



Introducción al *big data* biomédico II: Aplicaciones, repositorios y estándares

Grupo de Neurocomputación Biológica
Escuela Politécnica Superior
Universidad Autónoma de Madrid

<http://www.ii.uam.es/~gnb>

pablo.varona@uam.es

Campos de aplicación del big data biomédico

- Bioinformática
- Informática clínica
- Informática de imagen
- Informática de salud pública

Posible enfoque global o por patología

Bioinformática

- Analiza las variaciones de los sistemas biológicos en el nivel molecular:
- Abarca el uso de herramientas informáticas y de inteligencia artificial para abordar el alineamiento de secuencias, la predicción de genes, ensamblaje del genoma, alineamiento estructural de proteínas, predicción de estructura de proteínas, predicción de la expresión génica, interacciones proteína-proteína, y modelado de la evolución, etc.
- Conceptos básicos:
 - Genómica: estudia el genoma, el conjunto completo de ADN dentro de las células de un organismo y sus aplicaciones en terapia y biotecnología
 - Proteómica: estudio a gran escala de las proteínas, en particular de su estructura y función
 - Metabolómica: estudia los procesos químicos involucrados en el metabolismo

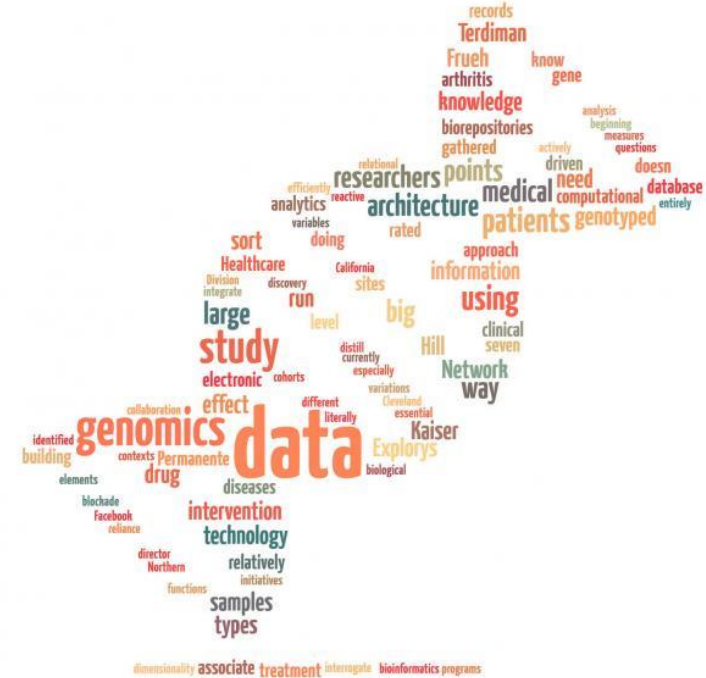
Clasificación de herramientas de big-data biomédico

- (1) Almacenamiento y recuperación de datos
- (2) Identificación de errores
- (3) Análisis de datos
- (4) Plataformas de integración y despliegue funcional

Estas categorías presentan interacciones y pueden solapar. Por ejemplo, la mayor parte de las aplicaciones de almacenamiento de datos pueden incluir análisis de simples y al revés.

Recordando: singularidad de los datos biomédicos

- La información biomédica es muy heterogénea, está almacenada de forma dispersa, suele ser voluminosa y recientemente involucra series temporales complejas, multivariadas.
- El reto es recoger esa información, almacenarla de forma en que se pueda integrar, compartir, analizar y utilizar fácilmente.



Beneficios de la integración de datos biomédicos

- Nuevos tratamientos, diagnósticos, mecanismos de alarmas, protocolos de seguimiento más eficaces y con un mayor grado de automatización, aplicaciones de telemedicina, etc.
- Medicina personalizada con tratamientos específicos para la singularidad de cada paciente en el instante adecuado.
- Descubrimiento de nuevas enfermedades.
- Identificación precisa del origen y control automático de epidemias.
- Mejoras en la administración de la gestión médica.
- Abaratamiento de los costes sanitarios.
- Desarrollo de tecnología bio-inspirada...

