

## Hacia una Arquitectura de Procesamiento de Datos del RADAR Meteorológico de INTA Anguil

Mario Diván<sup>1</sup>, Yanina Bellini<sup>2</sup>, María de los Ángeles Martín<sup>1</sup>, María Laura Belmonte<sup>2</sup>, Guillermo Lafuente<sup>1</sup> & Juan Marcelo Caldera<sup>2</sup>

<sup>1</sup> Facultad de Ingeniería, UNLPam, General Pico, La Pampa, Argentina

<sup>2</sup> INTA, Estación Experimental Anguil, La Pampa, Argentina.

{mjdivan, martinma, lafuente}@ing.unlpam.edu.ar  
{bellini.yanina, belmonte.maria, caldera.juan}@inta.gob.ar

**Abstract.** El Radar Meteorológico (RM) de la Estación Experimental Agropecuaria INTA Anguil produce en forma diaria un volumen de 17GB de datos, lo cual representa anualmente alrededor de 6,2 Tb. La utilización de dichos datos al momento en que son generados, como así también su posterior gestión, utilización y la posibilidad de brindar servicios al público representan un desafío por su volumen y complejidad. El enfoque integrado de procesamiento de flujos de datos centrado en metadatos de mediciones (EIPFDcMM), es un gestor de flujos de datos sustentado en un marco de medición y evaluación, el cual incorpora comportamiento detectivo y predictivo, mediante el empleo de las mediciones y metadatos asociados. Este trabajo propone una arquitectura de procesamiento que extiende el EIPFDcMM para incorporar el procesamiento de grandes datos almacenados. Ello permitiría dotar al RM de un comportamiento detectivo y predictivo sobre los datos en línea, como así también incluir una capa de servicios al público, que fomente el consumo de datos generados por el RM de INTA Anguil.

**Keywords:** Grandes Datos, Teledetección, Flujo de Datos, Memoria Organizacional.

### 1 Introducción

Actualmente, existen aplicaciones que procesan un conjunto de datos a medida, generados en forma continua, a los efectos de responder a consultas y/o adecuar su comportamiento en función del propio arribo de los datos [1], como es el caso de las aplicaciones para el monitoreo de signos vitales de pacientes; del comportamiento de los mercados financieros; entre otras. En dicho tipo de aplicaciones, se enmarca el Enfoque Integrado de Procesamiento de Datos centrado en Metadatos de Mediciones

(EIPFDcMM) [2,3], el cual sustentado en el marco de medición y evaluación C-INCAMI (*Context-Information Need, Concept model, Attribute, Metric and Indicator*) [4,5], incorpora metadatos al proceso de medición, promoviendo la repetitividad, comparabilidad y consistencia del mismo. Desde el punto de vista del sustento semántico y formal para la medición y evaluación (*M&E*), el marco conceptual C-INCAMI establece una ontología que incluye los conceptos y relaciones necesarias para especificar los datos y metadatos de cualquier proyecto de M&E. Por otra parte, y a diferencia de otras estrategias de procesamiento de flujos de datos [6,7,8], gracias a la incorporación de metadatos, el EIPFDcMM es capaz de guiar el procesamiento de las medidas provenientes de fuentes de datos heterogéneas, analizando cada una de ellas dentro de su contexto de procedencia, como así también su significado dentro del proyecto de M&E en el que se definió.

La Estación Experimental Agropecuaria (EEA) INTA Anguil tiene instalado un RM capaz de generar un flujo de datos de 17GB diarios aproximadamente, lo que representa un desafío para su almacenamiento, gestión y posterior servicio al público, principalmente considerando la importancia que los datos poseen para la región productiva de influencia.

Por otro lado, y si bien el EIPFDcMM incorporó en su esquema de procesamiento la posibilidad de manejar grandes volúmenes de datos como memoria organizacional para el entrenamiento de sus clasificadores [9], no estaba orientado al acopio de la totalidad de los datos procesados en línea y menos aún, al hecho de contar con la posibilidad de integrar la prestación de servicios de datos públicos a terceros.

