



CIECTI

Centro Interdisciplinario
de Estudios en Ciencia,
Tecnología e Innovación

BIG DATA APLICADA A LA SALUD

Big Data aplicada al sector Salud en Argentina. Definiendo una agenda de trabajo

Facundo Malvicino

Big Data aplicada al sector Salud en Argentina. Definiendo una agenda de trabajo.

Facundo Malvicino - CIECTI

fmalvicino@ciecti.org.ar

Justificación

El 26 de octubre de 2016 se realizó la primera edición de los Diálogos para el Desarrollo “Big Data aplicada a la Salud” en la Sala de Seminarios “Aldo Ferrer” del Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación (CIECTI). La iniciativa es impulsada por el CIECTI con el objetivo de contribuir al sistema de innovación nacional, dando lugar a procesos innovadores con impacto socioeconómico. En esta edición se buscó alcanzar la cooperación entre productores y usuarios de grandes datos, delineando ejes de discusión sobre las oportunidades y los desafíos en el aprovechamiento de Big Data en Salud.

La actividad arrojó diversos enfoques sobre el abordaje de la problemática planteada, los cuales admiten diferentes niveles de clasificación de acuerdo al grado de especificidad o generalidad de los mismos. Asimismo, reflejó la necesidad de crear capacidades para una adecuada generación y gestión de la Información, en un contexto de alta complejidad como es el del sistema de salud argentino. Por lo tanto, consideramos importante documentar la experiencia y proponer un marco conceptual de trabajo que permita ordenar futuras líneas de trabajo. El presente documento constituye el primer borrador circulado entre los participantes y referentes del tema y se encuentra en revisión y elaboración.

1. Introducción

El término Big Data se implementa a partir del año 2005 como expresión de una nueva fase del paradigma intensivo en información y comunicación¹. Junto a la computación en la nube e Internet de las Cosas (Internet of Things – IoT), constituyen la tercera plataforma de tecnología digital (CEPAL, 2015). Se trata de una nueva fase del paradigma intensivo en información y comunicación, precedida desde el comienzo del nuevo siglo por las innovaciones tecnológicas y los dispositivos digitales. Este período está caracterizado por un incremento exponencial en la cantidad y diversidad de datos digitales disponibles en tiempo real, producto de un mayor uso de equipos tecnológicos en la vida diaria, permitiendo alcanzar un mejor conocimiento del comportamiento humano y su entorno (UN Global Pulse, 2012). El aprovechamiento de los datos digitalizados en mayor frecuencia y variedad para la toma de decisiones, presenta desafíos significativos para la sociedad.

En el caso de salud, los datos digitales provienen de registros médicos electrónicos e imágenes, aunque también pueden considerarse datos farmacológicos, ambientales y hábitos de los pacientes, entre otros (Chen et al 2012; Manyika et al, 2013; Kambatla et al, 2014; Kayyali et al, 2013). De esta manera, Big Data permitiría mejorar la capacidad de respuesta del sistema de salud pública, incrementar la detección temprana de enfermedades y reducir los tiempos de investigación médica, o bien, avanzar hacia una medicina personalizada utilizando los datos de pacientes. Las mejoras en productividad se alcanzarían a partir de la reducción de costos por tratamientos poco efectivos, efectos adversos, errores médicos y criterios sesgados hacia intervenciones no adecuadas; y mejoras en la identificación de patrones en las patologías, monitoreo remoto de los pacientes y sistemas personalizados.

En Argentina, PAMI viene desarrollando el programa ITPAMI donde se promueve la recolección y análisis de datos para una mejor atención y mayor transparencia en el tratamiento de adultos mayores. El impacto de esta política sobre la comunidad es de gran envergadura dada la dispersión geográfica y la dimensión del organismo. El aprovechamiento de grandes datos permitió mejorar la atención médica, detectar fraudes de prestadores, reducir los costos de tratamientos, digitalizar las historias clínicas y mejorar el acceso de los pacientes a sus propios datos, brindando seguridad en la

¹ Para una revisión sobre la literatura y el estado del arte de Big Data en Argentina, consultar Malvicino y Yoguel (2014, 2015)

administración de los mismos. Simultáneamente, hospitales públicos de alta complejidad e instituciones privadas como la FLENI y el Hospital Italiano, están realizando esfuerzos significativos en investigaciones médicas, gestión de la información y capacitación para el análisis de Big Data. Otras instituciones de investigación como el Instituto Nacional del Cáncer (INC), la Fundación Instituto Leloir (FIL), el Instituto de Química Física de los Materiales, Medio Ambiente y Energía (INQUIMAE-FCEN-UBA) y el Instituto de Investigación en Biomedicina de Buenos Aires (IBIOBA)², entre otras, están realizando importantes avances en secuenciación genómica con destacadas vinculaciones internacionales, aunque enfrentando problemas vinculados al personal especializado en analítica de datos y la falta de infraestructura adecuada para el procesamiento y almacenamiento de los datos.

