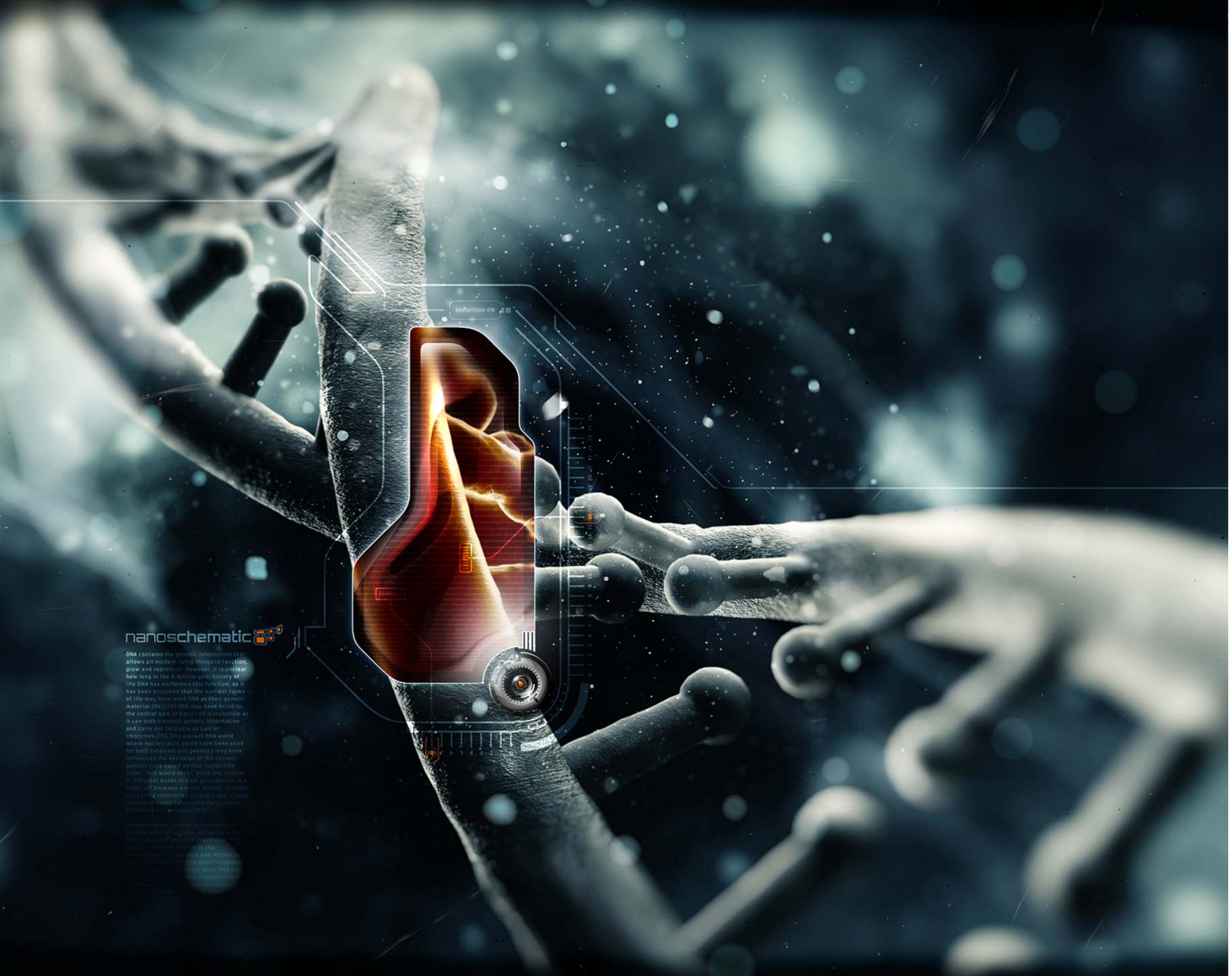


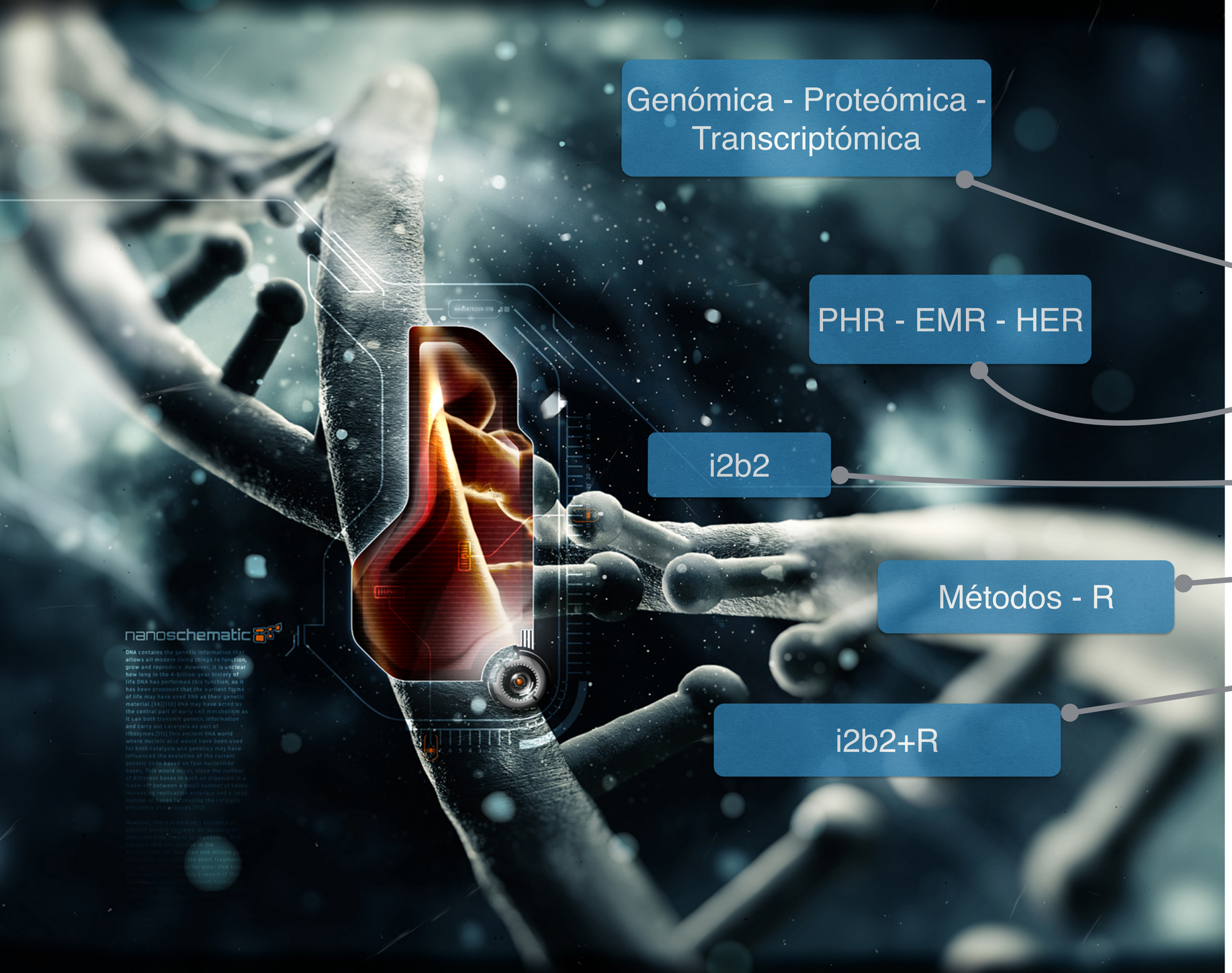
FACULTAD DE MEDICINA
UNIVERSIDAD DE CHILE

Análisis de Plataformas de Integración Clínico-Omicas y Métodos Estadísticos Predictivos para Apoyo al Diagnóstico y Tratamiento

Patricio Araneda García
Magíster (c) en Informática Médica

Prof. Dr. Rodrigo Assar

MIM 2014 MAGISTER EN INFORMATICA MEDICA



Genómica - Proteómica - Transcriptómica

PHR - EMR - HER

i2b2

Métodos - R

i2b2+R

FACULTAD DE MEDICINA
UNIVERSIDAD DE CHILE

Temario

- Datos Omicos
- Datos Clínicos
- Integración
- Apoyo Estadístico
- Plataforma
- Referencias

MIM 2014
MAGISTER EN INFORMATICA MEDICA

Introducción

La génesis y el desarrollo de enfermedades de alto impacto como el cáncer y otras esta relacionado a la disfunción de los genes.

La secuenciación del genoma humano ha impactado la investigación del cáncer a nivel molecular.

Medicina Traslacional

Facilitar la Transición desde la investigación básica hacia el tratamiento clínico al paciente.

Desarrollo de nuevas drogas y/o procedimientos que basados en data genómica promuevan un tratamiento personalizado.

Datos Ómicos

Genómica
Estudio integral e interdisciplinarios del genoma

- Definir función genica basado en su secuencia
- Determinar la interacción entre genes

Transcriptómica
Estudio y comparación de los conjuntos de ARNm o transcriptomas.

- Análisis de microarrays
- Determinación de expresion genica

Proteómica
Estudio del conjunto de proteínas, estructura y función expresados a partir de un genoma determinado.

