http://www.noraybio.com/es/productos/noraybanks
Biobank	BC Platforms	Comercial	http://bcplatforms.com/solutions-category/biobank/
bioGes	Alatel	Comercial	http://alatel.es/es/soluciones-eco-lims/biobancos-biogest#quees
Freezerworks	Dataworks	Comercial	http://www.freezerworks.com/index.php/freezerworks/
Nautilus LIMS	Thermo Fisher	Comercial	https://www.thermofisher.com/order/catalog/product/INF-12000
LabWare Biorepository Solution	LabWare	Comercial	http://www.labware.com/es/p/02170
OpenSpecimen	Krishagni Product	Comercial	http://www.openspecimen.org/
Onyx	oBiBa	Libre	http://www.obiba.org/pages/products/onyx/
cbsrbiobank	CBSR	Libre	http://biosample.ca/

La web, se convierte en la herramienta a través de la cual, las redes de biobanco llegan a diferentes actores:

investigadores, comunidad, pacientes, la industria y los estados miembros.

Algunas redes y asociaciones de biobancos, a través de su plataforma web, comparten y exigen a los socios de las organizaciones, buenas prácticas para biobancos, en pro de la normalización y armonización^{37,38}.

Se ofrecen espacios web de apropiación social del conocimiento con socialización de eventos, formación en temas pertinentes, documentos descargables sobre normas, buenas prácticas, proceso y legislación de biobancos, divulgación de literatura científica en el tema, así como boletines, noticias y convocatorias nacionales e internacionales en el área de interés, entre otros^{6-8,37,38}.

Las plataformas web divulgan el directorio de los biobancos adscritos a la red; promulgan el trabajo colaborativo y la cooperación en torno a proyectos que hagan uso de muestras y gestionan y facilitan recursos para la investigación médica y biomolecular; algunos ofrecen consultas para búsqueda y solicitud de servicio de las muestras y datos. En conclusión, promueven la gestión profesionalizada de los servicios, la promoción de colecciones de alto valor estratégico y la realización de actividades de I+D+i (Investigación, desarrollo e innovación) en el campo de los biobancos^{6-8,36,37}.

Finalmente, la armonización, estandarización e interoperabilidad de los datos de biobancos, requieren del apoyo, a través de proyectos priorizados a nivel internacional, no para la creación de nuevos sistemas, sino para la evaluación y mejora del material existente. Así mismo, se necesita un compromiso, desde la operación de recolección física de los datos y la administración estructurada de los mismos, para alcanzar una fusión, con un enfoque que permita la protección y seguridad de los datos, a través de sistemas flexibles y escalables con capacidad de respuesta analítica eficaz^{23,36,37}.

La estructura organizacional del programa Cardiecol, con su Unidad de Datos y Unidad de Biobanco, soportan el desarrollo de una aplicación propia, que cumpla con los requisitos para la gestión electrónica de los procesos de transporte, ingreso, almacenamiento, actividad del MBH y documentación de los procesos estandarizados por el biobanco, como garantía de la calidad, al mismo tiempo que se establece la interoperabilidad con las aplicaciones de captación de los datos clínicos de los participantes, soporte para la investigación de la enfermedad cardio-cerebrovascular en Colombia³⁹.

CONCLUSIONES

Es necesario la conexión en red de los biobancos, como modelo sostenible a largo plazo, con procesos estandarizados para la recopilación, procesamiento, almacenamiento, gestión y distribución del MBH con fines de investigación, bajo las mejores prácticas clínicas, bajo el esquema de marcos de trabajo interoperables, con el soporte de plataformas bioinformáticas robustas, flexibles, escalables y seguras, que respalden la armonización de las operaciones y los datos al mismo tiempo que garantiza los intereses de los participantes.

AGRADECIMIENTOS

Apoyos Recibidos: Colciencias bajo el contrato 696-2014, Programa CARDIECOL, Código 617399847755. El Biobanco CARDIECOL, como unidad estructural del programa, que tiene como objetivo diseñar y desarrollar una plataforma biológica con población colombiana, adulta e infantil, para el estudio de factores de riesgo convencionales y no convencionales (genéticos) en enfermedad cardiovascular. Esta plataforma biológica con fines de investigación se encuentra asociada con la Unidad de Datos del programa, la cual se encarga del manejo de datos clínicos electrónicos vinculados al material biológico almacenado.