Sin embargo, el aprovechamiento de las oportunidades que ofrece Big Data en el área de salud, demanda el diseño de políticas efectivas orientadas a sectores de alto contenido tecnológico e informacional, mejoras en la vinculación entre demandantes y proveedores de soluciones en Big Data, la formación de trabajadores especializados y nuevas formas de organización del trabajo. En las últimas décadas, Argentina ha impulsado el desarrollo del sector de servicios de informática e información y cuenta con una larga tradición en investigación médica. Sin embargo, se identifican carencias en trabajadores e investigadores especializados que combinen capacidades en informática y salud (ej.: médicos genetistas, bioinformáticos, etc.); también se observan dificultades para llevar a cabo proyectos multidisciplinarios. Si bien existen capacidades técnicas, reales y potenciales, aprovechar este avance informacional-tecnológico a los fines del desarrollo socioeconómico del país, requiere discutir el rol del Sector Público en la definición de políticas de largo plazo y la articulación intersectorial.

El documento se organiza de la siguiente manera. A continuación clasificamos los temas planteados en el Diálogo e identificamos las oportunidades para el desarrollo de proyectos de vinculación entre el sector salud y Big Data. Concluimos el trabajo con una reflexión sobre la experiencia y una agenda de trabajos a futuro.

2. Clasificación de Enfoques e Identificación de Oportunidades

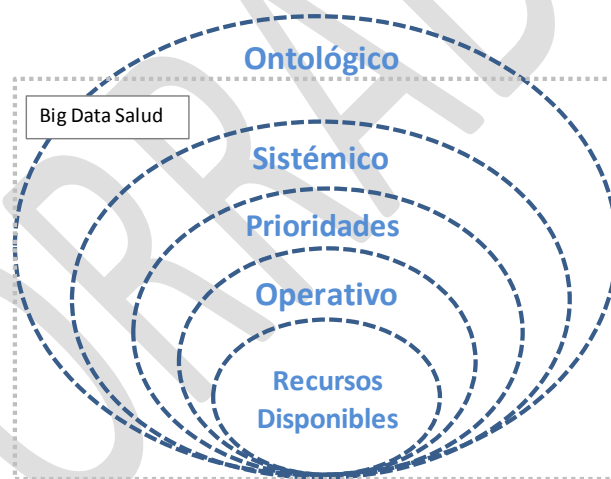
En forma separada, los ejes discutidos en el Diálogo no exceden a la experiencia internacional documentada. La importancia de la calidad y compartición de los datos, los problemas de infraestructura, la necesidad de mejorar la interoperabilidad de los distintos

² Siglas en Anexo

sistemas regionales y por supuesto, los riesgos asociados al uso de los datos personales, son ampliamente abordados por la literatura especializada (Cumbley y Church, 2013; Hilbert, 2013; OECD, 2013; SciDev.Net, 2014; Ciancarini et al, 2016, por mencionar algunos). Por lo tanto, lo interesante de esta actividad es hacer foco en la experiencia local, buscando capturar la riqueza de la problemática y buscando puntos de articulación para alcanzar proyectos de impacto social.

Identificamos cinco tipos de ejes que engloban los enfoques planteados en el Diálogo de acuerdo a su carácter: i) ontológico; ii) sistémico; iii) prioritario; iv) operativo; v) de disponibilidad (capacidades). En la Figura 1 se representa la interacción entre los ejes postulados. Las líneas punteadas indican que las diferentes dimensiones son permeables entre sí, en el marco de la generación y recolección de datos en salud. Los datos sirven y se valen de las dimensiones mencionadas mientras que el plano ontológico excede la discusión estrictamente asociada a Big Data en Salud (recuadro punteado).