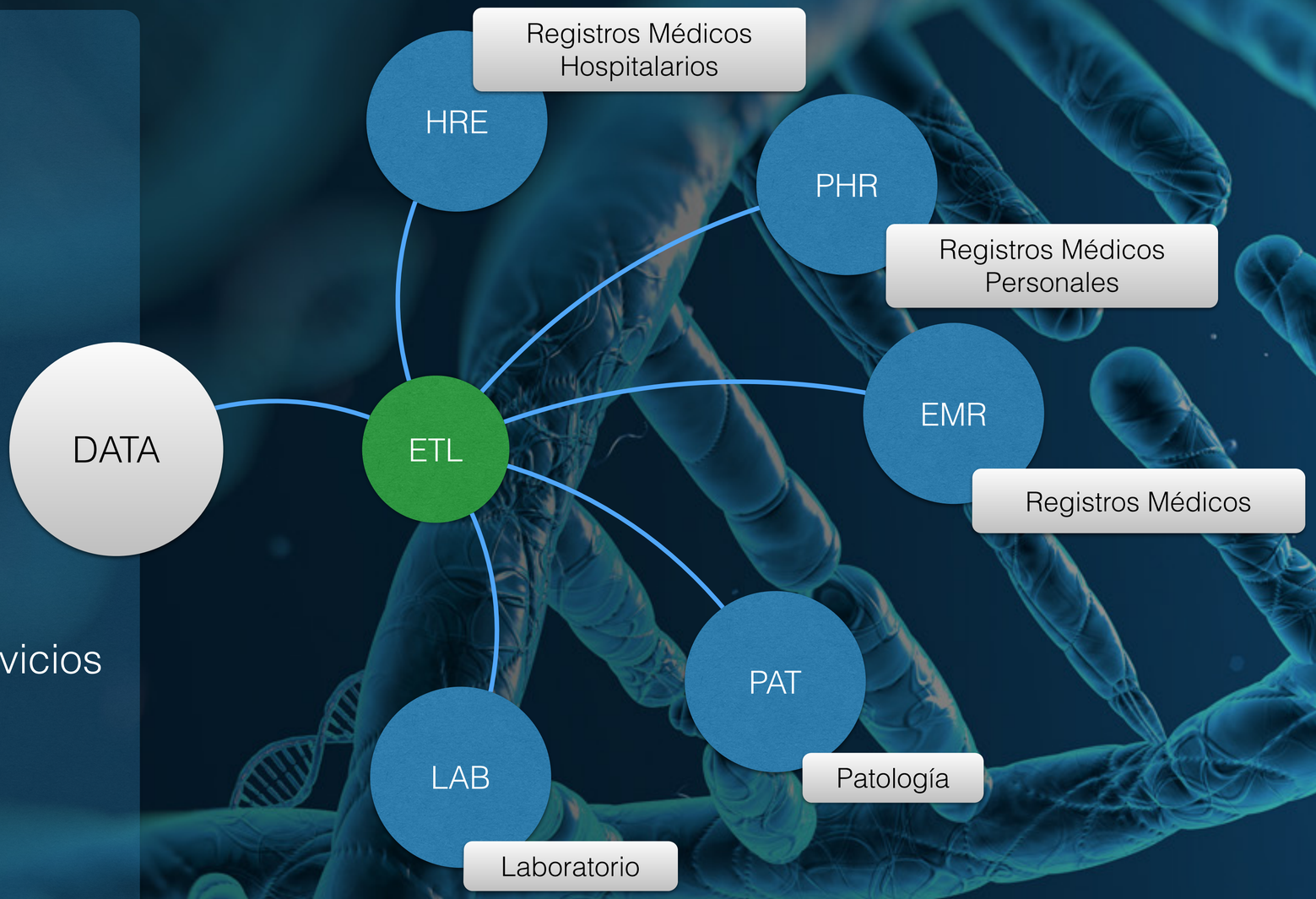
- Identificación de los componentes de proteína.
- Expresion diferencial
- Caracterización de las interacciones

