

AGRANDA 2015, 1º Simposio Argentino de Grandes Datos.

Descubriendo Big Data en Argentina. Encuesta Digital 2014

Facundo Malvicino, CIECTI

Gabriel Yoguel, UNGS

Abstract. El objetivo del presente documento es presentar los resultados preliminares de la Encuesta Big Data (2015) realizada desde el CIECTI a empresas del sector y grupos de investigación en el país. Se trata de un trabajo en curso que requiere ampliar la muestra de encuestados. En esta primera etapa buscamos caracterizar el sector en base a una selección de actores del sector académico y empresario, esbozando un primer diagnóstico en el marco de las políticas de desarrollo argentinas y el rol del Sector Público como dinamizador.

Antecedentes del tema o contexto

Big Data surge como una nueva fase del paradigma intensivo en información y comunicación que abarca no sólo su dimensión tecnológica, sino también una dimensión social, económica, política y cultural. Por lo tanto, una comprensión holística sobre el tema requiere un abordaje multidisciplinario. Al igual que con el surgimiento de las Tecnologías de la Información y Comunicación (TIC), existe una discusión acerca de si nos encontramos frente a un cambio de paradigma tecnológico, una nueva revolución industrial-tecnológica o si Big Data no ha hecho más que intensificar la competencia en mercados ya existentes, sin mostrar un impacto en las mejoras de la productividad general. Los autores que tienen una visión optimista, sostienen que Big Data estaría provocando cambios en la organización del trabajo con mejoras en la toma de decisiones y el acceso a los mercados. Estos cambios darían lugar a una reducción en las brechas de productividad a partir de una mayor competitividad de las unidades productivas con un impacto positivo sobre la calidad de vida de las personas (McKinsey, 2011; Brynjolfsson, Hitt, y Kim, 2011; UNGPulse, 2012; Chen y Chun-Yang, 2014). Entendido como un modelo de negocios, Big Data constituye un fenómeno global con un impacto económico real y potencial, a partir del aumento de la productividad del trabajo. La viabilidad de este proceso está sujeta a los cambios comunicacionales e informacionales que promueve la digitalización, la convergencia tecnológica e Internet a distintos niveles: a nivel operativo se trata del acceso remoto y en tiempo real a la información para la toma de decisiones, permitiendo un desanclaje territorial; los cambios estructurales tienen que ver con la mayor y mejor accesibilidad comunicacional (conectividad); mientras que nuevos niveles de propagación permiten variaciones en la escala comunicacional (Forte et al., 2012).

En las últimas décadas Argentina ha desarrollado el sector de Servicios de Informática e Información posicionándolo entre los más dinámicos de la región. Por lo tanto, existen capacidades técnicas reales y potenciales que permitirían aprovechar este avance informacional-tecnológico a los fines del crecimiento y desarrollo socio-económico del país (Barletta et al., 2013; CESSI, 2014). Dadas estas características, Argentina cuenta con la oportunidad de crear una plataforma para una industria de

Big Data, promoviendo políticas convergentes hacia un uso intensivo de datos digitales y TIC. Sin embargo, de acuerdo a los resultados de la Encuesta Big Data (2015), el mercado de este sector en Argentina muestra una demanda escasa y de baja complejidad, explicado en parte por la falta de conocimiento sobre el tema, problemas institucionales y limitaciones en la infraestructura. Simultáneamente, las instituciones científicas muestran una desconexión entre la producción académica y los desarrollos de las empresas locales. Frente a este desafío surgen las siguientes preguntas: ¿se encuentra el país en condiciones de aprovechar los beneficios que promete esta nueva fase digital?; si el análisis de grandes datos requiere de una base tecnológico-comunicacional: ¿está el país en condiciones de dar respuesta a esta exigencia?; ¿cómo impacta el escaso desarrollo local de infraestructura de alta complejidad para el desarrollo de Big Data?

Metodología de análisis

La Encuesta Big Data (2015) consistió en la realización de dos encuestas, una destinada a los grupos de investigación que se desempeñan en instituciones científicas y tecnológicas y otra dirigida a empresas, todas vinculadas a actividades de Big Data. Al no existir información sistematizada sobre el universo de grupos de investigación y de empresas, se requirió conformar los padrones a partir de la información provista por el CIECTI y la Fundación Sadosky, los cuales fueron utilizados en la implementación del operativo de las encuestas.