Así, y como contribuciones específicas se plantea, (i) *relacionado con la configuración de las fuentes de datos*: la posibilidad de incorporar dentro de la arquitectura de procesamiento del EIPFDcMM como una fuente heterogénea al RM de INTA Anguil en forma transparente, lo que permite incorporar un comportamiento detectivo y predictivo, a partir del proyecto de M&E que los expertos del INTA definan sobre las medidas observadas; (ii) *relacionado con la captación de medidas*: la posibilidad de definir formalmente un proyecto de M&E que permite detectar eventuales desvíos del funcionamiento del RM y promover los ajustes meteorológicos pertinentes, basando los cursos de acción probables a partir de la información disponible en la memoria organizacional; (iii) *relacionado con la arquitectura de procesamiento*: ahora se tiene una arquitectura actualizada para el EIPFDcMM que permite incorporar la gestión y utilización de grandes volúmenes de datos persistentes junto con el procesamiento de flujos de datos; (iv) *relacionado con los servicios de datos*: la arquitectura incorpora en forma transparente la posibilidad de prestar servicios de datos a terceros interesados. De este modo, los datos generados por el RM de la EEA Anguil, podrían ser consumidos en forma directa y sin mediar solicitud o intervención humana, por Centros de Investigación, Servicios Meteorológicos u otros organismos o personas.

Estas contribuciones representan un importante avance con respecto al modelo de procesamiento presentado en [3,9], ya que ahora se incorpora la posibilidad de gestionar grandes volúmenes de datos persistentes en forma adicional a su procesamiento en línea, y contar con la posibilidad de establecer capas de servicios para el eventual consumo de los mismos.

El presente artículo se organiza en seis secciones. La sección 2 resume el funcionamiento del RM y la tipología de los datos generados. La sección 3 resume el marco conceptual C-INCAMI. La sección 4 esquematiza la nueva arquitectura de procesamiento para el EIPFDcMM junto con la nueva perspectiva de capas de servicios. La sección 5 discute los trabajos relacionados, y por último, se resumen las conclusiones y trabajos a futuro.

## 2 Radar Meteorológico

Los RM son sensores activos de teledetección que emiten pulsos de energía electromagnética hacia la atmósfera en el rango de frecuencias de las microondas. Sus mediciones se basan, por un lado, en que la radiación electromagnética a medida que se propaga en la atmósfera es dispersada por los objetos y partículas en ella existentes, y por otro lado, en la capacidad de las antenas para emitir radiación direccionada y captar la radiación incidente procedente de una dirección determinada. Estos sensores representan herramientas que permiten monitorear en forma permanente variables medio ambientales, y específicamente, la identificación, análisis, monitoreo, pronóstico y evaluación de fenómenos hidrometeorológicos, como así también de los procesos físicos que estos involucran, considerando el análisis de los riesgos que pueden ocasionar los eventos severos. Entre sus principales aplicaciones pueden mencionarse: a) Descripción del estado del tiempo, generación de pronósticos a mediano y corto plazo (Nowcasting), b) Previsión y monitoreo de contingencias ambientales (granizo, lluvias torrenciales, tormentas severas, entre otros.), c) Seguridad en la navegación y aeronavegación, d) Estudios de física de la atmósfera, e) Estudios de riesgos agroclimáticos, f) Suministro de datos básicos para la investigación científica y tecnológica, y g) Suministro de datos de entrada para modelos hidrológicos (ej: inundaciones).

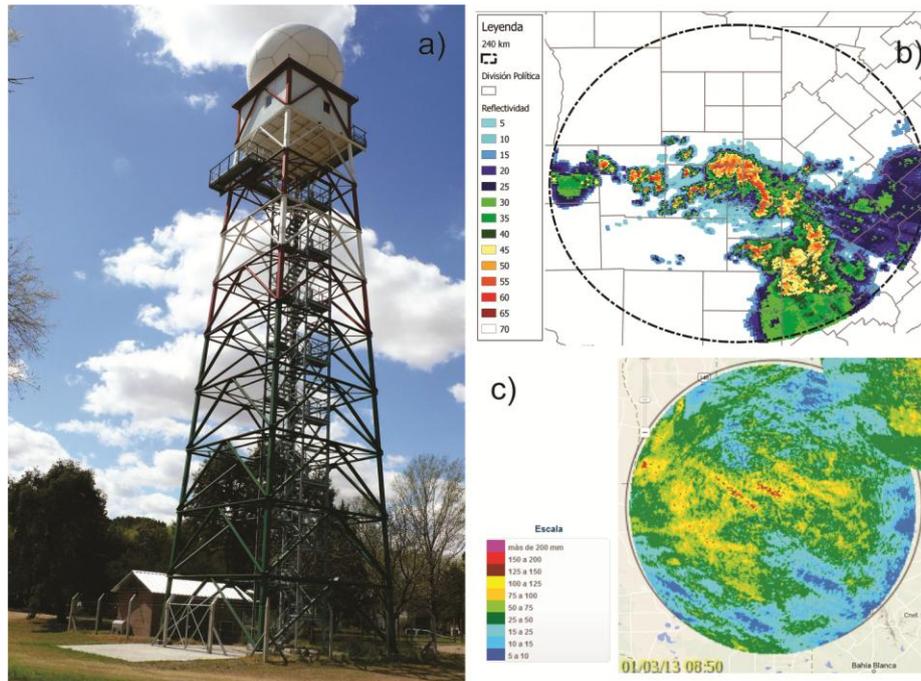
Los componentes básicos de un RM son: a) *Un transmisor*, que genera una señal de voltaje; b) *Una antena*, que direcciona esta señal concentrada en un haz y recibe la radiación retrodispersada por los blancos contenidos en un volumen de muestreo; y c) *Un receptor* que recoge la señal de voltaje inducida en la antena cuando ésta capta radiación incidente y la amplifica para facilitar su procesamiento.