Retos por superar:



- Existe una separación de datos biomédicos por la división tradicional de disciplinas (p.ej. fisiología, genética).
- Muchos médicos y gerentes son reacios a compartir datos e historiales clínicos por temor a futuras demandas, sobre todo en los países más desarrollados tecnológicamente.
- Hay pocas iniciativas públicas para crear las infraestructuras y el marco legal necesarios.
- Riesgos de malos usos e invasión no consentida de privacidad.

Big data en investigación biomédica:

- Disciplina relativamente reciente
- Pocas bases de datos con información parcial
- Enorme potencial de avance en el conocimiento biomédico.

Fisiología (EEG+fMRI)+Genética+Comportamiento

Necesidad de estándares para el almacenamiento, tratamiento y comunicación segura de datos biomédicos

- La información desestructurada dificulta enormemente su uso.
- La estandarización favorece la disponibilidad, comunicación y uso de los datos.
- Inconveniente: los distintos tipos de datos suelen tener estándares distintos
- En big data biomédico, agregar datos heterogéneos puede favorecer enormemente la comprensión sobre muchos tipos de patologías.

Estándar HL7



- HL7 International (Health Level Seven) <http://www.hl7.org/>
- Es una **organización de desarrollo de estándares** (especificaciones y protocolos) **para facilitar el intercambio electrónico en el ámbito de datos de salud.**
- Su sede central está en Ann Arbor, MI, Estados Unidos y tiene capítulos internacionales en más de 50 países.
- Su misión es proporcionar **estándares globales para dominios sanitarios**: clínico, asistencial, administrativo y logístico para lograr una interoperabilidad real y segura entre los distintos sistemas de información en el área de la salud: datos e informes clínicos, imágenes médicas, fármacos, transacciones con entidades aseguradoras, etc..
- Se fundó en 1987 como sociedad internacional de informática médica **sin ánimo de lucro.**
- En 1994 fue acreditada como sociedad de desarrollo de estándares por el ANSI.



Estándar HL7



➤ Tareas del Health Level Seven:

- 1. Desarrollar estándares de sintaxis coherente y extensible que permita **estructurar información en salud**, que permita apoyar los procesos de atención al paciente y ser intercambiada entre aplicaciones de software, conservando al mismo tiempo la semántica de la información.
- 2. Desarrollar una metodología formal para apoyar la creación de estándares HL7 a partir del **Modelo de Referencia de Información** (Reference Information Model, RIM) de HL7.
- 3. **Educar** al sector salud, los responsables de formular políticas y al público en general sobre los beneficios de la normalización de información sanitaria y específicamente sobre los estándares HL7.
- 4. **Promover** el uso los estándares HL7 en todo el mundo a través de la creación capítulos internacionales de HL7, que participen en la elaboración y adopción de estándares HL7.

Estándar HL7



➤ Tareas del Health Level Seven

- 5. Proporcionar **expertos** para que el sector salud y las organizaciones interesadas en participar en HL7 puedan desarrollar estándares de información en salud en su área de especialización.
- 6. Colaborar con otras organizaciones desarrolladoras de estándares y organismos nacionales e internacionales de estandarización (como ANSI e ISO) para desarrollar estructuras de dominios de información en salud para promover el uso de **estándares compatibles**.
- 7. Colaborar con los usuarios de tecnologías de información en salud para asegurarse que los estándares HL7 cumplan con los **requerimientos del mundo real** y que los esfuerzos de desarrollo iniciados por HL7 permitan satisfacer nuevos requerimientos en función del desarrollo tecnológico.