Se diseñaron los formularios de las encuesta en formato digital conteniendo preguntas abiertas y cerradas a fin de caracterizar a los grupos de investigación y a las empresas, mediante la conformación de indicadores generales e información particular de cada una de ellas. Esta actividad fue realizada por personal de la Dirección Nacional de Información Científica (DNIC-MinCyT) junto a los expertos de las instituciones ya mencionadas. Posteriormente se invitó a participar a todos los grupos de investigación y las empresas de los padrones conformados, dirigiéndose especialmente para su llenado a los directores y a los responsables de las actividades de Big Data, respectivamente.

La implementación de las encuestas permitió obtener respuestas completas de 15 grupos de investigación y de 18 empresas, representado el 75% y el 51% de los padrones, respectivamente. Es necesario destacar que la implementación del operativo no respondió a criterios muestrales de representatividad estadística. Por lo tanto, los resultados que se generarán corresponden solamente a los grupos y empresas que respondieron la encuesta.

Definición de Big Data

Por tratarse de un fenómeno reciente, Big Data no se halla todavía suficientemente delimitado, aunque podemos encontrar varios puntos en común entre las definiciones planteadas por la comunidad académica, los técnicos y los profesionales especializados en el área. McKinsey (2011: 1) ofrece una definición “intencionalmente subjetiva” y dinámica a partir de las limitaciones tecnológicas en la actualidad: “Big Data

refiere a un conjunto de datos cuyo tamaño está más allá de la capacidad que tienen los software de base de datos para capturar, almacenar, administrar y analizar”. De esta manera, a medida que la frontera tecnológica se va modificando, el tamaño del conjunto de datos que califica como Big Data también irá cambiando. Hablamos de Big Data cuando el tamaño se vuelve el principal problema. Si bien el volumen es un punto en común entre los especialistas que definen Big Data, otros conceptos incorporan más dimensiones al análisis. Gartner (2011) y UN Global Pulse (2012) suelen caracterizar el tema a partir de los desafíos que Big Data representa no sólo a partir del volumen, sino también de la variedad y la velocidad de la información (3Vs), incluso dándole mayor importancia a estas últimas. Con variedad se suele hacer referencia a la heterogeneidad de la representación e interpretación semántica de los datos, es decir que no están directamente listos para ser integrados a una aplicación; mientras que con velocidad (o timeliness según el término en inglés) se hace referencia tanto a la frecuencia de los datos como al tiempo de respuesta: la importancia reside en la velocidad del feedback, utilizando los datos para la toma de decisiones. En algunas ocasiones también se incluye una cuarta V que remite a valor, variabilidad o virtual (Armah, 2013; OECD, 2013; Chen y Chun-Yang 2014). Según la publicación White Paper (2012) restaría incluir las dimensiones de privacidad de las personas y facilidad de uso (usability) para una cabal comprensión del fenómeno.

Principales Resultados

Al ser consultadas sobre la complejidad de la demanda de proyectos de Big Data en Argentina, el 63% (sobre 16) de las empresas respondieron que es de complejidad media; el 25% respondió que es de baja complejidad y sólo una empresa la considera de alta complejidad. Las principales razones brindadas por los participantes remiten al volumen de datos generados, el desconocimiento sobre el tema y el desarrollo incipiente de proyectos. Respecto a países desarrollados el volumen generado es bajo, incidiendo las deficiencias de infraestructura de procesamiento, almacenamiento y comunicaciones. En segundo lugar, los posibles usuarios del análisis de Big Data desconocen de qué se tratan estos desarrollos y los potenciales beneficios, mientras que las empresas que pueden ofrecer servicios relacionados encuentran dificultades a la hora de mostrar el retorno de las inversiones (ROI, por sus siglas en inglés) de los proyectos. En tercer lugar, nos encontramos en las primeras etapas de proyectos vinculados a Big Data, los cuales muestran todavía baja complejidad. Esta característica no difiere de las primeras etapas del sector TIC en Argentina. Un enfoque integral de la oferta y la demanda fue realizada por Borello et al (2005).