Datos Clínicos

Gestión Clínica

- Privacidad
- Exactitud
- Completitud
- Estandarización

- Integración funcional a nivel de servicios
- Interoperabilidad entre sistemas

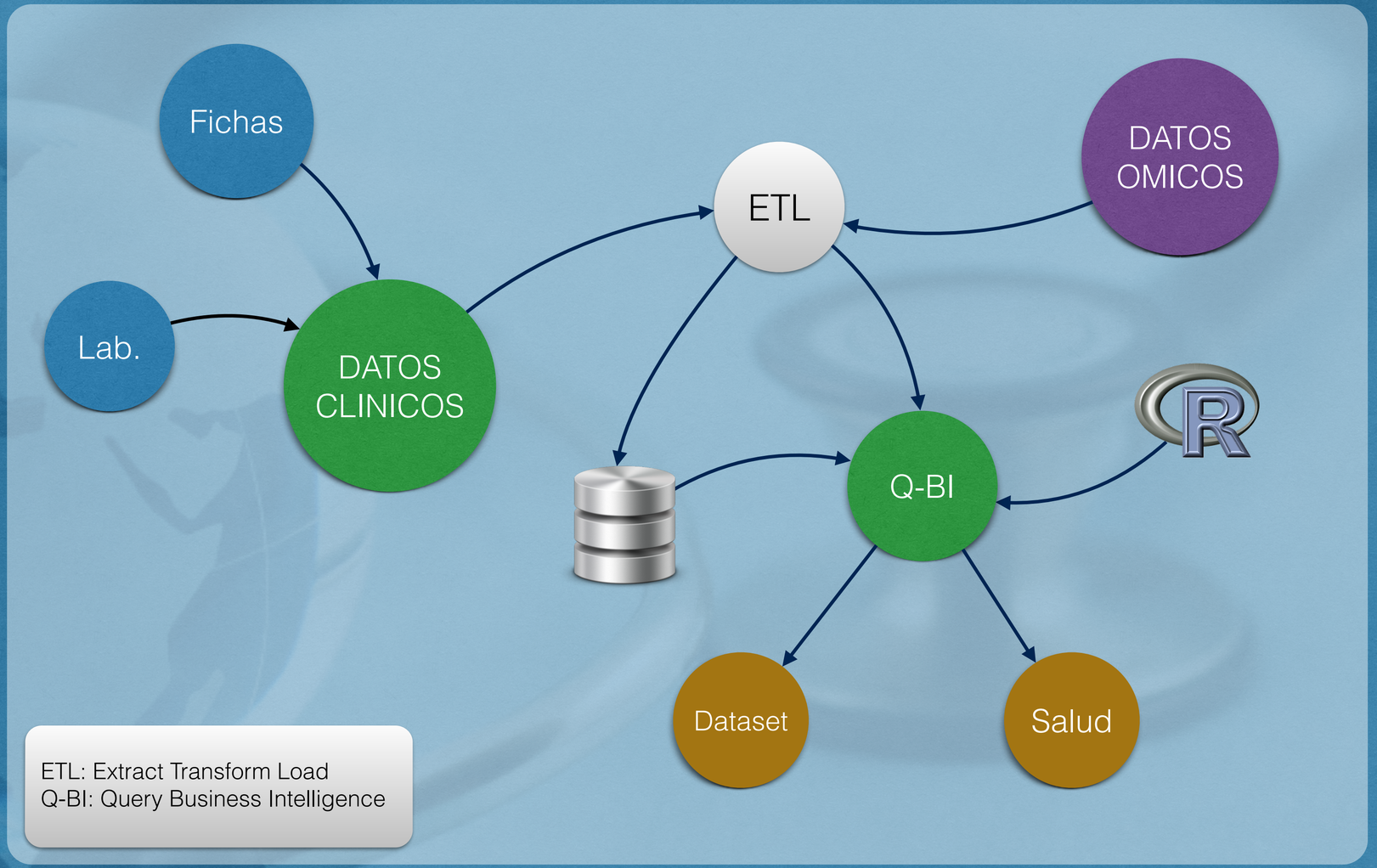


Integración

Plataformas de integración

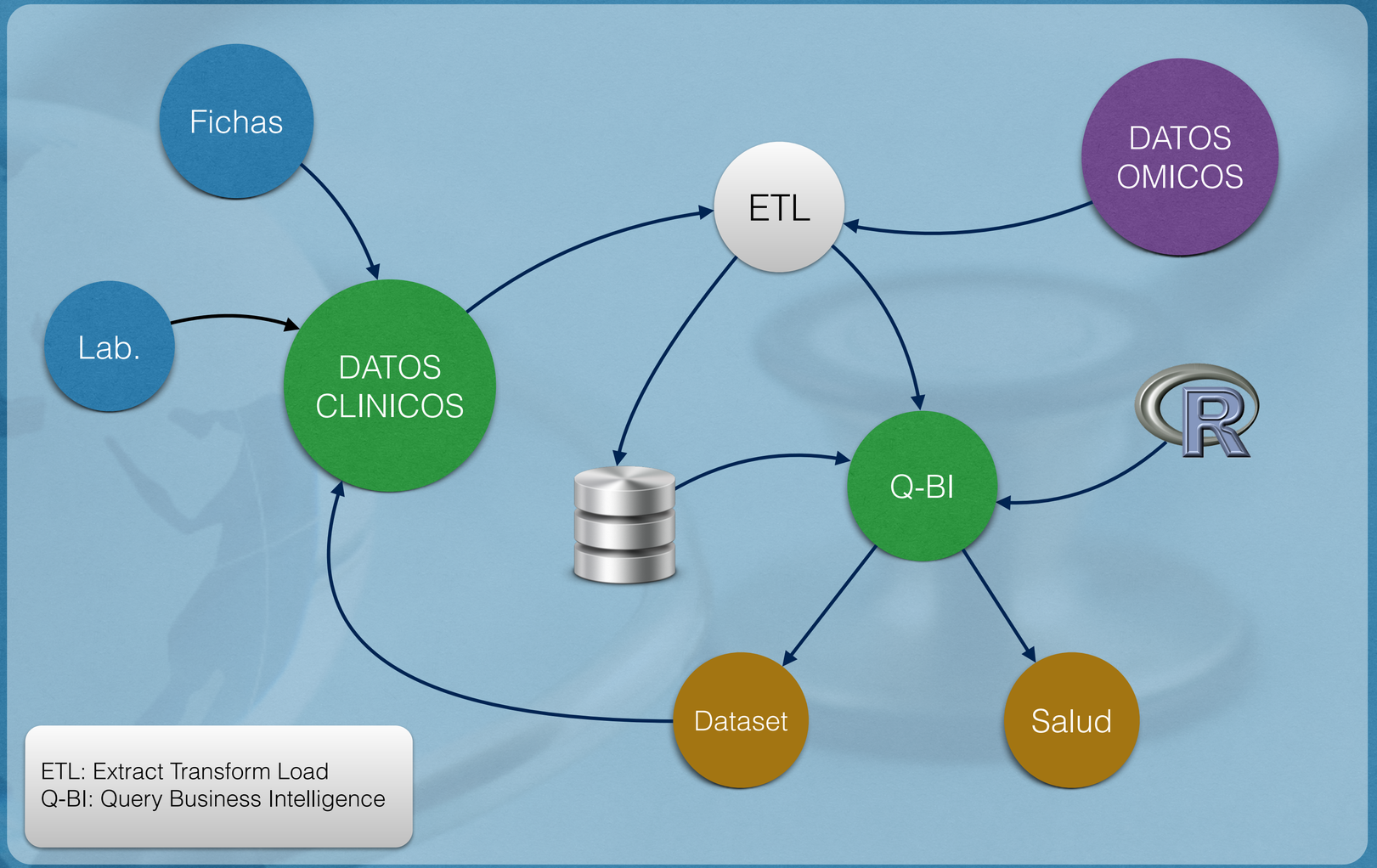
Metas

- Almacenar e integrar datos clínicos y ómicos
- Cruce de información
- Marco de análisis (generación de hipótesis médica)



Desiderata

- Privacidad
- Interoperabilidad
- Granularidad
- Mantenimiento
- Personalización
- Análisis Predictivo



Plataformas de integración [8]

- **i2b2** (Informatics for Integrating Biology & the Bedside).
- **ICOD** (Integrated Clinical Omics Database).
- **IDASH** (Integrating data for analysis, anonymization and sharing).
- **BRISK** (Biology-Related Information Storage Kit).

Características

Soporte de interoperabilidad limitada o nula.

Uso de estándares limitado.

Integración con herramientas estadísticas de terceros (R).

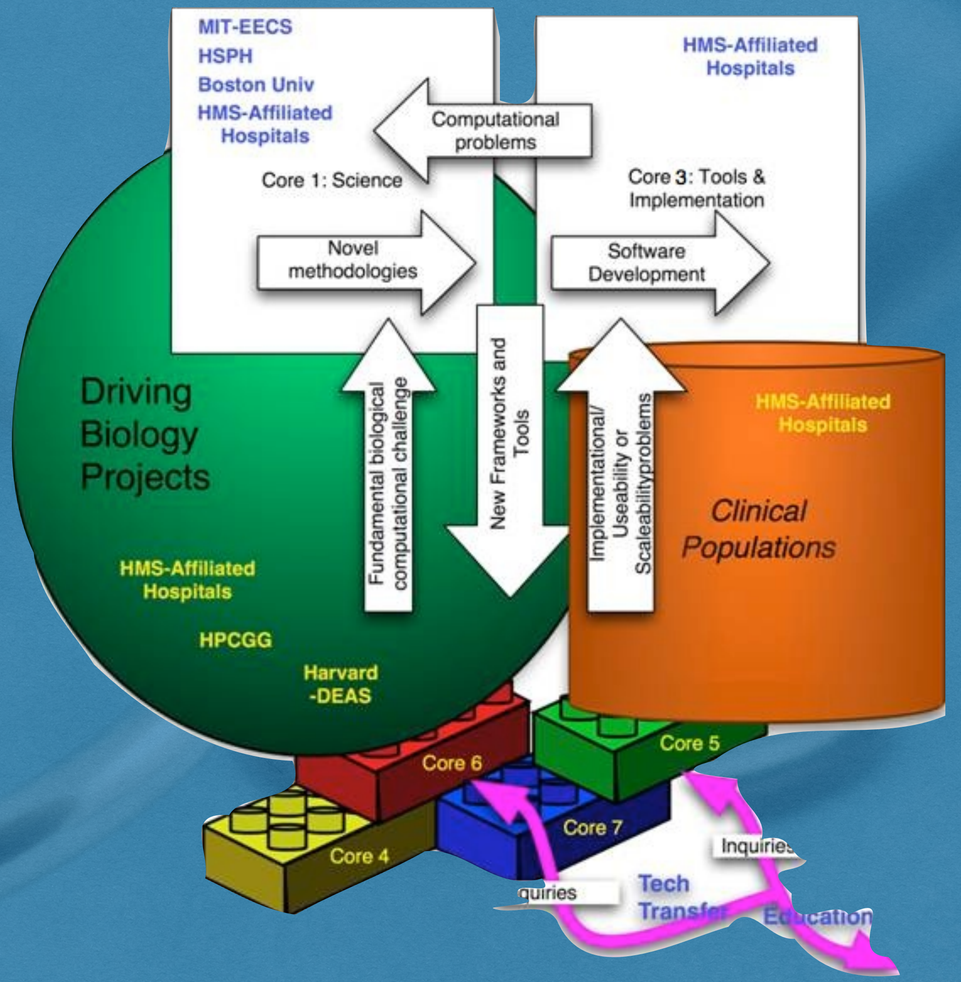
En almacén local o cloud.

Privacidad de datos (anonimicidad)

[8] Canuel *et al.*

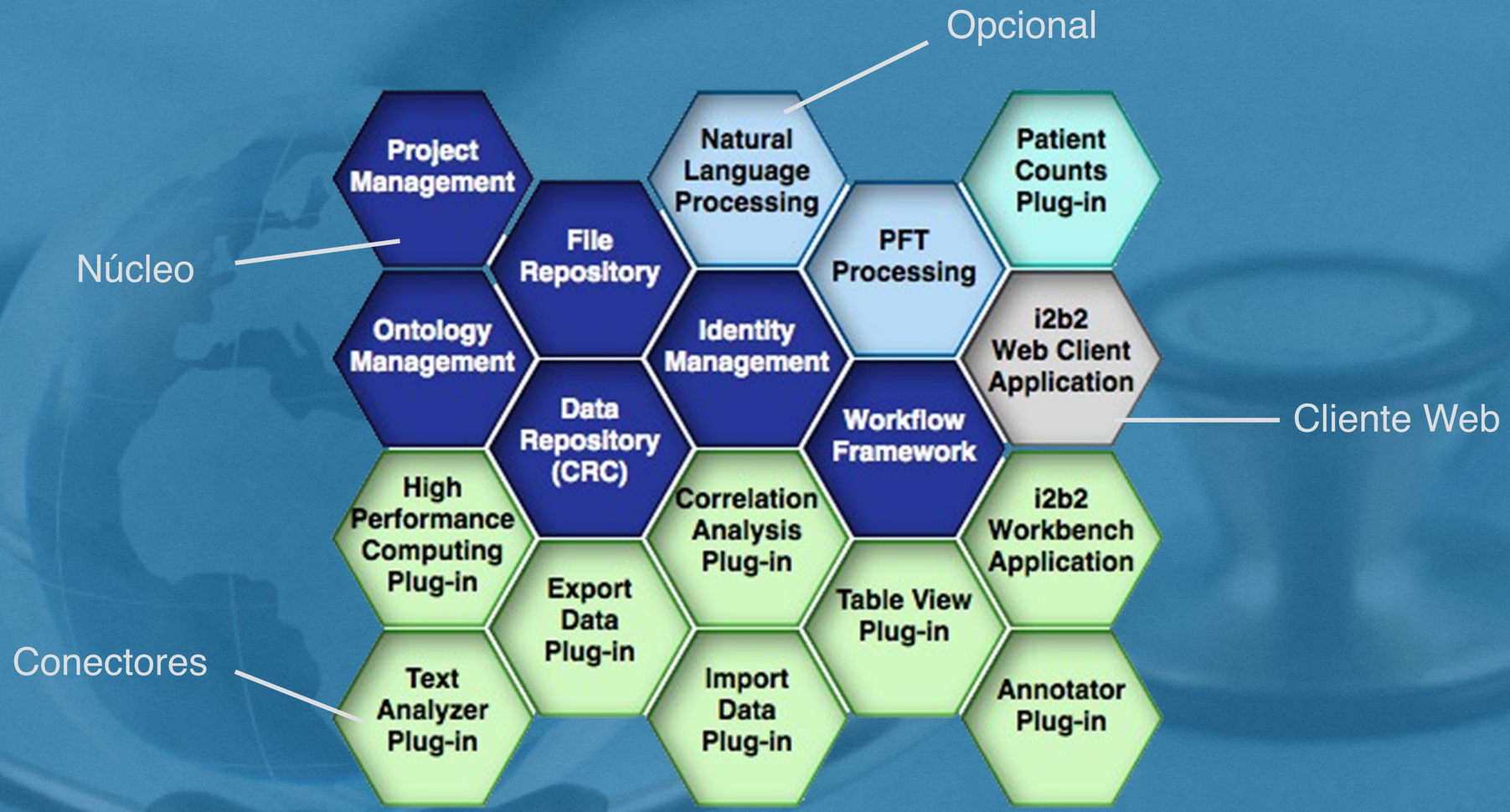
i2b2 [14]

- Plataforma Opensource
- National Center for Biomedical Computing (NCBC)
- Cores
  - Harvard Hospitals
  - MIT
  - Harvard School of Public Health
  - Harvard Medical School
  - Harvard/MIT Division of Health Sciences and Technology