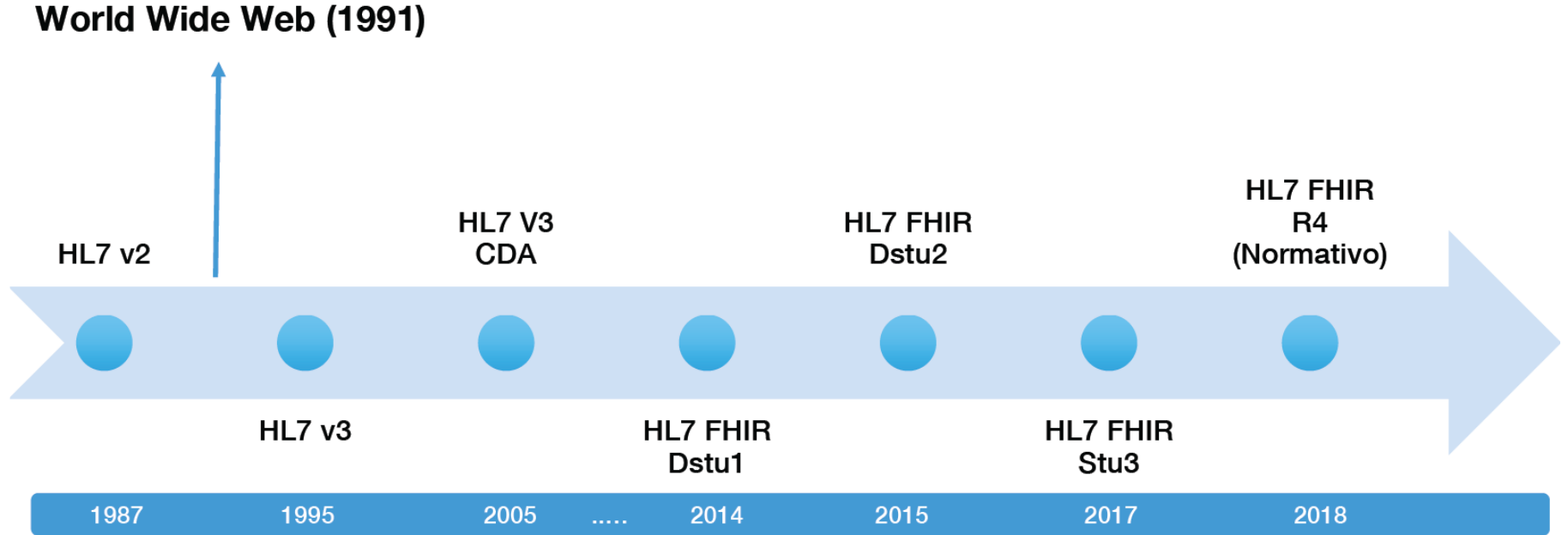
Estándar HL7



➤ Health Level Seven

- NO es una aplicación
- NO es una estructura de datos o una especificación de base de datos
- NO es una arquitectura para diseñar aplicaciones hospitalarias
- NO es una especificación para enrutar mensajes
- **Es un “conjunto de protocolos” para el intercambio de información médica**
- Su especificación más utilizada es un estándar de mensajería para el intercambio electrónico de datos en salud.
- “Level Seven” (Nivel Siete) hacen referencia al último nivel del modelo de comunicaciones para interconexión de sistemas abiertos (Open Systems Interconnection, OSI) de la Organización Internacional para la Estandarización (International Organization for Standardization, ISO).
- El “Nivel Siete” dentro del modelo es el nivel aplicación, que se ocupa de la definición y la estructura de los datos que van a ser intercambiados.

Versiones del Estándar HL7



Ejemplos de estándares HL7 International

Mensajería Versión 2.x

- Es el estándar internacional de mensajería para el intercambio electrónico de datos en los ámbitos clínico, asistencial, económico y logístico, más ampliamente utilizado en el mundo de la salud.
- La última versión incorpora esquemas basados en XML y un progresivo alineamiento con la metodología de desarrollo de la versión 3.

Ejemplo de HL7 v2

```
MSH|^~\&|HOSPITAL|CMQUI|EXTSYS|CMQUI|20180906175029||ADT^A31^ADT_A05|79854440|P|2.5||AL|NE||8859/1
EVN|A31|20180906162458|||HL7_cib^^^^^^^^^^^^^^^^
PID|1|127912359^^^CAMD^JHN^^^MD&&|ISO3166-2|1235340^^^PI^|NCHR123^^^MS^HC^^^ESP&&|ISO3166|ANCA^FRANCISCO
JUAN^PEREZ|GARCIA|19930803000000|M|||CALLE&PRINCIPAL&1^2 D
2^D^79^28^28018^724^16||912233595^PRN^PH^^^^^^^^12333595~12328569^ORN^PH^^^^^^^^634728569~^PRN^Internet^|1237
28569^PRS^CP|||||28/13668662-41|12399976Y|||DESCONOCIDA|||ESP^ESPANA||||N
PD1|||^^^FI^^16012810~HOSPITAL CENTRAL^^^^XX^^281270|^^^^^^^^^^^^1231280107M
PV1|1|N
```

En estos mensajes, cada línea corresponde a un segmento que se identifica por sus tres primeras letras. A continuación figuran los campos de ese segmento. Estos campos, a su vez, están formados por componentes y subcomponentes.