La potencia recibida por un RM está relacionada con las propiedades retrodispersoras de los hidrometeoros contenidos en cada volumen de muestreo. Esta medida de la capacidad de los blancos de interceptar y devolver la energía del RM, se denomina reflectividad y es una medida de la intensidad de los ecos recibidos por la antena radar. La reflectividad depende de los parámetros físicos del objeto a medir, como su tamaño, forma, orientación y composición [10].

A seguir, se presentará en la sección 2.1 una breve reseña de la ubicación y posibilidades del RM de la EEA Anguil, mientras que en la sección 2.2 se resume el modo en que los datos son provistos desde el RM.

## 2.1 Radar Meteorológico de la EEA Anguil

El RM ubicado en Anguil (La Pampa), es de industria alemana marca Gematronik modelo Meteor 600C (Ver figura 1.a). Posee sistema doppler y es de doble polarización (DP). Opera en banda C a una frecuencia de 5,64 Ghz y longitud de onda de 5,4 cm [11].



**Fig. 1.** a) Infraestructura del RM instalado en la EEA Anguil, b) Imágen de reflectividad de la primera elevación (0,5°) del 15-01-2011 a las 23:40hs generado con Software de INTA, c) PAC (*Precipitation Accumulation*) de Febrero de 2013 generado con software Rainbow 5 de Gematronik

La antena permite un giro en el sentido horizontal (azimut) y puede elevarse en ángulo vertical hasta 45°. Este RM está configurado para completar una serie de giros a 360° que se repite para 12 ángulos de elevación, entre 0,5° de base y 15,1° de tope, en rangos de 120 km, 240 km y 480 km [12], un ejemplo para la primera elevación es expuesto en la figura 1(b). La frecuencia de este escaneo completo está programada cada 10 minutos, totalizando, bajo funcionamiento normal, 144 adquisiciones diarias.

Desde el momento de su instalación en marzo de 2009, INTA cuenta con un convenio con el Servicio Meteorológico Nacional (SMN) para su mantenimiento, por lo que es factible su empleo continuado a lo largo de diferentes proyectos de investigación propios e interdisciplinarios con otras Instituciones tales como la Facultad de Ingeniería de la UNLPam, mediante comisiones de estudio y/o cooperaciones técnicas.

## 2.2 Recolección de Datos

La información registrada por el RM es recolectada a través de escaneos volumétricos y los datos son almacenados en archivos separados llamados volúmenes. Los datos contienen las distintas variables: factor de reflectividad (Z), reflectividad diferencial (ZDR), coeficiente de correlación polarimétrica (RhoHV), desplazamiento de fase diferencial (PhiDP), desplazamiento de fase diferencial específica (KDP), velocidad radial (V) y anchura del espectro (W) [11].

Se distinguen dos tipos de datos: a) Datos “*crudos*” y b) Datos con algún nivel de procesamiento o “*productos*”. En ambos casos, la unidad de muestreo es de 1 km<sup>2</sup> y 1°, mientras que cada variable o producto se almacena en archivos separados [11].