REFERENCIAS

1. Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica.
2. Serrano-Díaz N, Páez-Leal MC, Luna-González ML, Guío-Mahecha E. Biobanco: Herramienta fundamental para la investigación biomédica actual. *Rev Univ Ind Santander Salud*. 2016; 48(2): 97-117.
3. Cabrera LA, Wiesner C, Serrano ML, Peñaranda Correa NR, Huertas A. Consideraciones éticas y legales de los biobancos para investigación. *Rev Col Bioética*. 2010; 5(1): 121-141
4. Zika E, Paci D, Schulte T, Braun A, Rijkers Defrasne S, Deschênes M, et al. Biobanks in Europe: prospects for harmonisation and networking. (Joint Research Centre, European Commission, 2010). DOI: 10.2791/41701.
5. Blackman G. (2003-2016). Saving for the future.

- Europe Science. Cambridge, UK. <https://www.scientific-computing.com/feature/saving-future>
6. Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-ERIC) <http://bbmri-eric.eu/>
 7. Red de Biobancos de Latinoamérica y Caribe (REBLAC). http://www2.rinc-unasur.org/wps/wcm/connect/RINC/site/home/grupos_operativos/grupo_operativo_de_bancos_de_tumores
 8. Red Nacional de Biobancos, Plataforma Biobancos. Instituto de Salud Carlos III. España. <http://www.redbiobancos.es/Default.aspx?i=10>
 9. Mora M, Angelini C, Bignami F, Bodin AM, Crimi M, Di Donato, et al. The EuroBioBank Network: 10 years of hands-on experience of collaborative, transnational biobanking for rare diseases. *Eur J Hum Genet.* 2015 Sep; 23 (9): 1116-1123. DOI: 10.1038/ejhg.2014.272.
 10. Mascalconi D, Dove E, Rubinstein Y, Dawkins H, Kole A, McCormack P, et al. International Charter of principles for sharing bio-specimens and data. *Eur J Hum Genet.* 2015; 23: 721-728. DOI: 10.1038/ejhg.2014.197.
 11. Real Academia Española, RAE. <http://dle.rae.es/>.
 12. International Organization for Standardization. <http://www.iso.org/iso/home/standards.htm>
 13. Fred Coulson. Tutorial: Las 3 Formas Normales. 2009. <http://phlonx.com/resources/nf3/>.
 14. Calleros L, Cortés M, Luengo A, Mora I, Guijarro B, Martín P, et al. Puesta en marcha de una plataforma de proceso, almacenamiento y gestión de muestras clínicas: organización y desarrollo del Biobanco de REDinREN. *Nefrología.* 2012; 32(1): 28-34. DOI: 10.3265/Nefrologia.pre2011.Oct.11121.
 15. Angelow A, Schmidt M, Weitmann K, Schwedler S, Vogt H, Havemann C, et al. Methods and implementation of a central biosample and data management in a three centre clinical study. *Computer methods and programs in biomedicine.* 2008; 91: 82-90. DOI: 10.1016/j.cmpb.2008.02.002.
 16. Prokosch HU, Beck A, Ganslandt T, Hummel M, Kiehntopf M, Sax U, et al. IT infrastructure components for biobanking. *Appl Clin Inform.* 2010; 1(4): 419-429. DOI: 10.4338/ACI-2010-05-RA-0034.
 17. Norlin L, Fransson MN, Eriksson M, Merino-Martinez R, Anderberg M, Kurtovic S, et al. A Minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobank.* 2012; 10(4): 343-348. DOI: 10.1089/bio.2012.0003.
 18. Prokosch HU, Mate S, Christoph J, Beck A, Köpcke F, Stephan S, et al. Designing and implementing a biobanking IT framework for multiple research scenarios. *Stud Health Technol Inform.* 2012; 180: 559-563. DOI: 10.3233/978-1-61499-101-4-559.