Los campos, componentes y subcomponentes se separan mediante caracteres separadores especiales:

- «|»: separador de campos
- «^»: separador de componentes
- «&»: separador de subcomponentes.
- «~»: separador en iteraciones de campos.
- «\»: carácter de escape.

Ejemplos de estándares HL7 International

Mensajería Versión 2.x Norma XML

- Esta especificación, conocida por HL7V2.xml facilita que los mensajes V2 sean procesados a través de un esquema XML
- Incrementa los niveles de validación para reducir errores en la construcción y en el envío/recepción del mensaje.

Ejemplos de estándares HL7 International

Versión 3 Metodología basada en Modelos de Referencia

Aborda la interoperabilidad clínica con el apoyo de modelos de referencia. A partir de un evento, aplicación emisora y receptora, se construye un mensaje o documento en base a un modelo restringido que procede, con sucesivas abstracciones, de un núcleo común, el RIM (*Reference Information Model*).

- El modelo común permite una interoperabilidad semántica.

Ejemplos de estándares HL7 International

Clinical Document Architecture (CDA)

- Es el núcleo de la historia clínica electrónica global de un paciente. Establece la composición de cualquier documento clínico (el informe de alta hospitalaria, el informe de resultados de una analítica o de una imagen diagnóstica, un resumen de situación clínica).
- La estructura de un documento CDA define una cabecera y un cuerpo con unas entradas en XML normalizadas que facilitan el procesamiento de su contenido y a la vez su visualización a través de cualquier navegador con una personalización del formato para los usuarios.

Ejemplos de estándares HL7 International

Clinical Context Management Specification (CCOW)

- CCOW está orientado a facilitar la integración de aplicaciones en un punto de actuación.
- Complementa el enfoque de HL7 en el intercambio de datos y el flujo de procesos en una organización de salud.
- Garantiza un acceso consistente a la información del paciente desde múltiples fuentes, a través de una interfaz de usuario única y ajustada con los estándares de seguridad internacionales.

Ejemplos de estándares HL7 International

Genómica Clínica

- Este estándar, basado en los esquemas HL7 V3 y V2, permite el intercambio de datos clínicos personalizados sobre genómica entre proveedores asistenciales, laboratorios de genética y centros de investigación biomédica.
- Se utiliza para registrar las observaciones sobre datos de genómica en los ensayos clínicos.
- Dispone de un modelo de análisis en el contexto de la genómica clínica.

Ejemplos de estándares HL7 International

SPL (Etiquetaje Estructurado de Productos/Fármacos)

- Structured Product Labeling (SPL) es una especificación que incluye la descripción detallada de un modelo de información basado en XML para etiquetar productos de una manera estructurada.
- El documento clínico que implementa está basado en HL7 CDA que define su estructura y la semántica de su contenido. Se usa principalmente para el etiquetaje de fármacos y la información sobre su uso y dosificación.

Bases de datos de investigación científica

- Informan del progreso de las distintas disciplinas desde la investigación básica, la investigación trasnacional y la investigación clínica
- Explican la relevancia y el uso real de datos biomédicos para el diagnóstico de enfermedades.
- Proporcionan datos sobre el porcentaje de éxito de tratamientos establecidos y experimentales.
- Reflejan en general el progreso del conocimiento científico.
- Sirven para identificar artículos que describen bases de datos disponibles en repositorios.