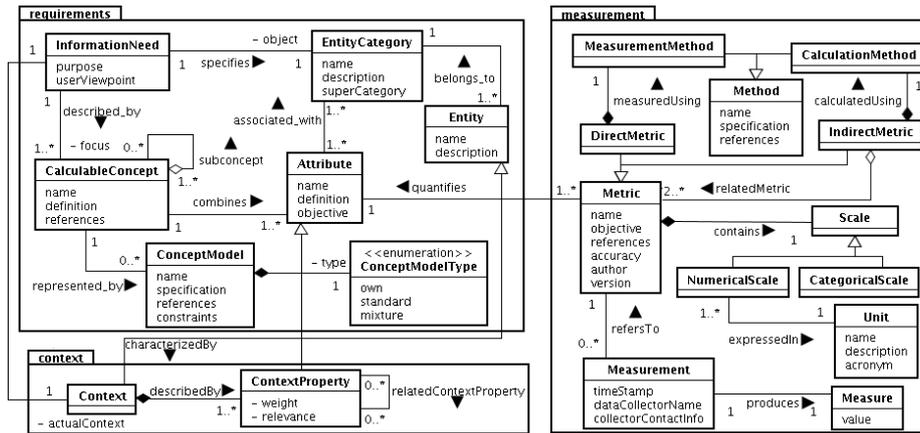
Las variables a registrar se configuran de acuerdo al rango, y posteriormente, se almacenan en volúmenes que contienen: el valor de la variable medida para cada elevación, y la posición de cada unidad muestreada con respecto del RM en coordenadas polares [12]. Los volúmenes tienen un formato propietario, con una sección XML que almacena datos del contexto de la toma de datos, y una sección de datos en formato binario con compresión que contiene el dato plano, pero que también presenta una estructura al estilo XML [12,13], e impone una restricción al uso de herramientas estándares para manipular este tipo de archivos [12]. Ante esta situación, se desarrollaron un conjunto de programas que permiten descargar los volúmenes, transformarlos a dos formatos estándares (*ASCII* y *GeoTIFF*) y procesarlos [14] (Ver figura 1.b y 1.c). De este modo, se genera un archivo *ASCII* por cada elevación con el formato: latitud, longitud y valor de la variable. Bajo funcionamiento normal, en un día completo (00:00h a 23:50h), se generan 8.640 archivos sólo para el rango de 240 kilómetros. Luego, a partir de los *ASCII*, se generan archivos raster en una matriz de 487x505 píxeles de 1 km<sup>2</sup> de resolución, en coordenadas geográficas (latitud y longitud) y *Datum* WGS84 [14].

A partir de datos “*crudos*” y mediante el software propietario Rainbow 5 [11], se pueden obtener diferentes procesamientos, por ejemplo, algunos productos hidrológicos que estiman características de la precipitación como SRI (*Surface Rain Intensity*), que genera valores de intensidad o PAC (*Precipitation Accumulation*), que calcula un acumulado de lluvia en un intervalo de tiempo predefinido. Estos productos pueden tener formato al estilo XML o de imagen raster [11]. Asimismo, INTA desarrolló un software que permite generar más productos a partir de datos crudos radar, por ejemplo aplicaciones de modelos para estimar ocurrencia de granizo y daño en cultivos [14].

Así, pueden mencionarse como beneficiarios directos (o consumidores) de los datos y productos generados, con acceso libre y abierto a: i) Sistema Nacional de Radares Meteorológicos (SiNaRaMe), ii) Servicio Meteorológico Nacional (SMN), iii) Subsecretaría de Recursos Hídricos de la Nación (SSRH), iv) Instituto Nacional del Agua (INA), v) Defensa Civil, vi) Fuerza Aérea Argentina, vii) Aviación comercial y general, viii) Dirección de Agricultura y Contingencias Climáticas (DACC, Mendoza), ix) Oficina de Riesgo Agroclimático (ORA), xi) Universidades, xii) Grupos de Investigación y Desarrollo de productos relacionados, xiii) Compañías de Seguros, xiv) Medios de comunicación, e xv) INTA.

### 3 Panorama de C-INCAMI

C-INCAMI es un marco conceptual [4,5] que define los módulos, conceptos y relaciones que intervienen en el área de M&E, para organizaciones de software.



**Fig. 2.** Principales conceptos y relaciones de los componentes Especificación de Requerimientos no Funcionales, Especificación del Contexto y de la Medición

Se basa en un enfoque en el cual la especificación de requerimientos, la medición y evaluación de entidades y la posterior interpretación de los resultados están orientadas a satisfacer una necesidad de información particular. Está integrado por los siguientes componentes principales: 1) Gestión de Proyectos de M&E; 2) Especificación de Requerimientos no Funcionales; 3) Especificación del Contexto del Proyecto; 4) Diseño y Ejecución de la Medición; y 5) Diseño y Ejecución de la Evaluación. La mayoría de los componentes están soportados por los términos ontológicos definidos en [5]. En la figura 2, se muestra un diagrama con los principales conceptos y relaciones para los componentes de requerimientos, contexto y medición.

Los flujos de medidas que se informan desde las fuentes de datos al EIPFDcMM, se estructuran incorporando a las medidas, metadatos basados en C-INCAMI tales como la métrica a la que corresponde, el grupo de seguimiento asociado, el atributo de la entidad que se mide, entre otros. Dentro del flujo, se etiquetan conjuntamente con cada medida asociada al atributo, las medidas asociadas a cada propiedad de contexto. Gracias a la formalización del proyecto de M&E en base a C-INCAMI, el hecho de procesar el flujo etiquetado, permite la estructuración del contenido de un modo consistente y alineado con el objetivo del proyecto. Esta estructuración de las mediciones dentro del EIPFDcMM, mantiene el concepto con el que se asocia cada medida; por ejemplo, si es una medida de atributo o bien de propiedad contextual. De este modo, se enriquece el análisis estadístico dado que es posible en forma directa, verificar la consistencia formal y sintáctica de cada medida contra su definición formal, en forma previa a avanzar con técnicas estadísticas más complejas.