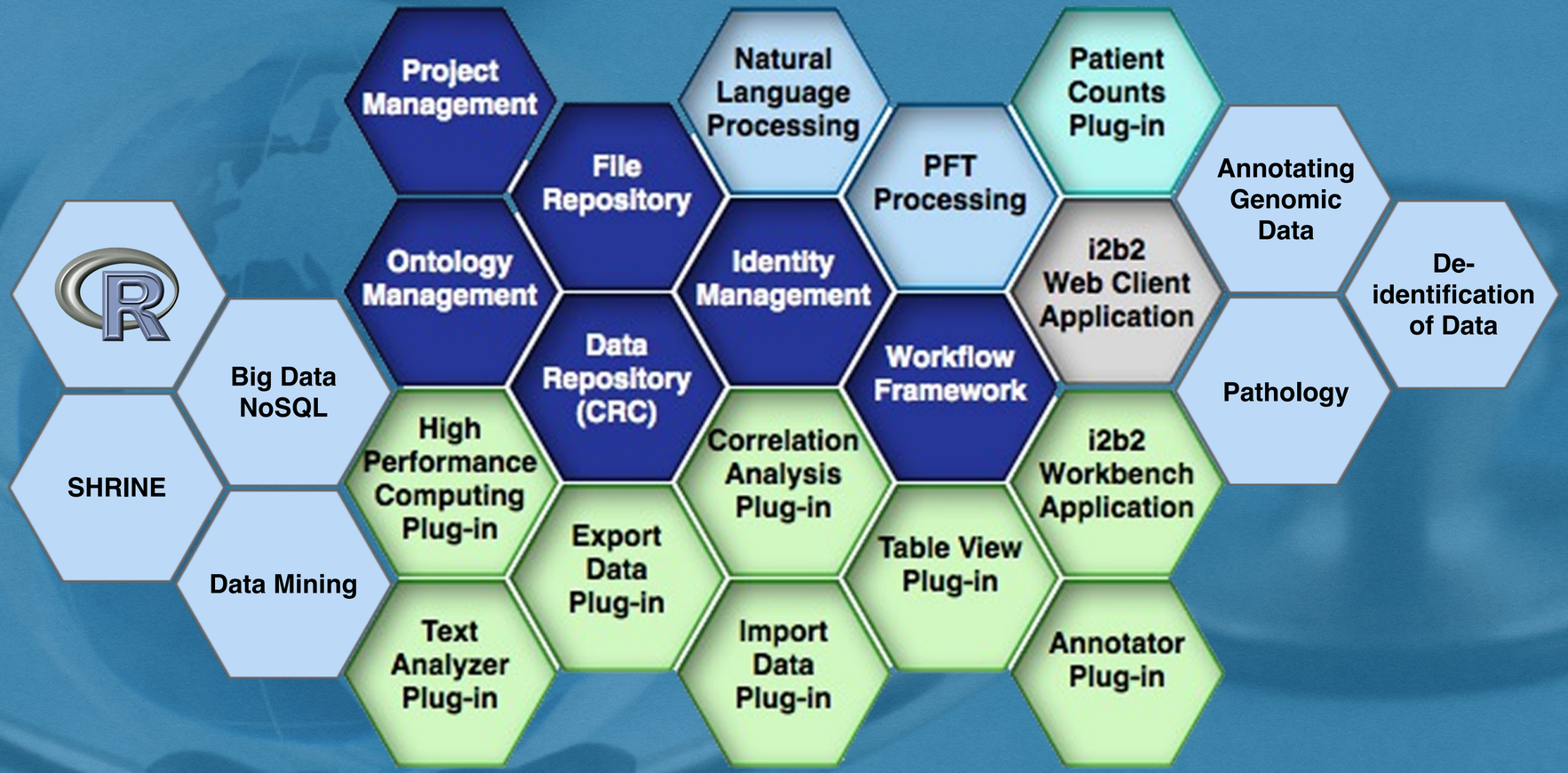


[14] Natter *et al.*

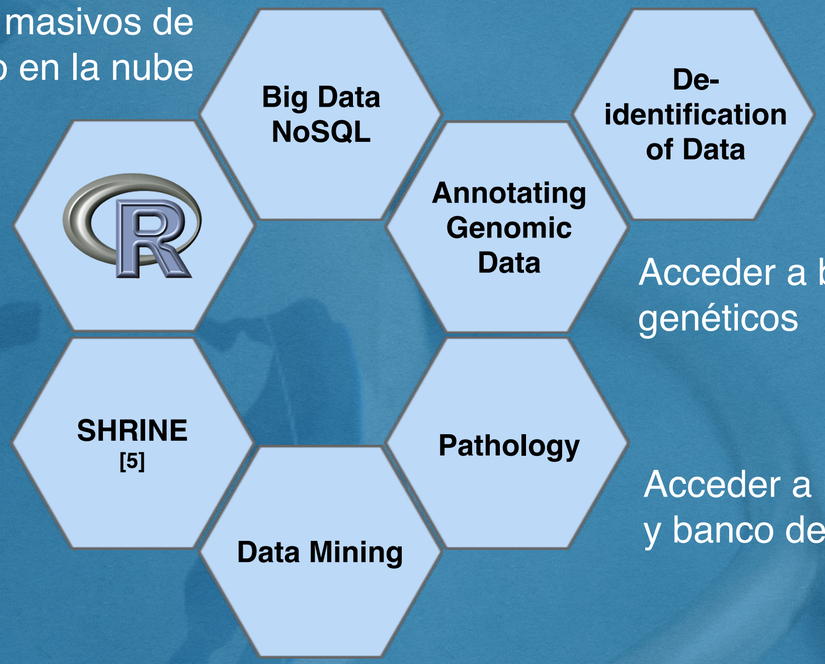
i2b2 - Core


Opcional
Project Management
Natural Language Processing
Patient Counts Plug-in
Núcleo
File Repository
PFT Processing
Ontology Management
Identity Management
i2b2 Web Client Application
Cliente Web
Data Repository (CRC)
Workflow Framework
High Performance Computing Plug-in
Correlation Analysis Plug-in
i2b2 Workbench Application
Export Data Plug-in
Table View Plug-in
Conectores
Text Analyzer Plug-in
Import Data Plug-in
Annotator Plug-in

i2b2 - Core


Project Management
Natural Language Processing
Patient Counts Plug-in
File Repository
PFT Processing
Annotating Genomic Data
Ontology Management
Identity Management
i2b2 Web Client Application
De-identification of Data
Big Data NoSQL
Data Repository (CRC)
Workflow Framework
Pathology
SHRINE
High Performance Computing Plug-in
Correlation Analysis Plug-in
i2b2 Workbench Application
Data Mining
Export Data Plug-in
Table View Plug-in
Text Analyzer Plug-in
Import Data Plug-in
Annotator Plug-in

i2b2 - Personalización



Acceder a volumenes masivos de datos local o en la nube

Anonimizar / re-identificar data de paciente

Realizar análisis predictivo

Acceder a banco de datos genéticos

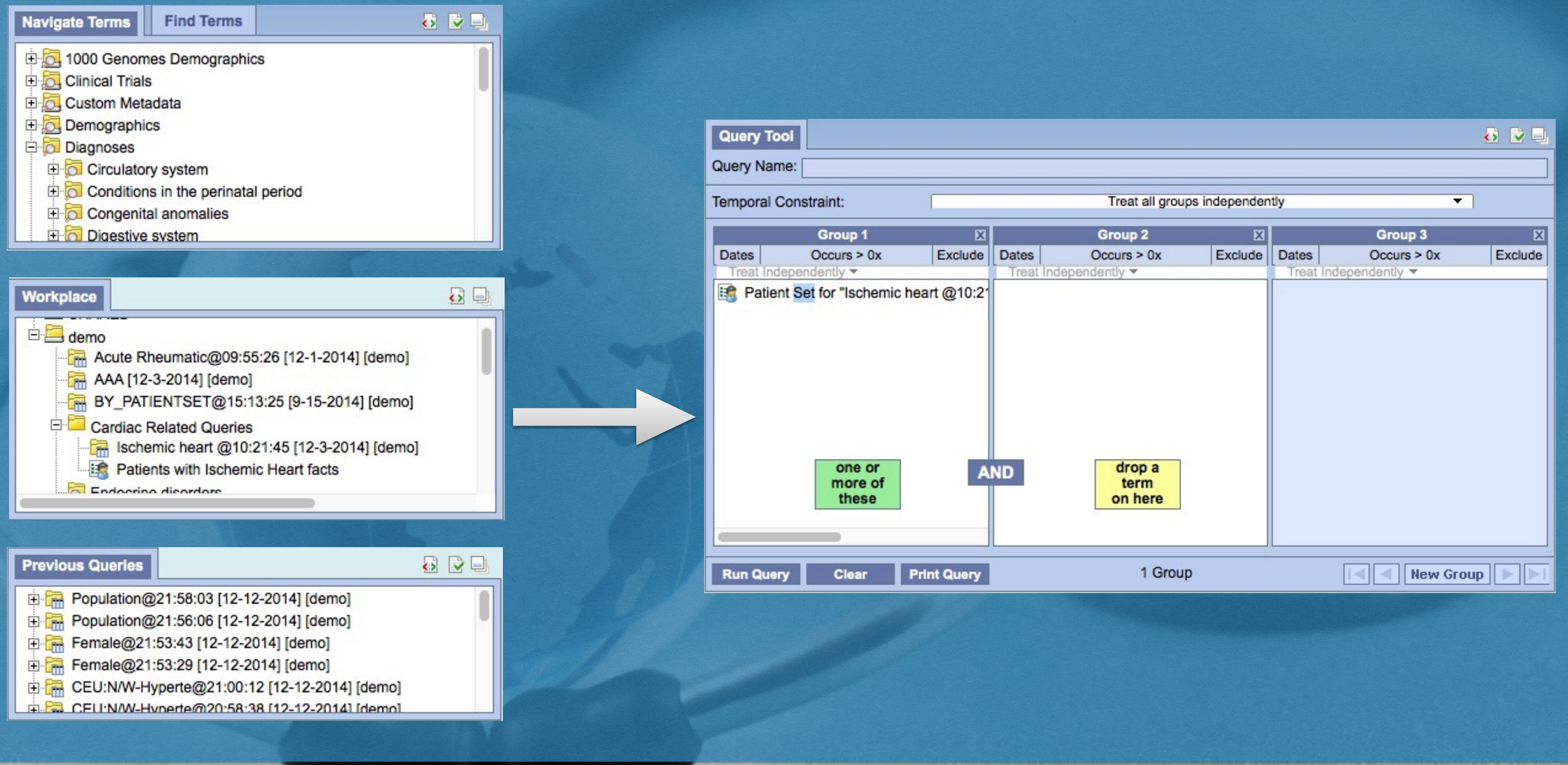
Integrar investigaciones en forma regional / nacional