Figura 1. Ejes en Big Data Salud



El intercambio entre los distintos planos no debe entenderse como un sistema armonioso. El sistema de salud argentino es particularmente complejo con una desigual difusión tecnológica y de capacidades, por lo que es de esperar que Big Data profundice estas características al mismo tiempo que habilite el desarrollo nuevas oportunidades.

La definición de los ejes se apoya en dos proposiciones: por un lado, la adecuada explotación de Big Data exige haber alcanzado capacidades de alta calificación y organizaciones multidisciplinarias debido al carácter transversal del nuevo paradigma; por otro, la escasa articulación en temas de Big Data Salud entre empresas, miembros del

SNCTI y el sector público, demandan estándares y protocolos para el intercambio de información que promuevan el intercambio de saberes en base a relaciones de confianza y reciprocidad. En el próximo acápite, desarrollamos las características de estas cinco dimensiones, identificando potenciales líneas de trabajo a futuro.

2.1. Ejes de Big Data en Salud

2.1.1. Ontología

Esta dimensión surge como consecuencia de las transformaciones que se producen en la sociedad en relación a la concepción del conocimiento, a partir de los cambios tecnológicos. Evitando todo reduccionismo tecnocrático, Big Data provoca cambios en la creación de conocimiento a partir de nuevas formas de hacer teoría (ej.: taxonomías en epidemiología). Esto representa un cambio epistemológico significativo más allá del reduccionismo que ha ganado fuerza en círculos empresariales, propicio para una narrativa orientada a negocios basados en el conocimiento (ej., *brokers* de datos, consultorías, etc.). Dentro de un marco empirista extremo, la analítica de datos promete un conocimiento “perspicaz, objetivo y rentable”, sin la ciencia o los científicos, los gastos asociados, las contingencias y la búsqueda de explicación. Más allá de la validez de las técnicas de ciencias de datos, esto opera como dispositivo para simplificar un enfoque epistemológico complejo, en base a una estrategia de venta de servicios basados en la analítica de datos.

Durante el Diálogo, surgió la necesidad de abordar este desafío a partir de considerar a Big Data como un fenómeno social-informacional. Esto requiere pensar los conceptos *datos*, *información* y *conocimiento*, en la era Big Data y las implicancias epistemológicas en el campo científico, ya que la mirada sobre estos temas condiciona el abordaje de los problemas y por lo tanto, las soluciones. A un nivel más concreto, se planteó la necesidad de definir las sub-áreas vinculadas a Big Data Salud, ya que con el término *salud* se hace referencia a áreas vinculadas aunque muy disímiles como ser: industria farmacéutica, secuenciación genómica, gestión hospitalaria y diagnóstico clínico, epidemiología, entre otros. Por lo tanto, un posible proyecto de trabajo en este sentido podría ser de orden epistemológico en cuanto a los conceptos información y conocimiento, por un lado, y la adopción de una definición y clasificación de las áreas de aplicación de Big Data en Salud, por el otro.

¿Cómo concebimos la información y el conocimiento? ¿Big Data representa el “fin de la teoría”? ¿Qué es Big Data en Salud y qué áreas abarca?

Probablemente, estas preguntas no puedan ser respondidas con total precisión y requieran de una revisión permanente en base al desarrollo de las experiencias más concretas. Sin embargo es necesario asumir el desafío para delimitar el alcance de los proyectos de campo, proponer políticas públicas y estimar su impacto.

2.1.2. Sistémico

Durante el Diálogo se evidenció la necesidad de coordinar la relación de los actores, la definición de políticas estratégicas que marquen el norte y mayores definiciones en las regulaciones, particularmente en lo que refiere a la protección de datos personales y el uso de los datos.

La definición de áreas de estrategias a ser desarrolladas queda, al menos en última instancia, en manos del Estado. Sin embargo, es posible realizar propuestas en base a la identificación de oportunidades. Por otro lado, organismos descentralizados de apoyo gubernamental pueden tener mayor flexibilidad para promover las relaciones entre los actores, tanto públicos como privados, que desarrollen actividades en Big Data y Salud. Las barreras administrativas pueden demorar proyectos conjuntos que involucren organismos públicos y privados, aunque también pueden ser útiles para dar un buen tratamiento a datos sensibles: el desafío es un equilibrio de modo que se puedan desarrollar proyectos en base a datos sensibles sin caer en el mal uso de los mismos. La dimensión “operativa” (2.1.4) abordará la necesidad de definir estándares y protocolos de buenas prácticas.