Esta característica del mercado local es reforzada por el destino de los productos y servicios relacionados, ya que el 46% de los clientes son empresas privadas nacionales y otro 46% son firmas extranjeras. La alta participación de empresas privadas extranjeras en la facturación de un sector emergente indicaría la existencia de capacidades locales competitivas orientadas hacia el exterior por falta de un mercado interno suficientemente demandante. Esto implica una débil co-evolución entre la demanda local y la oferta de productos y bienes en Big Data que puede explicarse por falta de capacidades desde la demanda e instituciones poco desarrolladas. La competi-

tividad de las firmas les permite insertarse en el mercado internacional, pero a partir de salarios bajos. Por lo tanto, una demanda local sostenida y diferenciada de estos proyectos se vería acompañada de un aumento de la productividad de las empresas a partir de aprender haciendo, mejorando incluso su inserción internacional.

En cuanto a la forma en que surgen los proyectos, las firmas manifestaron el sesgo privado de la actividad o bien la escasa participación del sector público en la demanda de proyectos de Big Data. Los resultados muestran que en 11 (69%) de 16 empresas los proyectos surgen a partir de una idea generada por la empresa y que luego intenta vender; asimismo, 11 firmas inician los desarrollos a partir de las demandas de clientes o usuarios que se contactan para resolver un problema puntual; mientras que sólo una empresa mencionó responder a demandas del sector público. Esto último encuentra dos explicaciones posibles. Por un lado, se trata de un área nueva y la producción de grandes volúmenes de datos todavía es incipiente en el sector público o se desconocen las potenciales aplicaciones de los mismos: el nivel de incertidumbre sobre Big Data estaría limitando la demanda de estos productos. Por otro lado, se trataría de un problema institucional: la característica transversal de este subsector afecta las actividades de diferentes organismos públicos, lo cual requiere de coordinación y comunicación entre los mismos. Estos resultados indican la existencia de margen para la intervención dinamizadora del Estado, demandando proyectos que redunden en desafíos de aumento de la productividad en sectores considerados estratégicos.

Otro tema de consulta fueron las políticas públicas para el estímulo de proyectos del sector. Se les solicitó a los participantes que jerarquizaran en base a sus criterios cinco posibles ejes de políticas públicas para impulsar el sector. Las prioridades para ambos grupos son la Oferta académica y formación de profesionales en Ciencia de Datos y los Incentivos fiscales y financiamiento en primer lugar, seguido por los Desarrollos de proyectos desde el sector público. Las Políticas de Datos Abiertos y el Marco regulatorio y los derechos de propiedad son los menos valorados, indicando preferencias por políticas públicas de incentivos antes que defensivas.

Finalmente, la red de vinculaciones entre las empresas y los grupos de investigación entrevistados muestra poca densidad. Las 14 firmas que respondieron dieron indicios de una red poco desarrollada todavía. La mitad de las firmas participantes afirmaron no haber trabajado conjuntamente ni con empresas ni con centros de investigación; el 29% (4) se vincularon con instituciones de Ciencia y Tecnología; mientras que sólo el 14% (2) declararon relaciones productivas con empresas nacionales y extranjeras en cada caso. Una sola empresa mantuvo relaciones tanto con grupos de investigación como empresas extranjeras (aunque no se especifica que haya sido para el mismo proyecto). En general, las firmas muestran un buen conocimiento de su entorno en cuanto a firmas de Big Data, aunque poco trabajo conjunto. Por su parte, los grupos de investigación declararon vincularse con siete empresas locales. Los principales obstáculos identificados para formalizar una actividad conjunta son factores institucionales, reglamentación y tiempos de gestión, ya sea porque las contrataciones están directamente prohibidas o porque los tiempos de gestión exceden los que la industria suele considerar razonables. La vinculación de grupos y empresas es un factor importante para la transferencia de conocimiento y apropiación local de los mismos. Si bien es deseable la vinculación de grupos de investigación con firmas o

grupos extranjeros, se corre el riesgo de que la capitalización de los resultados de las investigaciones sea realizada en otros países (Kern 2010). A esto debemos agregar que según la Encuesta Big Data (2015), la relación entre aplicaciones de los proyectos de investigación y los desarrollos de empresas privadas muestra desajustes. Las diferencias extremas se hallan en el Sector Financiero con mayor producción por parte de las empresas; mientras que los grupos de investigación cuentan con mayor presencia en el Sector Agropecuario/Agroindustrial. Tanto grupos como firmas persiguen una lógica diferente. Los primeros estarían abocados a la publicación de documentos, mientras que las firmas se orientan a la búsqueda de rentabilidad.