Acceder a información de biopsias y banco de imágenes digitales

Analizar y recuperar información

[5] McMurry *et al.*

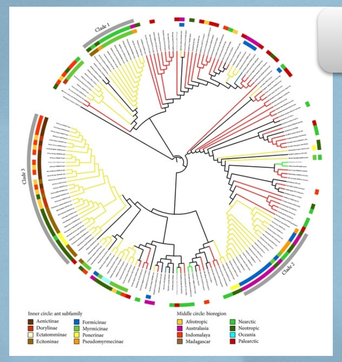
i2b2 | Query

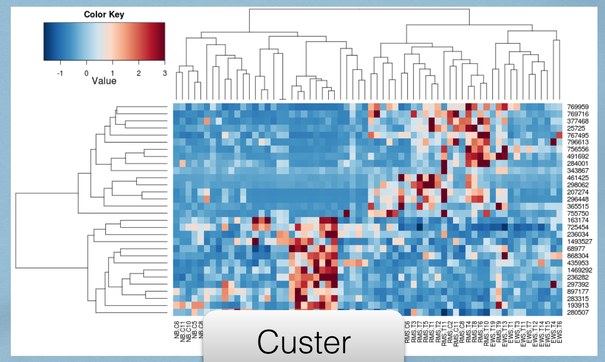

Navigate Terms
Find Terms
1000 Genomes Demographics
Clinical Trials
Custom Metadata
Demographics
Diagnoses
Circulatory system
Conditions in the perinatal period
Congenital anomalies
Digestive system
Workplace
demo
Acute Rheumatic@09:55:26 [12-1-2014] [demo]
AAA [12-3-2014] [demo]
BY_PATIENTSET@15:13:25 [9-15-2014] [demo]
Cardiac Related Queries
Ischemic heart @10:21:45 [12-3-2014] [demo]
Patients with Ischemic Heart facts
Previous Queries
Population@21:58:03 [12-12-2014] [demo]
Population@21:56:06 [12-12-2014] [demo]
Female@21:53:43 [12-12-2014] [demo]
Female@21:53:29 [12-12-2014] [demo]
CEU:N/W-Hyperte@21:00:12 [12-12-2014] [demo]
Query Tool
Query Name:
Temporal Constraint:
Treat all groups independently
Group 1
Group 2
Group 3
Dates
Occurs > 0x
Exclude
Treat Independently
Patient Set for "Ischemic heart @10:2
one or more of these
AND
drop a term on here
Run Query
Clear
Print Query
1 Group
New Group

Análisis Predictivo | R [12]

Integración de información para estudios estadísticos analíticos y predictivos.

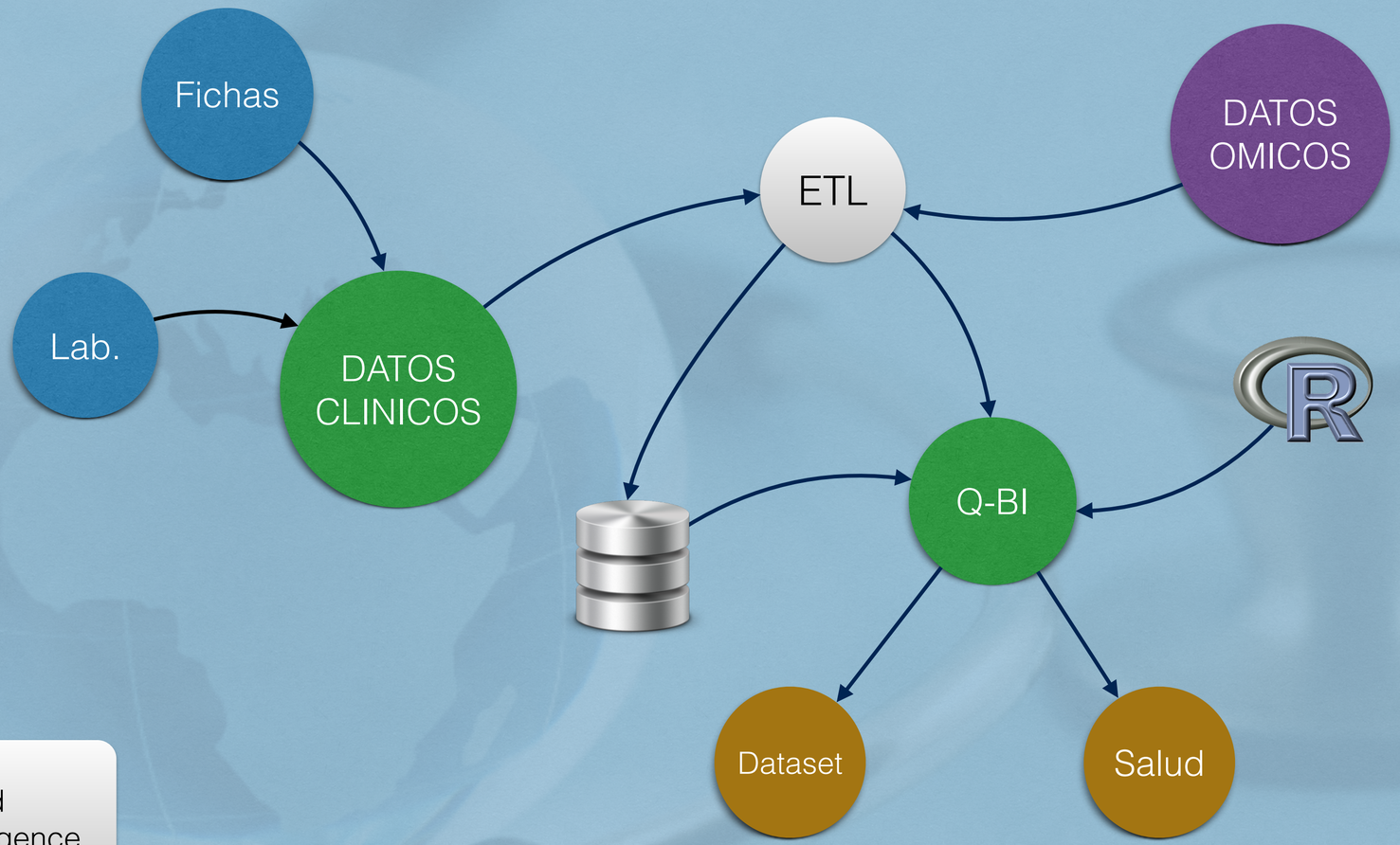
Estudio de grandes volúmenes de datos.