 19. Sang H, Sun H, Cho H. Development of a Customized SNOMED-CT Toolkit for efficient clinical term searches and mapping biobank resources in Korea. *Adv Sc Technol Letters.* 2015; 88: 212-217. DOI: <http://dx.doi.org/10.14257/astl.2015.88.45>.
 20. Sang H, Sun H, Cho H. Development of a SNOMED-CT Mapping Framework for the Interoperability of Biobank Resources in Korea. *International Journal of Bio-Science and Bio-Technology.* 2015; 7(5):93-104. DOI: 10.14257/ijbsbt.2015.7.5.09.
 21. Gostev M, Fernandez-Banet J, Rung, J, Dietrich J, Prokopenk I, Ripatti S, et al. SAIL—a software system for sample and phenotype availability across biobanks and cohorts. *Bioinformatics,* 2011, 27(4), 589–591. DOI: 10.1093/bioinformatics/btq693.
 22. Spjuth O, Krestyaninova M, Hastings J, Shen HY, Heikkinen J, Waldenberger M, et al. Harmonising and linking biomedical and clinical data across disparate data archives to enable integrative cross-biobank research. *Eur J Hum Genet.* 2016; 24(4), 521-528. DOI: 10.1038/ejhg.2015.165.
 23. Schneider D, Riegman P, Cronin M, Negrouk A, Moch H, Balling R et al. Accelerating the development and validation of new value-based diagnostics by leveraging biobanks. *Public Health Genomics* 2016; 19(3): 160-169. DOI: 10.1159/000446534.
 24. Litton JE. Biobank Informatics: connecting phenotypes and phenotypes. *Methods Mol Biol.* 2011; 675: 343-361. DOI: 10.1007/978-1-59745-423-0_21.
 25. Huppertz B, Holzinger A. Biobanks – a source of large biological data sets: open problems and future challenges. *Knowledge Discovery and Data Mining,* 2014. LNCS 8401: 317-330.
 26. Merino-Martinez R, Norlin L, van Enckevort D, Anton G, Schuffenhauer S, Silander K, et al. Toward global biobank integration by implementation of the minimum information about biobank data sharing (MIABIS 2.0 Core). *Biopreserv Biobank.* 2016; 14(4): 298-306. DOI: 10.1089/bio.2015.0070.
 27. Noraybanks (2013). *NorayBio.* <http://www.noraybio.com/es/productos/noraybanks>
 28. Biobank (2016). Software platforms for next-generation sequencing – BC Platforms. <http://bcplatforms.com/solutions-category/biobank/>
 29. BioGest. (2016). *Alatel.* <http://alatel.es/es/soluciones-eco-lims/biobancos-biogest>
 30. Freezerworks (2015). *Dataworks.* <http://www.freezerworks.com/>
 31. Nautilus LIMS. (2016) Thermo Fisher Scientific

- Inc. <https://www.thermofisher.com/order/catalog/product/INF-12000>
32. LabWare Biorepository Solution (2016). LabWare <http://www.labware.com/es/p/021703>
 33. OpenSpecimen. Krishagni Product. <http://www.openspecimen.org/>
 34. Onyx. (2016). OBiBa. <http://www.obiba.org/>
 35. Cbsrbiobank. The Canadian BioSample Repository. <http://biosample.ca/>
 36. ISBER. (2016) International Society for Biological and Environmental Repositories (ISBER). <http://www.isber.org/>
 37. Biobanking and Biomolecular Resources Research Infrastructure. BBMRI D5.4 Requirements for a general information management system for European biobanks Deliverable Version: v1.2. 2009. <http://bbmri-eric-eu/>
 38. Lineamientos de la OCDE sobre Buenas Prácticas en Biocustodia para CRBs. Federación Latinoamericana de Colecciones de Cultivos-FELACC. <http://www.oecd.org/sti/biotech/oecdbestpracticeguidelinesforbiologicalresourcecentres.htm>
 39. Conocimiento y Acción para Reducir la Dimensión de la Enfermedad Cardiovascular en Colombia. (2016). Cardiecol. <http://www.cardiecol.org/>