En este caso, la entidad bajo monitoreo la constituye la situación meteorológica

con un radio de hasta 480km con centro en EEA Anguil, y dentro de los atributos que desean monitorearse pueden encontrarse, por ejemplo, la cantidad de agua presente en la atmósfera, para lo que el valor del factor de reflectividad (Z, ver sección 2.2) constituye una métrica plausible de análisis. Asociado a él, un indicador de precipitaciones puede definirse, junto con criterios de decisión y cursos de acción que favorezcan la emisión temprana de alertas.

## 4 Grandes Volúmenes de Datos Persistentes y Capas de Servicios en el EIPFDcMM

### 4.1 Arquitectura de Procesamiento desde la Perspectiva del Flujo de Datos

El EIPFDcMM es un gestor de flujos semi-estructurados de mediciones, enriquecidos con metadatos sustentados en C-INCAMI, especializado en proyectos de M&E, que incorpora comportamiento detectivo y predictivo en línea junto con la capacidad de gestionar y proveer grandes volúmenes de datos sobre demanda. Como puede apreciarse en la Figura 3, la idea que subyace al modelo en términos de procesamiento de flujos [15] es la siguiente.

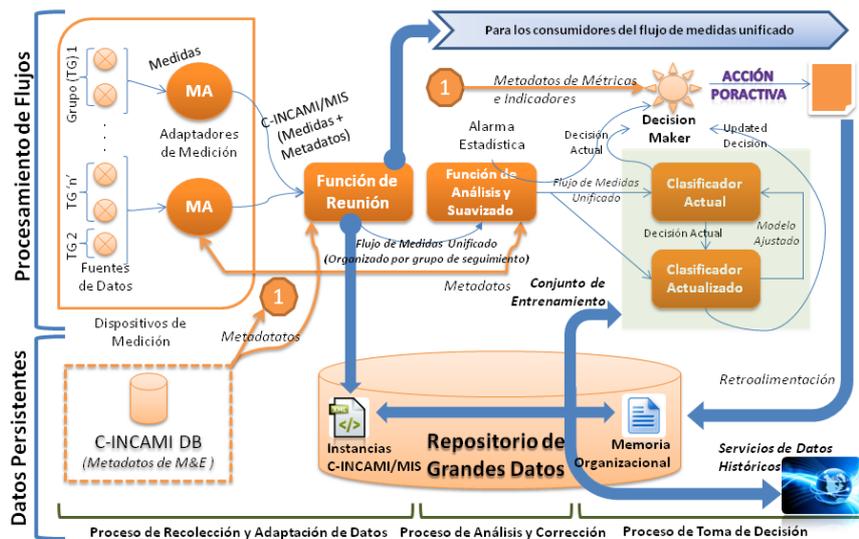


Fig. 3. Esquema Conceptual del EIPFDcMM con Big Data

Las mediciones se generan en las fuentes de datos heterogéneas (por ejemplo, el RM del INTA), las cuales abastecen a un módulo denominado adaptador de mediciones (MA en la Figura 3) generalmente embebido en los dispositivos de mediciones. MA incorpora junto a los valores medidos, los metadatos del proyecto de medición y los informa a una función de reunión central (*Gathering Function –GF*). GF incorpora

los flujos de mediciones paralelamente en: a) El repositorio de grandes datos en forma persistente, b) En un flujo C-INCAMI/MIS para los consumidores suscriptos que deseen procesar la información al momento en que se genera (por ejemplo, si fueren datos del RM del INTA, un consumidor podría ser el Servicio Meteorológico Nacional), y c) Dentro de un buffer organizado por grupos de seguimiento –modo dinámico de agrupar a las fuentes de datos definido por el director del proyecto de M&E- con el objeto de permitir análisis estadísticos consistentes a nivel de grupo de seguimiento o bien por región geográfica donde se localicen las fuentes de datos, sin que ello implique una carga adicional de procesamiento. Adicionalmente, GF incorpora técnicas de load shedding [16] que permiten gestionar la cola de servicios asociada a las mediciones, mitigando los riesgos de desborde independientemente el modo en que se agrupan.