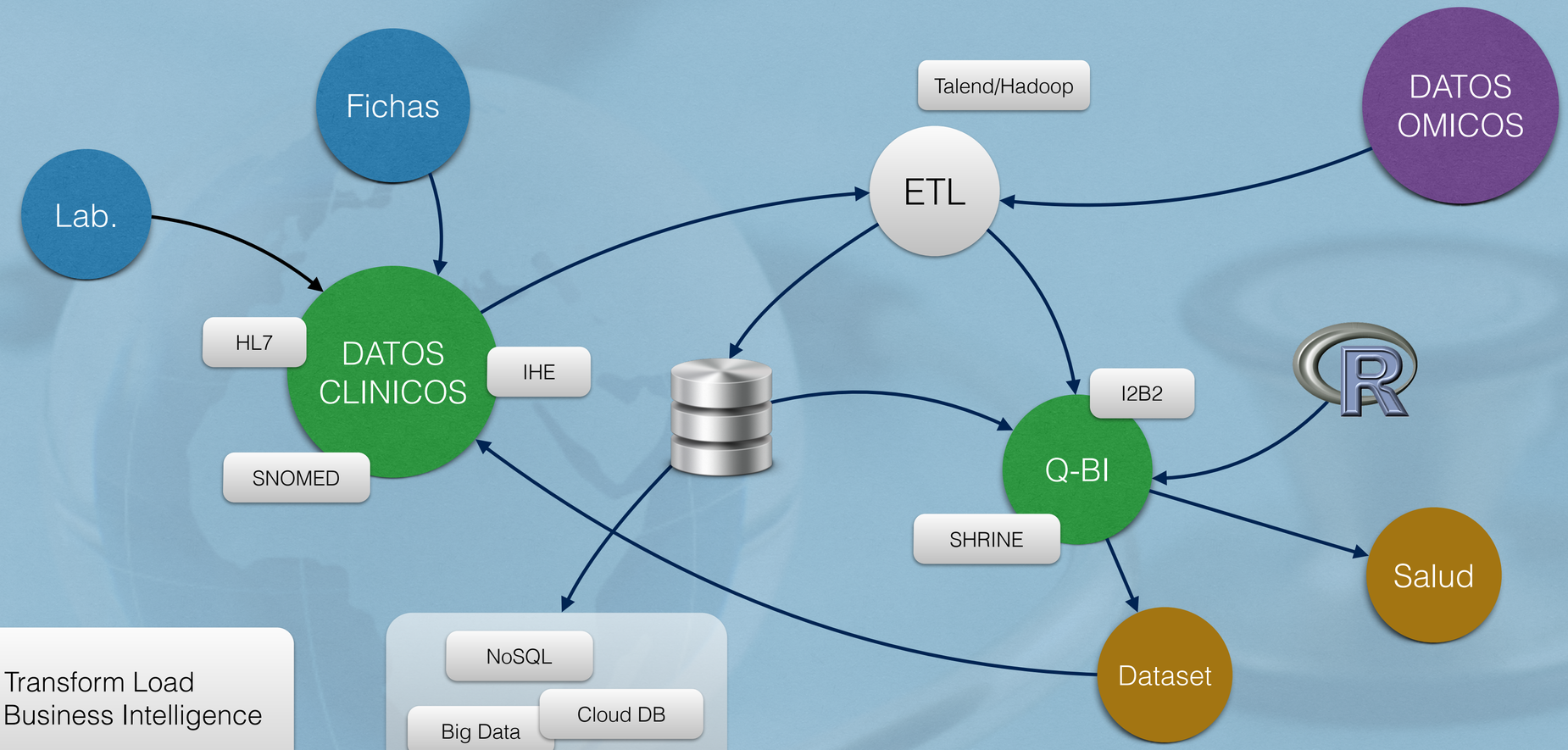
[12] Segagni *et al.*

Plataformas de integración



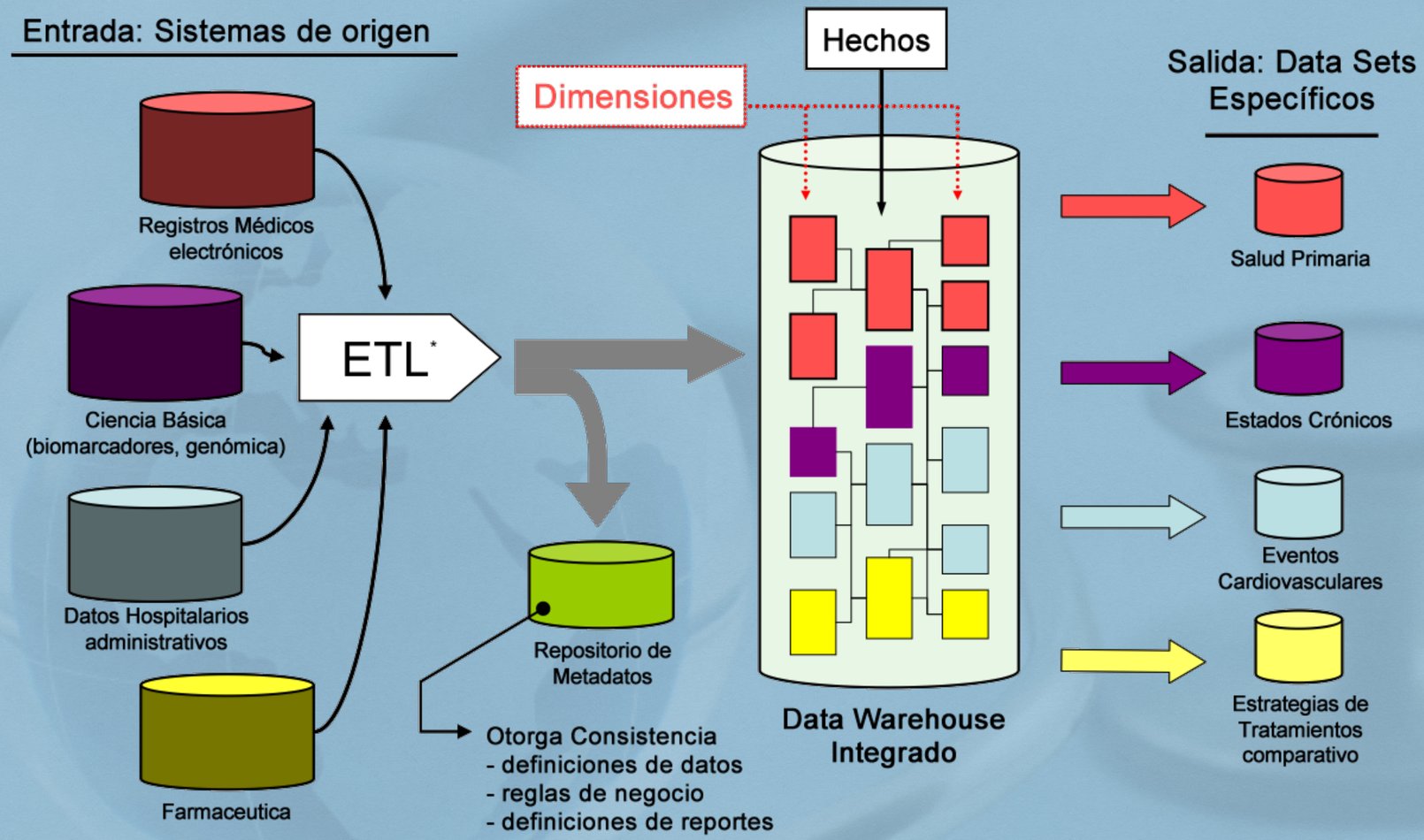
ETL: Extract Transform Load
Q-BI: Query Business Intelligence

Plataformas de integración



ETL: Extract Transform Load
Q-BI: Query Business Intelligence
IHE:
SHRINE:

ETL (Extract, Transformation and Load)



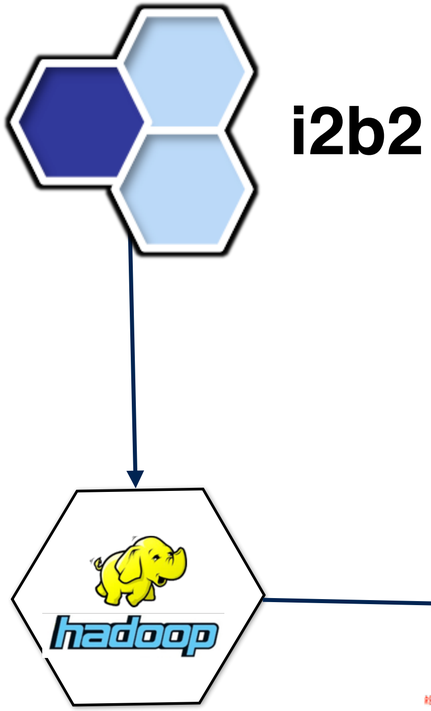
- Basado en el modelo de datos de i2b2 (Hechos=medidas, dimensiones=conceptos o descripciones)
- Arquitectura escalaba, el contenido es mareado a los conceptos en repositorio de metadatos
- Rápido, incremental, con un proceso de desarrollo aditivo

Perspectivas

i2b2 y Big Data

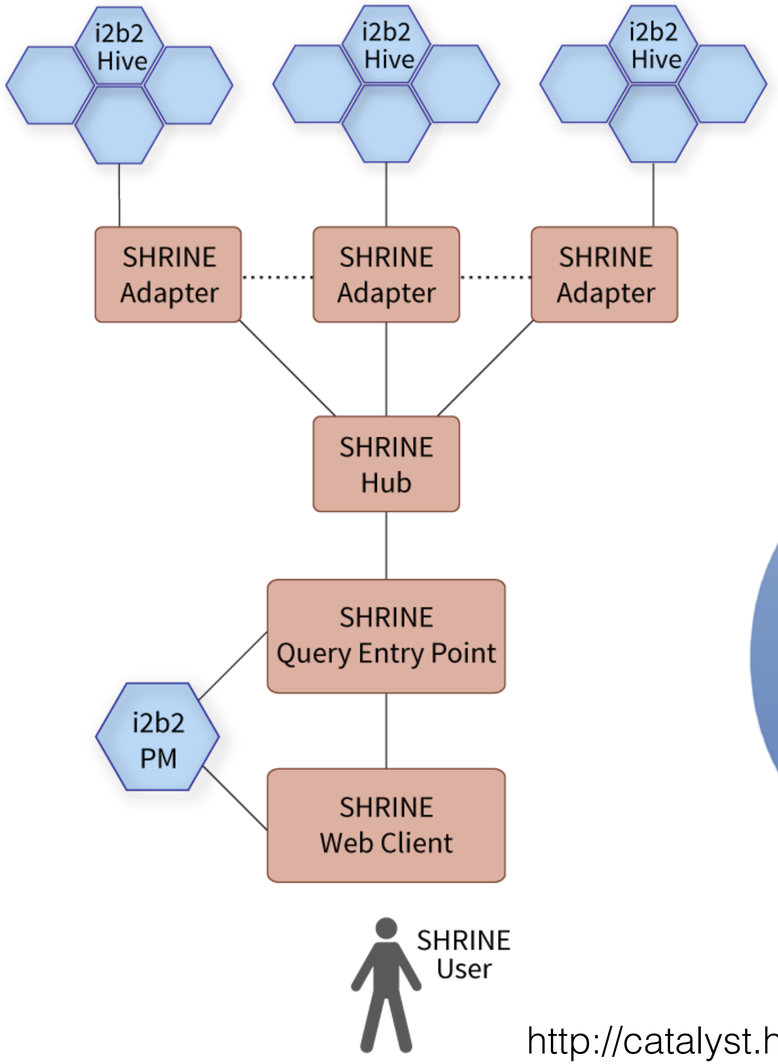
Integración de datos no relacionales y no estructurados.