Necesidad de políticas públicas que promuevan la coordinación entre actores, la creación de conocimiento en un marco regulatorio que asegure el buen uso de los datos.

Sin embargo, es menester tener en cuenta que Big Data es un proceso histórico comunicacional que involucra a toda la sociedad y los riesgos en cuanto al uso de datos sensibles no pueden ser considerados como un desafío adicional. Se trata de un aspecto intrínseco al fenómeno ya que constituyen la esencia del mismo: si no se “comparten” los datos personales, Big Data no tendría tal importancia.

2.1.3. Prioridades: jerarquías de las problemáticas en salud

Considerando el carácter multidisciplinario del abordaje, el impacto que los cambios en salud pueden generar en la sociedad y la falta crónica de una plataforma homogénea de datos, el eje articulador de enfoques debería ser una serie de problemáticas claves, jerarquizadas de acuerdo a prioridades claramente definidas. De esta manera, el

problema a solucionar se pone por encima de las particularidades y permite avanzar hacia una interdisciplinariedad, donde ninguna disciplina ejerza dominancia absoluta sobre las restantes.

Por lo tanto, se coincide en un esquema dinámico de promoción de la analítica de grandes datos a partir de una demanda específica. Dicho de otra manera, cómo a través de los datos se puede tener un impacto real en la población. A modo de ejemplo, un buen indicador sería conocer las principales causas de mortalidad en Argentina y a partir de ahí identificar cómo los grandes datos pueden contribuir a reducir causas de muerte. Incluso pueden considerarse otras variables para ordenar las prioridades, como las sociales donde la ciencia de datos puede hacer mucho (ej.: mapeo de epidemias, indicar poblaciones vulnerables, etc.). La ciencia de datos permite una mayor focalización de políticas públicas que, a partir de la salud, exceden su área de conocimiento:

Dilucidar los problemas a resolver como estrategia de articulación y a partir de ahí, organizar los datos.

La estrategia de centralidad del problema puede ser abordada en forma sistémica por las autoridades de políticas o bien como modelo de abordaje al problema de investigación.

2.1.4. Operativo

La dimensión operativa contempla los mecanismos que permiten llevar a cabo la generación, captura, limpieza y análisis de datos vinculados a salud. Necesariamente, contempla el carácter comunicacional del proceso al requerir un marco contextual para la compartición de los datos.

Se distinguen al menos dos obstáculos en cuanto a la generación e intercambio de datos en general y particularmente en salud. Por un lado, la calidad de los datos y la disponibilidad para su re-utilización; por otro, la reticencia de los agentes involucrados a compartir los datos, convirtiéndose en silos de información en el marco de una comunicación relativamente menos fluida. Ambos casos tienen sus fundamentos. La calidad en la generación y captura de datos está asociada a la infraestructura disponible y las capacidades de la población tanto para generar el dato, como para realizar las preguntas orientadoras de la captura. En relación a la débil tendencia hacia la compartición de la información, esto involucra riesgos asociados al uso de datos sensibles (particularmente en salud), un marco regulatorio poco definido y la posición de poder que otorga la concentración de información.

Para valorizarse, la información debe circular: mientras más se difunde, mayor es la probabilidad de crear valor, aunque no necesariamente de la apropiación de dicho valor.

En este sentido, durante el Diálogo se planteó la necesidad de desarrollar estándares y protocolos de buenas prácticas que garanticen la calidad de los datos para la toma de decisiones y establezcan un marco de referencia para el desarrollo de relaciones de largo plazo basadas en la confianza. Este posible proyecto requerirá considerar las etapas del análisis de Big Data y los intereses -económicos y no económicos- de los actores involucrados.

2.1.5. Disponibilidad de Capacidades

Esta dimensión puede ser analizada a partir de la infraestructura instalada, la disponibilidad de trabajadores especializados y las bases de datos existentes. La infraestructura en telecomunicaciones es un recurso estratégico para el desarrollo de tecnologías de la información. Las oportunidades vinculadas a los grandes datos están sujetas a los cambios comunicacionales e informacionales que promueve la digitalización masiva de la sociedad, la convergencia tecnológica e Internet. Adicionalmente, la capacidad de almacenamiento y procesamiento es determinante para un mejor desarrollo de la analítica de datos para la toma de decisiones. En este sentido, para países en desarrollo la brecha digital es suplementaria de la brecha tecnológica.