Así, el flujo C-INCAMI/MIS es incorporado al repositorio de grandes de datos con mediciones y metadatos, y permanece disponible para atender peticiones de servicios asociados con datos de medición históricos (*Big Data Repository* y *The Historical Data Services* en la figura 3). Por otro lado, y además de que el flujo de mediciones es remitido a los consumidores suscriptos, una copia de este continúa dentro del procesador de flujo de datos y se le aplica análisis descriptivo, de correlación y componentes principales (*Analysis & Smoothing Function –ASF-* en Figura 3) guiados por sus propios metadatos, a los efectos de detectar situaciones inconsistentes con respecto a su definición formal, tendencias, correlaciones y/o identificar las componentes del sistema que más aportan en términos de variabilidad. De detectarse alguna situación en ASF, se dispara una alerta estadística al tomador de decisiones (*Decision Maker -DM*) para que evalúe si corresponde o no disparar la alerta externa (vía e-mail, SMS, etc) que informe al personal responsable de monitoreo sobre la situación (Por ejemplo, en el caso del RM del INTA, podría señalar un eventual error de dispositivo o descalibración del RM). En paralelo los nuevos flujos de mediciones son comunicados al clasificador vigente (*Current Classifier –CC-*), quien deberá clasificar las nuevas mediciones si corresponden o no a una situación de riesgo e informar dicha decisión al DM. Simultáneamente, se reconstruye en línea el CC incorporando las nuevas mediciones al conjunto de entrenamiento y produciendo con ellas un nuevo modelo (*Updated Classifier -UC*). El UC clasificará las nuevas mediciones y producirá una decisión actualizada que también será comunicada al DM. El DM determinará si las decisiones indicadas por los clasificadores (*CC* y *UC*) corresponden a una situación de riesgo y en cuyo caso con qué probabilidad de ocurrencia, actuando en consecuencia según lo definido en el umbral mínimo de probabilidad de ocurrencia definido por el director del proyecto. Finalmente, independientemente de las decisiones adoptadas, el UC se torna en CC sustituyendo al anterior, en la medida que exista una mejora en su capacidad de clasificación según el modelo de ajuste basado en curvas ROC (*Receiver Operating Characteristic*) [17].

Las decisiones a las que arriba el DM (sean a partir de una alerta estadística o bien de los clasificadores), permite retroalimentar la memoria organizacional basada en casos de la arquitectura (*Organizational Memory –OM-* en Fig. 3). De este modo, OM considera las variables bajo monitoreo de las distintas entidades bajo análisis, a través del flujo de medición C-INCAMI/MIS, incorporando las probabilidades de ocurrencia

de los distintos eventos configurados, a partir de los criterios de decisión de los indicadores definidos para el proyecto de M&E.

Adicionalmente, el hecho de contar con un repositorio de grandes volúmenes de flujos C-INCAMI/MIS en OM, retroalimentado en forma permanente a partir del procesamiento de los distintos flujos de datos, hace posible disponer de una base de entrenamiento consistente para el start-up inicial de los clasificadores.

#### 4.2 Arquitectura de Procesamiento desde la Perspectiva de Capas de Servicios

Desde la perspectiva de la capa de servicios, el sistema que dará soporte al procesamiento de datos del RM y la provisión de servicios a terceros desde la EEA Anguil, sigue un modelo arquitectónico distribuido en capas (Ver Fig. 4), constituyendo una estructura modular, flexible, segura y extensible de la aplicación. Consiste en una arquitectura genérica, que esta basada en Servicios Web y *Streaming*. Dicha arquitectura está pensada para ser usada como núcleo de aplicaciones de gestión datos del RM con distintas finalidades y fue desarrollada teniendo en cuenta tres aspectos principales:

- I. *Es necesario integrar distintos niveles de datos e información*, a saber: 1) Los datos “crudos” (Ver sección 2.2), 2) Datos validados, clasificados y corregidos por el procesador de flujos de mediciones (GF y ASF en Fig. 3), 3) Los metadatos y reglas de validación necesarias para el procesamiento de flujos (C-INCAMI DB y MA en Fig. 3), y 4) Una Memoria Organizacional con conocimiento adquirido de casos con datos del RM previos (Ver *Organizational Memory* en Fig. 3). Cualquiera de estos cuatro niveles de datos puede ser accedido a través de servicios básicos de consultas que brinda el sistema (Ver Fig. 4).
- II. *Facilitar la escalabilidad y extensibilidad*. Las tres capas lógicas se distribuyen en al menos tres capas físicas basadas en nodos con capacidad de procesamiento paralelo e integración dentro de uno o más clusters. De este modo, es posible regular, escalar y extender las prestaciones de las capas de: a) *Servicios externos*, la cual posee contacto directo con el RM, usuario, servicios y herramientas de terceros; b) *Acceso a Datos*, responsable de regular el acceso, actualización de datos y aplicación de las políticas de datos en general, constituyéndose en un cluster de servicios y c) *Persistencia*, responsable por el resguardo de los datos mediante un cluster de almacenamiento distribuido.
- III. *Incorporar Seguridad y Protección de los datos*, sea para los accesos permitidos o bien para la actualización de los datos del RM efectuados mediante la capa de servicios externos.