Bases de datos de investigación científica

- Pubmed: <https://www.ncbi.nlm.nih.gov/pubmed/>
- Web of Science/Web of Knowledge: <https://apps.webofknowledge.com>
- Scopus: <https://www.scopus.com>
- Google scholar: <https://scholar.google.com/>

Google genomics

- Capacidad para procesar y analizar datos genómicos en **tiempo real** con BigQuery, en programación literaria con Cloud Datalab, por lotes con GATK, con Apache Spark o Cloud Dataflow, o con un clúster de Grid Engine.
- Alta escalabilidad: carga y procesa de forma eficiente petabytes de lecturas de secuencias, variantes, referencias y anotaciones.

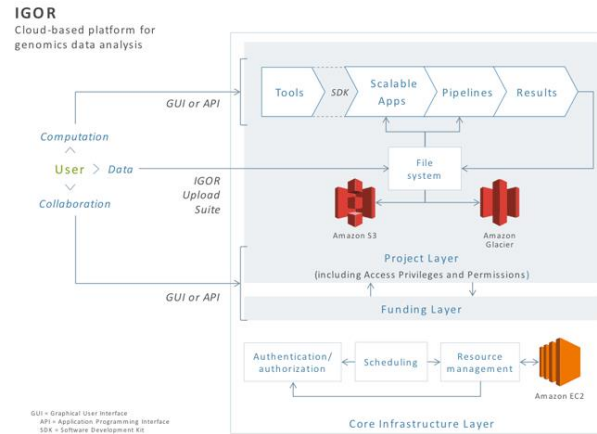


Google
Genomics

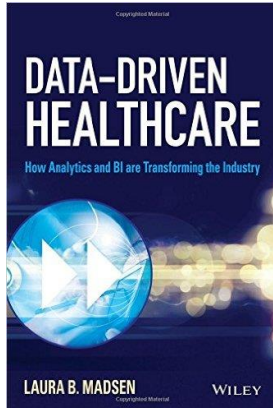
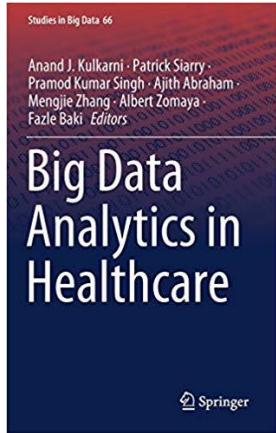
- <https://cloud.google.com/genomics/>
- Las extensiones de Google Cloud Platform para almacenar, procesar, analizar y compartir grandes conjuntos de datos de genómica
- Estándar abierto de la Global Alliance for Genomics and Health que funciona en varios repositorios de genomas.
- Máquinas virtuales, con almacenamiento escalable y bases de datos SQL y NoSQL administradas

Amazon web services (AWS) y gnómica

- Proporciona servicios de Seven Bridges Genomics con tecnología de next-generation sequencing (NGS) a través de su plataforma IGOR.
- IGOR es una plataforma de análisis y descubrimiento de datos genéticos basada en la nube.



Más información:



- Cirillo, D., & Valencia, A. (2019). Big data analytics for personalized medicine. [*Current Opinion in Biotechnology*, 58, 161-167.](#)
- Bui, A.A.T. Van Horn J.D., [*Envisioning the future of 'big data' biomedicine.* Journal of Biomedical Informatics 2017; 69:115-117.](#)
- Celi LA, Citi L, Ghassemi M, Pollard TJ (2019) [*The PLOS ONE collection on machine learning in health and biomedicine: Towards open code and open data.* PLOS ONE 14\(1\): e0210232.](#)
- Costa F, [*Big Data in Biomedicine.* Drug Discovery Today 2014 19\(4\): 433–440.](#)
- Alyass A, Turcotte M, Meyre D. [*From big data analysis to personalized medicine for all: challenges and opportunities.* BMC Medical Genomics 2015; 8:33.](#)
- Raghupathi W, Raghupathi V. [*Big data analytics in healthcare: promise and potential.* Health Inf Sci Syst. 2014;2\(1\):3.](#)
- Hersh W, Jacko JA, Greenes R, Tan J, Janies D, Embi PJ, et al. [*Health-care hit or miss?* Nature 2011;470\(7334\):327–329.](#)
- Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2017). [*Health-CPS: Healthcare cyber-physical system assisted by cloud and big data.* IEEE Systems Journal, 11\(1\), 88-95.](#)