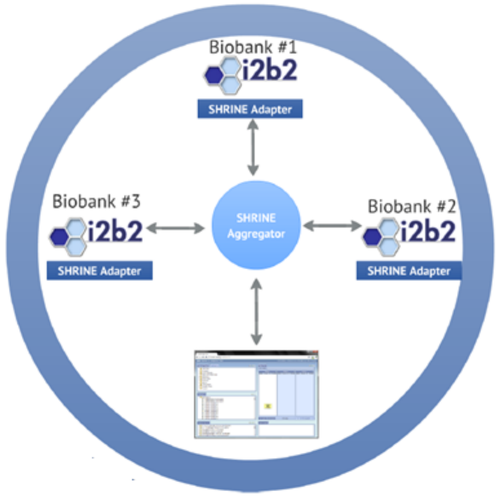
- Sistemas de monitoreo de paciente.
- Experimentos Next Generation Sequencing.



SHRINE (Shared Health Research Information Network)

- Basada en plataforma i2b2.
- Desarrollada por Harverd Catalyst: the Harvard Clinical and Translational Science Center
- Compila grandes grupos de pacientes bien caracterizados.
- Determina la cantidad de pacientes de cada hospital participante definidos en base a criterios de inclusión/exclusión.
- La información es agregada, protege la privacidad del paciente.
- Genera investigaciones a nivel regional o nacional.

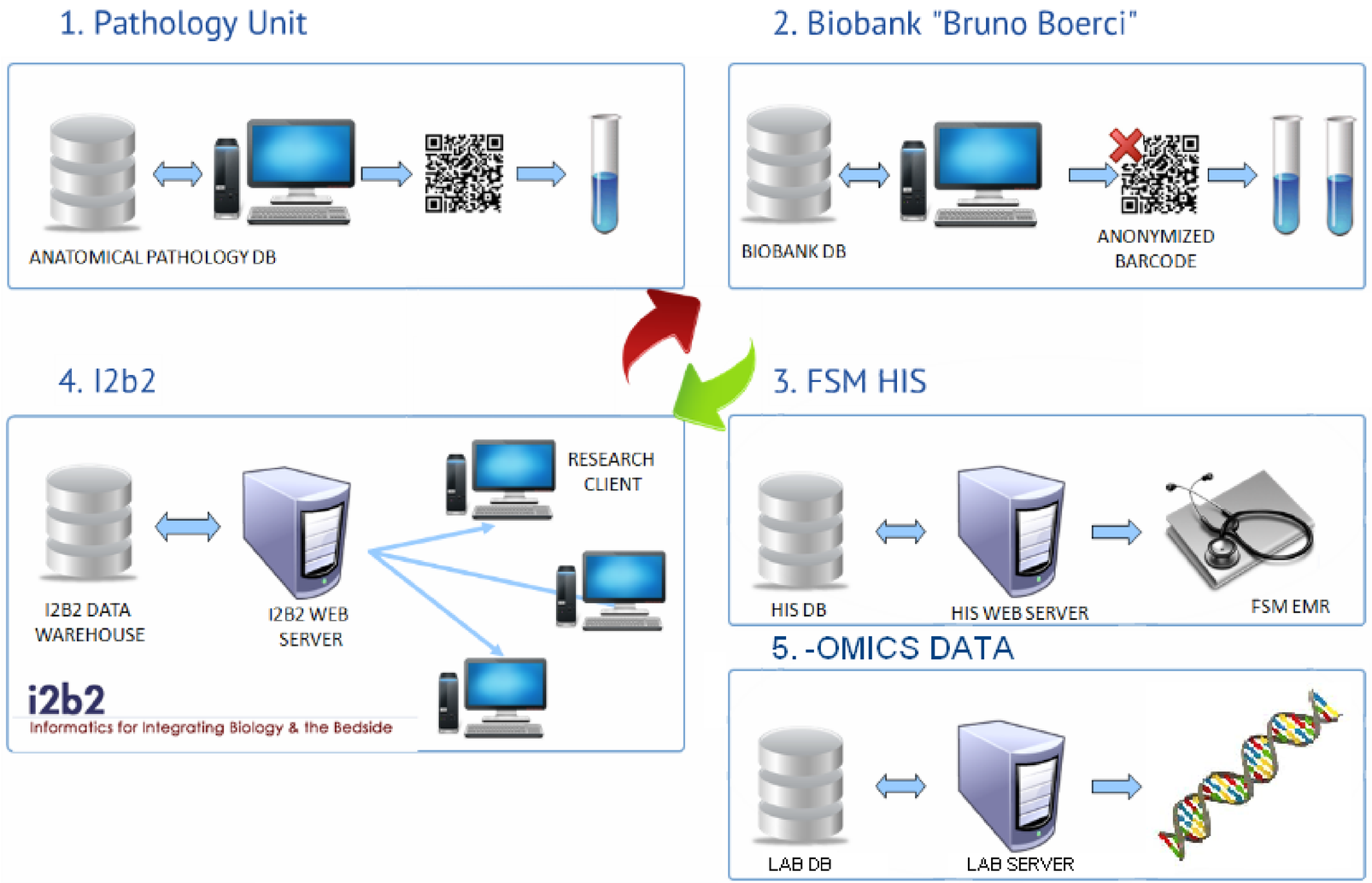




http://catalyst.harvard.edu/spotlights/shrine.html

Onco-i2b2

Laboratory of Biomedical Informatics for Clinical Research
Fondazione Salvatore Maugeri - FSM - Hospital, Pavia, italy



Personalización

7 web plugins.

1 Módulo de integración con R.

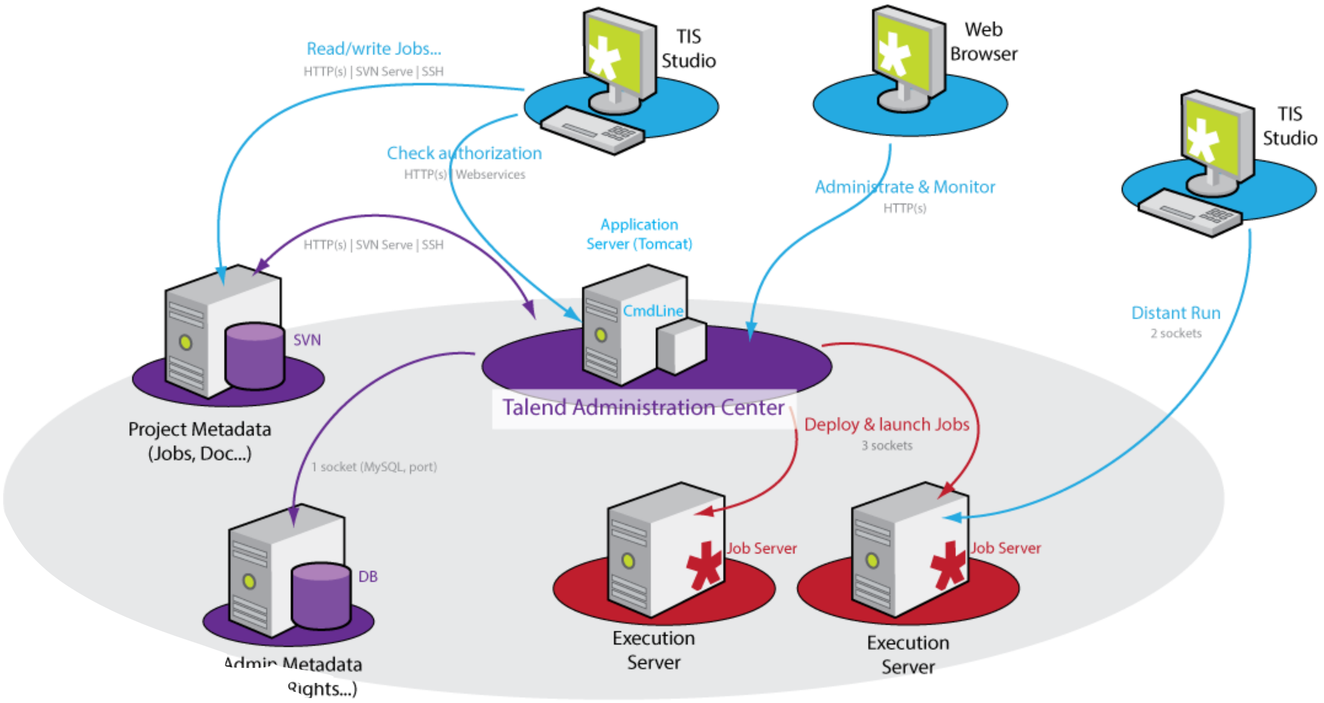
1 Módulo de procesamiento de lenguaje natural.

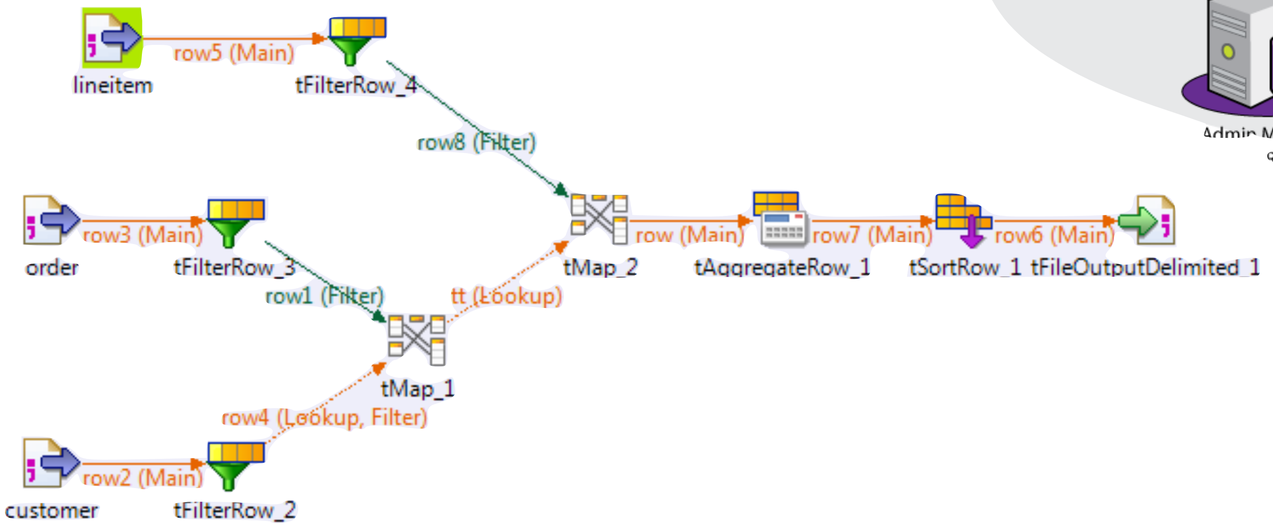
Múltiples tareas ETL.