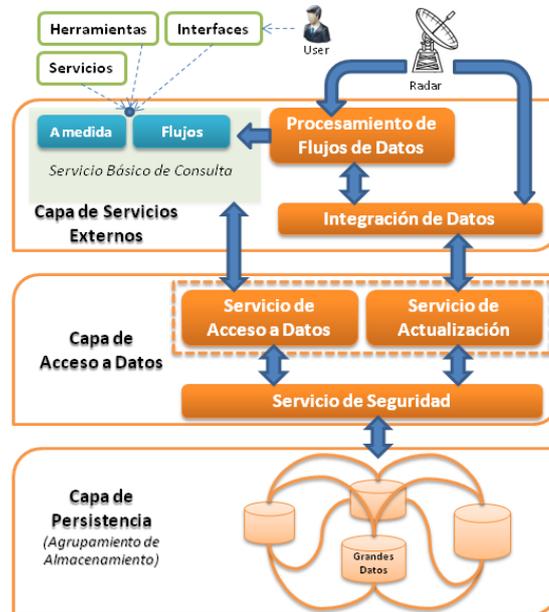


Fig. 4. Arquitectura de Procesamiento desde la Perspectiva de las Capas de Servicios

En la figura 4 se sintetiza la arquitectura procesamiento de datos del RM desde la perspectiva de las capas de servicios (Ver sección 4.1 para una vista secuenciada de procesamiento). De este modo, para el diseño de la arquitectura, se ha elegido un estilo arquitectónico de tres capas, a saber:

- **Capa de Persistencia:** almacena de manera persistente tanto los datos del RM sin procesar (tal y como son obtenidos del RM), como los datos del RM ajustados y validados mediante el procesamiento en línea.
- **Capa de Acceso a Datos:** Brinda la funcionalidad de acceso a los datos del RM en forma uniforme y controlada para todos los servicios. Esto permite aplicar políticas de seguridad que resguarden los datos del RM y garanticen su consistencia e integridad. Se compone de los siguientes servicios a) *Servicio de Acceso a Datos (Data Access Service)*, el cual permite consultas sobre los datos e información del RM almacenada en el *Big Data*, b) *Servicio de Actualización de Datos (Update Service)*, permite actualizar los datos del RM de la capa de persistencia, brindando funcionalidad al módulo de integración de datos de la capa de servicios externos, y c) *Servicios de Seguridad (Security Service)*, el cual además de verificar los usuarios (ideales o reales) que desean acceder a los datos del RM, se verifican sus permisos asociados y se aplican las políticas de seguridad en forma uniforme a todos los módulos de la capa actual, como así también a las que dependen de esta.
- **Capa de Servicios Externos:** Por un lado y a través de los servicios básicos de consulta (*Basic Service of Query*), implementa la funcionalidad de consultas ad-hoc o consumo de *streaming* de mediciones del RM para los usuarios, los agentes u

otras herramientas que los requieran. Por otro lado, mediante el procesador de flujos de mediciones y las capacidades de transformación, se permite el procesamiento en línea (con comportamiento detectivo y predictivo) sobre el propio flujo de mediciones del RM, incorporando la posibilidad de transformar los datos (guiado por sus metadatos embebidos) mediante el esquema de integración de datos (*Data Integration*) para su posterior actualización.

## 5 Trabajos relacionados y Discusión

Existen trabajos que enfocan el procesamiento de flujos de datos desde una óptica sintáctica, donde se permite el modelado del flujo de procesamiento y la consulta continua sobre el flujo, es realizada en términos de atributos y sus valores asociados mediante CQL (*Continuous Query Language*) [18,19]. Nuestra estrategia, incorpora la capacidad de introducir metadatos basados en un marco formal de M&E, que guían la organización de las medidas (datos) en el buffer, facilitando análisis consistentes y comparables desde el punto de vista estadístico, con la posibilidad de disparar alertas en forma proactiva a partir de los diferentes análisis estadísticos o bien de la decisión a la que arriben los clasificadores. Adicionalmente, cuenta con los procesos formalizados mediante *SPEM*, lo que promueve una especificación bien establecida, comunicable y extensible.

El Servicio Climático de Estados Unidos [20] posee una aplicación web en los que es posible consumir datos provenientes desde diferentes RM distribuidos a lo largo del país. Nuestra propuesta, incorpora metadatos de medición al momento de la generación de datos del RM, lo que permite un estudio de la medición en base a su semántica y no solo en términos del valor cuantitativo arrojado, con las posibilidades de automatización que ello implica. Más aún, la base C-INCAMI prevé la posibilidad de gestionar distribuciones de probabilidad para una medida dada, si es que un valor determinístico no puede ser obtenido.