Conocer las limitaciones de la infraestructura tecnológica como recurso estratégico para el desarrollo de capacidades vinculadas a las tecnologías de la comunicación e información.

En segundo lugar, la disponibilidad de trabajadores especializados representa un factor a tener en cuenta para el incremento de productividad y la creación de conocimiento. Durante el Diálogo se manifestó el déficit de especialistas tales como médicos genetistas y bioinformáticos, pero también la necesidad de dotar de herramientas informáticas a todos los profesionales de la salud (ej.: enfermeras). Si bien la oferta académica se ha incrementado en la zona centro del país en los últimos años, la definición de áreas prioritarias en el marco de una política de salud pública digital, permitiría generar los incentivos para orientar la formación de profesionales en estas áreas. Asimismo, es necesario incrementar las capacidades en relación a las herramientas informáticas a nivel población en general, ya que la primera y última etapa (generación y utilización) de la cadena del análisis de grandes datos, involucran a a toda la sociedad.

Definir las capacidades necesarias (cuantificar la cantidad de profesionales especializados, conocimientos mínimos, etc.), una vez que se hayan definido las áreas estratégicas.

Finalmente, existen proyectos de bases de datos nacionales, algunos ya en funcionamiento y otros programados³, en el marco de la promoción de la ciencia abierta por la ley N° 26.899 que regula el desarrollo de repositorios digitales de acceso abierto e insta a los organismos del Sistema Nacional (SN) de Ciencia, Tecnología e Innovación (CTI) a establecer políticas para el acceso público a datos primarios de investigación. Actualmente existen cinco SN de Bases de Datos: del Mar, Biológicos, Climáticos, Repositorios Digitales y Genómicos. Este último fue creado en 2014 (Res. Ministerial N°761/14), cuenta con 13 centros adheridos y 94 conjuntos de datos.

Relevar las bases de datos disponibles, tanto nacionales como regionales, para determinar los datos disponibles, su posible integración y evitar duplicación del trabajo.

3. Reflexiones finales y propuesta de trabajo

El Diálogo en Big Data y Salud resultó una experiencia enriquecedora y dejó planteado el desafío de mantener las actividades de intercambio entre las partes con el objetivo de concretar proyectos en base a los ejes discutidos. Por tal motivo, decidimos documentar la actividad y realizar las propuestas de trabajo que detallamos a continuación. La intención es promover equipos de trabajo en función de los intereses y disponibilidad de los participantes, promoviendo un diálogo abierto. Los posibles proyectos propuestos, sujetos a modificación, serían:

- *Epistemológico/conceptual*: discutir el concepto de información y conocimiento; nuevas taxonomías; la ética en el uso de datos sensibles.
- *Áreas temáticas de Big Data Salud*: definir las áreas Big Data en Salud e identificar las áreas de aplicación de la analítica de grandes datos.
- *Políticas Públicas en Big Data Salud*: identificar áreas de vacancia y proponer políticas públicas para el fomento del análisis de grandes datos en salud, señalando las potencialidades y riesgos que conlleva el uso de datos sensibles.
- *Problemáticas articuladoras*: proponer indicadores para establecer prioridades de atención de problemas como estrategia de articulación y organización de los datos.

³ La iniciativa de gobierno abierto a nivel nacional es una iniciativa que tiene a algo más amplio que lo que establece la Ley N° 26.899, que es más específica para ciencia.
<https://www.argentina.gob.ar/modernizacion/ejes/gobiernoabierto>.

- *Estándares y protocolos de buenas prácticas:* proponer un modelo para intercambio de datos, el resguardo de la calidad y el fomento de relaciones de largo plazo.
- *Recursos disponibles:* reseña de problemas de infraestructura TIC (bastante documentado); evaluación de las capacidades requeridas para la explotación de datos; relevamiento de bases de datos disponibles y fuentes de generación.

Se espera que la conformación de los grupos se formalice de acuerdo a la decisión de los participantes en base al compromiso asumido, pudiendo adquirir la forma de un espacio de intercambio de ideas o bien formalizarse en un proyecto de investigación con posibilidades de financiamiento.

