

AGRANDA 2016, 2° Simposio Argentino de Grandes Datos

## Propuesta para una Infraestructura de Datos Agropecuarios del Instituto Nacional de Tecnología Agropecuaria (INTA)

Yanina Bellini Saibene

Estación Experimental Agropecuaria Anguil. Ruta Nacional N° 5, Km 580, (6326) Anguil, La Pampa, Argentina. [bellini.yanina@inta.gob.ar](mailto:bellini.yanina@inta.gob.ar)

**Keywords:** integración de datos, datos abiertos, big data, IDE, geodatabase

El INTA como institución de ciencia y técnica produce datos que se pueden enmarcar en el paradigma de Big Data definido por las 3 Vs [1]: 1) Volumen: grandes cantidades de datos, debido a la continuidad de actividades (Ej: ensayos territoriales, red agro meteorológica, laboratorios, etc.) a la escala (desde la parcela al país) y potenciado por nuevos métodos de adquirir, almacenar y transferir datos (Ej: monitores de rendimientos, dispositivos móviles, drones); 2) Velocidad: dispositivos que registran un flujo continuo de datos en tiempo casi real (ej: radares, estaciones automáticas); 3) Variedad: la diversidad de temáticas y las particularidades territoriales, hacen que los datos tengan orígenes dispares y naturaleza multivariada, algunos sistematizados y otros no. Esta complejidad en su gestión lleva a: 1) integración y acceso difícil, imposibilitando, en muchos casos, que los datos presenten una imagen completa de la realidad a analizar; 2) subutilización de este costoso y valioso activo de información histórica, limitando la capacidad de apoyar al análisis y a la toma de decisiones y 3) pérdida de información por degradación normal a través del tiempo. Estos datos tienen un enorme potencial si se mejora su resguardo, acceso y análisis, tanto para las decisiones de corto plazo de las diferentes audiencias de INTA en el manejo de las actividades productivas, como para el mediano y largo plazo asistiendo a la planificación estratégica y a la formulación de políticas, permitiendo analizar la historia, modelarla y generar escenarios potenciales.

En este contexto se propuso desarrollar La Infraestructura de Datos Agropecuarios (IDA) del INTA cuya visión es generar *La más amplia colección de datos técnicos del INTA archivada, catalogada y accesible al mayor número posible de usuarios* y sus objetivos: 1) Integrar datos, metadatos, servicios e información del tipo agropecuario para facilitar y promover su uso; 2) Generar el conjunto tecnologías, políticas, estándares, herramientas y estructura para adquirir, procesar, almacenar, distribuir y mejorar el uso de la información agropecuaria institucional tanto dentro como fuera del INTA y 3) Promover el acceso abierto y completo a los datos del INTA. La figura 1 presenta la arquitectura de la IDA. Debido a que la diversidad y demanda de los usuarios de datos de INTA se ha incrementado, no es posible satisfacerlos publicando sólo datos “crudos, es necesario estructurar los datos en productos de datos (tabla 1) que tienen diferente niveles de procesamiento y agregado de valor, de acuerdo a las necesidades de las diversas audiencias. La IDA se encuentra en desarrollo y se ha

avanzado en pruebas de conceptos de integración de sistemas [3] y el desarrollo de soluciones de almacenamiento y catalogación de datos para la red de radares meteorológicos del INTA [4].

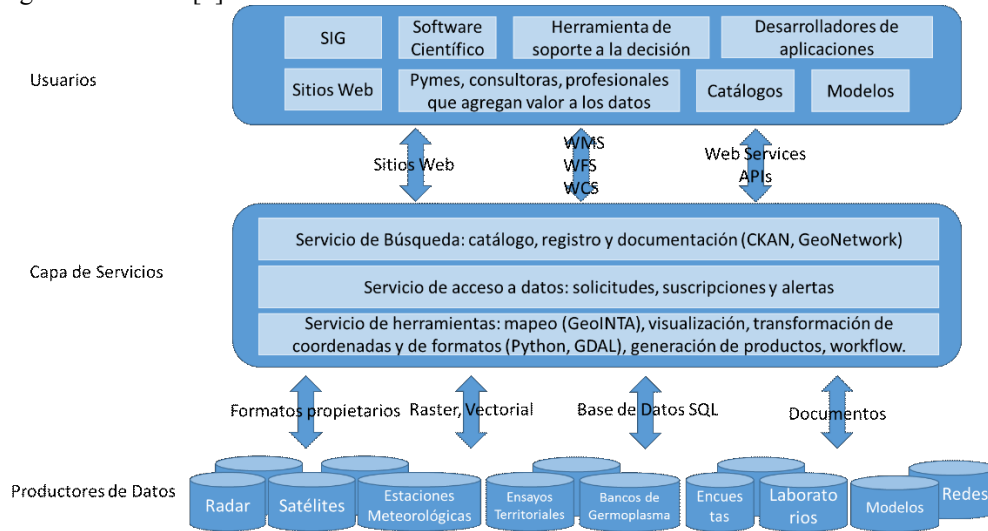


Fig. 1. Arquitectura de la IDA armada en base a [2]

Table 1. Productos de datos con diferentes niveles de acuerdo a sus características

Nivel 0	Los datos «crudos» de registros de algún instrumental o personas. Ej: volúmenes de los radares.
Nivel 1	Datos calibrados generalmente de un solo instrumento/observador/área de muestreo. Incluyen información de calidad de los datos. Ej: resultados de redes de ensayos.
Nivel 2	Datos con procesamiento básico (filtros, agrupamientos, estadísticas básicas) y herramientas utilizadas para ese procesamiento. Ej: cantidad de cabezas de bovinos de carne, por categoría.
Nivel 3	Datos de Nivel 1 y/o 2 mapeados en una cuadrícula espacio-tiempo uniforme.
Nivel 4	Derivados de los niveles 1, 2 y/o 3. Pueden combinar observaciones de diversas fuentes. Incluyen herramientas para trabajarlos. Ej: Mapas de fósforo extractable, materia orgánica y pH.
Nivel 5	Herramientas de visualización, análisis, modelado y sus resultados utilizando los productos de los niveles 1 a 4. Ej: Raster de 1 km <sup>2</sup> con clasificación de ocurrencia de granizo, cálculo de superficie afectada, cruce con información de uso de la tierra, cálculo de posible impacto económico.

1. S. Sagiroglu and D. Sinanc, "Big data: A review," Collaboration Technologies and Systems (CTS), 2013 International Conference on, San Diego, CA, 2013, pp. 42-47.
2. Davis, S. D. NOAA Environmental Data Management Framework. 2012.
3. Sistema Integrado de Información Institucional. <http://rian.inta.gov.ar/prets/>.
4. Divan, M., Bellini Saibene, Y. N., Martín, M. D. L. A., Belmonte, M. L., Lafuente, G., & Caldera, J. M. (2015). Hacia una Arquitectura de Procesamiento de Datos del RADAR Meteorológico de INTA Anguil. In Jornadas Argentinas de Informática. 44. Simposio de Informática en el Estado. 9. 2015.