Bibliografía

- Armah, N. A. (2013). "Big Data Analysis: The Next Frontier". Bank of Canada Review, 2013 (Summer), p. 32-39.
- Barletta, Pereira, Robert y Yoguel (2013), "Argentina: Dinámica reciente del sector de software y servicios informáticos", Revista CEPAL N° 110. Agosto 2013.
- Barletta, F., Pereira, M., Rodríguez, S., Yoguel, G., (2015). "Performance of ICT research groups from Argentina: Scientific vs. Transfer Activities?" Documento Interno Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación (CIECTI). Buenos Aires.
- Borello, J., Erbes, A., Robert, V., Roitter, S., & Yoguel, G. (2005). Competencias técnicas de los trabajadores informáticos. Revista de la CEPAL, 87, 131.
- Brynjolfsson, Erik and Hitt, Lorin M. and Kim, Heekyung Hellen (2011). "Strength in Numbers: How Does Data-Driven Decision making Affect Firm Performance?" Abril 2011.
- CESSI (2014). "Reporte semestral sobre el Sector de Software y Servicios Informáticos de la República Argentina". Observatorio Permanente de la Industria del Software y Servicios Informáticos de la República, Cámara de Empresas de Software y Servicios Informáticos. Agosto 2014.
- Chen, C. L., y Chun-Yang Zhang (2014). "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data." Information Sciences. 2014.
- Encuesta Big Data (2015). Primera Encuesta a Grupos de Investigación y Empresas de Big Data. Argentina 2014. CIECTI – Ministerio de Ciencia, Tecnología e Innovación, Buenos Aires 2014.
- Forte, M. Á., Pignuoli Ocampo, S., Calise, S., Palacios, M., & Zitello, M. (2012). "Las TIC como problemática de la teoría sociológica. Una problematización teórica de la comunicación digitalizada desde la teoría general de sistemas sociales autorreferenciales y autopoieticos." Revista de la Carrera de Sociología. Entramados y Perspectivas, (2), 205-226.
- Gartner (2011). "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Junio 2011
- Kern, A. (2010). Relaciones entre ciencia, tecnología y política en procesos de cooperación internacional. Análisis de un caso de cooperación entre Argentina y

Alemania en el campo de las Tecnologías Informáticas (Disertación Doctoral, Tesis de Doctorado en Ciencias Sociales de FLACSO-Argentina).

- McKinsey (2011). “Big data: The next frontier for innovation, competition, and productivity”. McKinsey Global Institute. Junio 2011.
- Milesi, D., Petelsky, N., Verre, V. (2014). Apropiación privada de los resultados de innovación. Tópicos de economía evolutiva. Barletta, F., Roberts, V., Yoguél, Y. (comp.). Ed. Miño y Dávila y UNGS, Buenos Aires.
- 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.
- Saviotti, P. P. (2001). Variety, growth and demand. *Journal of Evolutionary economics*, 11(1), 119-142.
- Saviotti, P. P., & Pyka, A. (2004). Economic development by the creation of new sectors. *Journal of evolutionary economics*, 14(1), 1-35.
- UN Global Pulse 2012. “Big Data for Development: Challenges & Opportunities” United Nations. Mayo 2012.
- White Paper (2012): “Challenges and Opportunities with Big Data. A community white paper developed by leading researchers across the United States”. Documento creado a partir de conversaciones mantenidas entre miembros de universidades de EEUU (MIT, UC Berkeley, Stanford Univ., etc.) y empresas tecnológicas líderes (Yahoo!, Google, Microsoft, IBM, HP). Nov. 2011-Feb. 2012.