BORRADOR

Sobre el presente documento

Este documento es el resultado del encuentro entre investigadores y profesionales de las Ciencias de Datos y el Sistema de Salud argentino, convocados por el CIECTI a la primera edición de los Diálogos para el Desarrollo “Big Data aplicada a la Salud”. Participaron del diálogo representantes de las siguientes instituciones:

Centro de Investigaciones en Bionancias (CIBION) - CONICET

Departamento de Informática – Hospital Italiano

Facultad de Ingeniería de la Universidad Católica de Córdoba

Facultad Latinoamericana de Ciencias Sociales (FLACSO)

Fundación Argentina de Nanotecnología (FAN)

Fundación Sadosky

Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA)

Instituto de Química Física de los Materiales, Medio Ambiente y Energía (INQUIMAE) - UBA – CONICET

Instituto de Salud Colectiva de Lanús - UNLa

Maestría de Explotación de datos y descubrimiento del Conocimiento, FCEN, UBA.

Ministerio de Ciencia y Tecnología de la Provincia de Córdoba

Plataforma Nacional de Bioinformática

Programa SUMAR - Ministerio de Salud de la Nación

Referencias bibliográficas

- CEPAL, N. (2015). "La nueva revolución digital: de la Internet del consumo a la Internet de la producción".
- Chen, C. L., y Chun-Yang Zhang (2014). "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data." Information Sciences. 2014.
- Ciancarini, Paolo, Francesco Poggi, y Daniel Russo (2016). Big Data Quality: A Roadmap for Open Data. 210–15. IEEE. doi:10.1109/BigDataService.2016.37.
- Cumbley, Richard y Peter Church (2013). "Is "Big Data" creepy?". Computer Law & Security Review 29.5: 601-609. Londres 2013
- Hilbert, Martin (2013) "Big Data for Development: From Information- to Knowledge Societies" United Nations ECLAC. Enero 2013.
- Kayyali, Basel, Knott, David y Van Kuiken, Steve (2013). The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey & Company, 1-13.
- Malvicino, F. y Yoguel, G. (2014). [Big Data. Avances Recientes a Nivel Internacional y Perspectivas para el Desarrollo Local. Documento de Trabajo N° 3](#). Centro Interdisciplinario de Estudios en Ciencia Tecnología e Innovación (CIECTI). Buenos Aires.
- Malvicino, F. y Yoguel, G. (2015). Descubriendo Big Data en Argentina. Encuesta Digital 2014. AGRANDA 1era ed. 44va Jornadas de Informática (JAIIO). Rosario, septiembre 2015. ISSN: 2451-7569
- Manyika, James, Chui, Michael, Brown, Brad, Bughin, Jacques, Dobbs, Richard, Roxburgh, Charles y Byers, Angela Hung. (2011). Big data: The next frontier for innovation, competition, productivity. McKinsey Global Institute. Report May 2011.
- OECD (2013), "Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"", OECD Digital Economy Papers, No. 222, OECD Publishing
- SciDev.Net (2014). "Big Data for Development. Big obstacles ahead for big data for development". SciDev.Net, abril 2014. Recuperado el 19/12/2016.
- UN Global Pulse (2012). Big Data for Development: Challenges & Opportunities. United Nations. Mayo 2012.


Anexo. Siglas Utilizadas

- CEPAL: Comisión Económica para América Latina y el Caribe
- CIECTI: Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación
- CTI: Ciencia Tecnología e Innovación
- FCEN: Facultad de Ciencias Exactas y Naturales
- FIL: Fundación Instituto Leloir
- FLENI: Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia
- IBioBA: Instituto de Investigación en Biomedicina de Buenos Aires
- INC: Instituto Nacional del Cáncer
- INQUIMAE: Instituto de Química Física de los Materiales, Medio Ambiente y Energía
- OECD: Organisation for Economic Co-operation and Development
- PAMI: Programa de Atención Médica Integral - Obra social de adultos mayores del Instituto Nacional de Servicios Sociales para Jubilados y Pensionados
- TIC: Tecnologías de Información y Comunicación.
- UBA: Universidad de Buenos Aires



CIECTI

Centro Interdisciplinario
de Estudios en Ciencia,
Tecnología e Innovación

Seguinos en  @ciecti
Buscanos en  /ciecti

Godoy Cruz 2390 - PB (C1425FQD), CABA,
[54-11] 4899-5000, int. 5684
info@ciecti.org.ar

ciecti.org.ar