Oncología (2010): 28.838
Cardiología (2009): 6.334

ETL: Talend Open Studio

- Herramientas de análisis.
- Integración de operaciones / procesos.
- Sincronización o replicación de bases de datos.
- Transformaciones de datos complejos y carga.
- Calidad de datos.





www.talend.com

Conclusiones

- La interrelación de datos clínicos requiere estandarización de la información.
- Procesos de ETL (extraccion, transformación y carga) es determinante en la obtención de datos relevante.
- i2B2 se presenta como la opción mas ampliamente aceptada.
- i2B2 Requiere desarrollo de modelos específicos a la necesidad.
- El gran volumen de datos requiere adopción de plataformas de Big Data para almacenamiento y análisis.

Referencias

[1] L. J. Frey, L. Lenert, and G. Lopez-Campos, “EHR Big Data Deep Phenotyping. Contribution of the IMIA Genomic Medicine Working Group.,” Yearb. Med. Inform., vol. 9, no. 1, pp. 206–11, Jan. 2014.

[2] S. Jauhari and S. Rizvi, “Mining Gene Expression Data Focusing Cancer Therapeutics: A Digest,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 11, no. 3, pp. 533–547, 2014.

[3] M. Lawrence and M. Morgan, “Scalable Genomics with R and Bioconductor,” Stat. Sci., vol. 29, no. 2, pp. 214–226, May 2014.

[4] M. G. de Lecea and M. Rossbach, “Translational genomics in personalized medicine – scientific challenges en route to clinical practice,” Hugo J., vol. 6, no. 1, p. 2, 2012.

[5] McMurry AJ, Murphy SN, MacFadden D, Weber G, Simons WW, et al. (2013) SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies. PLoS ONE 8(3).

[6] A. R. Post, T. Kurc, S. Cholleti, J. Gao, X. Lin, W. Bornstein, D. Cantrell, D. Levine, S. Hohmann, and J. H. Saltz, “The Analytic Information Warehouse (AIW): a platform for analytics using electronic health record data.,” J. Biomed. Inform., vol. 46, no. 3, pp. 410–24, Jun. 2013.

[7] J. F. Hurdle, S. C. Haroldsen, A. Hammer, C. Spigle, A. M. Fraser, G. P. Mineau, and S. J. Courdy, “Identifying clinical/translational research cohorts: ascertainment via querying an integrated multi-source database.,” J. Am. Med. Inform. Assoc., vol. 20, no. 1, pp. 164–71, Jan. 2013.

[8] V. Canuel, B. Rance, P. Avillach, P. Degoulet, and A. Burgun, “Translational research platforms integrating clinical and omics data: a review of publicly available solutions.,” Brief. Bioinform., Mar. 2014.

[9] R. Sabatier, A. Gonçalves, and F. Bertucci, “Personalized medicine: present and future of breast cancer management.,” Crit. Rev. Oncol. Hematol., vol. 91, no. 3, pp. 223–33, Sep. 2014.

[10] R. L. Schilsky, “the future is now,” vol. 9, no. May, pp. 363–366, 2010.

[11] D. Segagni, V. Tibollo, A. Dagliati, A. Zambelli, S. G. Priori, and R. Bellazzi, “An ICT infrastructure to integrate clinical and molecular data in oncology research.,” BMC Bioinformatics, vol. 13 Suppl 4, no. Suppl 4, p. S5, Jan. 2012.

[12] D. Segagni, F. Ferrazzi, C. Larizza, V. Tibollo, C. Napolitano, S. G. Priori, and R. Bellazzi, “R engine cell: integrating R into the i2b2 software infrastructure.,” J. Am. Med. Inform. Assoc., vol. 18, no. 3, pp. 314–7, May 2011.

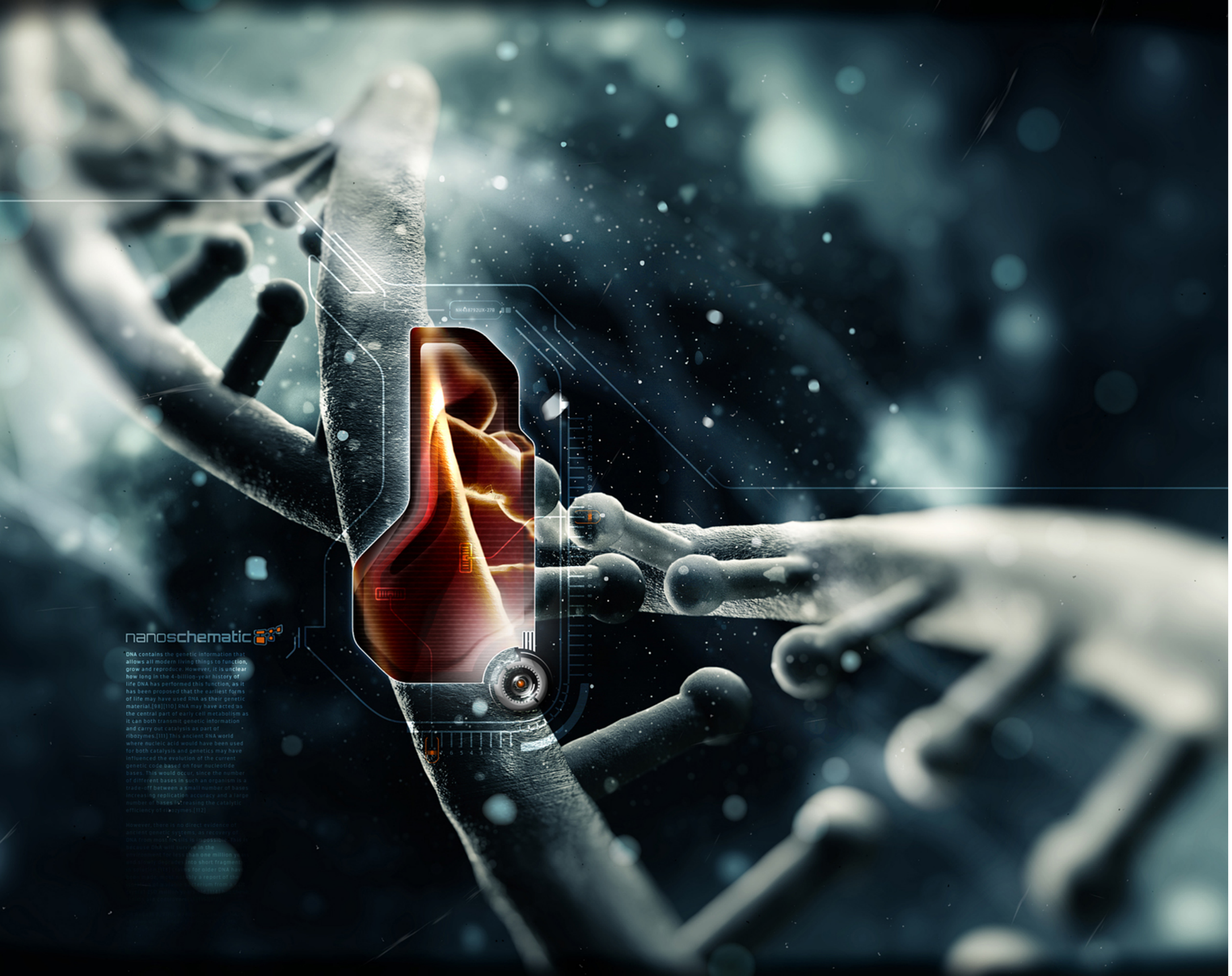
[13] E. Zapletal, N. Rodon, N. Grabar, and P. Degoulet, “Methodology of integration of a clinical data warehouse with a clinical information system : the HEGP case,” in MEDINFO 2010, 2010, pp. 193–197.

[14] M. D. Natter, J. Quan, D. M. Ortiz, A. Bousvaros, N. T. Ilowite, C. J. Inman, K. Marsolo, A. J. McMurry, C. I. Sandborg, L. E. Schanberg, C. a Wallace, R. W. Warren, G. M. Weber, and K. D. Mandl, “An i2b2-based, generalizable, open source, self-scaling chronic disease registry.,” J. Am. Med. Inform. Assoc., vol. 20, no. 1, pp. 172–9, Jan. 2013.

[15] B. Knoppers, “Framework for responsible sharing of genomic and health-related data,” Hugo J., vol. 8, no. 1, p. 3, 2014.

[16] S. Jakka and M. Rossbach, “An economic perspective on personalized medicine,” Hugo J., vol. 7, no. 1, p. 1, 2013.

[17] R. L. Richesson, M. M. Horvath, and S. a Rusincovitch, “Clinical research informatics and electronic health record data.,” Yearb. Med. Inform., vol. 9, no. 1, pp. 215–23, Jan. 2014.



FACULTAD DE MEDICINA
UNIVERSIDAD DE CHILE

Análisis de Plataformas de Integración Clínico-Omicas y Métodos Estadísticos Predictivos para Apoyo al Diagnóstico y Tratamiento

Patricio Araneda García
Magíster (c) en Informática Médica

Prof. Rodrigo Assar

MIM 2014 MAGISTER EN INFORMATICA MEDICA