La Agencia Estatal de Meteorología (AEMET) del Gobierno de España, tiene diversos mecanismos por el cual hace público los datos climáticos provenientes de los radares de su región [21]. Si bien AEMET posee una política de datos abiertos, nuestra propuesta incorpora metadatos a los datos desde su propio origen para guiar su procesamiento sin alterarlo, y adicionalmente, permite al suscriptor y-o usuario final decidir qué tipos de datos necesita consultar (sea por *streaming*, o bien, por demanda a partir del repositorio *Big Data*), siempre dentro de un marco de seguridad asociado.

Por otro lado, el Instituto Nacional de Investigación Espacial (Ministerio de Ciencia, Tecnología e Innovación) de Brasil dispone también de aplicaciones Web mediante las cuales hacen públicos los datos provenientes de sus RM, contando con la posibilidad incluso de seleccionar el RM de origen que se desee analizar en particular [22]. Nuestra propuesta prevee el aprovisionamiento por suscripción de datos generados en tiempo real desde el RM, adicionalmente guiado por metadatos sustentado en una base ontológica de M&E, que permiten determinar si los mediciones arrojadas son consistentes con la definición del objeto bajo estudio.

En Argentina y adicionalmente a INTA, existen entidades que cuentan con RM en

funcionamiento tales como el Servicio Meteorológico Nacional (SMN), la Dirección de Agricultura y Contingencias Climáticas (DACC, Mendoza), como así también proyectos tales como el Sistema Nacional de RM (SINARAME), con reciente instalación en Córdoba, de su primer RM con tecnología doppler de doble polarización de fabricación Argentina [23]. Dichas entidades ponen a disposición de los usuarios información en tiempo real a través de diferentes productos de sus RM en un formato de imagen (GIF, PNG, JPG) mediante diferentes sitios web [24,25,26]. A diferencia de los servicios actuales mencionados, nuestra propuesta incorpora la posibilidad de acceder tanto a datos “crudos” en línea, como a datos “procesados” (o productos) vigentes y-o históricos, como así también a su detalle y a los mecanismos empleados para la generación de productos.

## 6 Conclusiones y Trabajo Futuro

En el presente artículo hemos discutido la incorporación del RM de la EEA Anguil como fuente de datos heterogénea del EIPFDcMM, a los efectos de incorporar comportamiento detectivo y predictivo sobre las mediciones a medida que son generadas. Adicionalmente, se planteó la actualización de la arquitectura de procesamiento, para poder no solo gestionar servicios por suscripción en tiempo real sobre los datos que genera el RM, sino también poder satisfacer a partir de repositorios *Big Data* la petición de datos del RM ad-hoc dentro de un marco de seguridad dado. Esto permite y promueve adicionalmente, la capacidad de extender el procesamiento a otros RM de INTA (o incluso otras Instituciones), ya que sus incorporaciones serían en forma transparente como una fuente de datos heterogénea más, posibilitándose un monitoreo detectivo y predictivo en línea de la red de RM.

Por otro lado, la arquitectura de procesamiento también ha sido expuesta desde la perspectiva de la capa de servicios, lo que incorpora en forma transparente la prestación de datos a terceros interesados o beneficiarios en general. Esto permitiría que los datos fuesen consumidos sin mediar solicitud o intervención humana, por diferentes interesados.

De este modo, el hecho de poder definir un proyecto de M&E sobre una base ontológica dada e incorporarla dentro de la arquitectura de procesamiento como se ha expuesto, permite que las mediciones provenientes del RM sean monitoreadas ante eventuales desvíos en su funcionamiento, promoviendo los ajustes meteorológicos pertinentes, junto con los cursos de acción pre-definidos dentro del proyecto de M&E.

Uno de los principales aspectos en cualquier proceso de medición, reside en la comparabilidad de sus medidas a lo largo del tiempo. En tal sentido, nuestra estrategia, *permite incrementar la confiabilidad en el procesamiento con respecto al proyecto de M&E, haciendo consistente el cómputo sobre las medidas, y promoviendo así, la interoperabilidad con respecto a las diferentes fuentes de datos y-o los destinatarios que deseen emplear tal información*, gracias a: a) encontrarse sustentada en un marco formal de M&E como C-INCAMI, b) contar con una ontología subyacente de M&E, c) guiar su procesamiento de datos en base a los metadatos asociados con la entidad

bajo análisis, incluyendo también a su contexto, y finalmente, d) contar con los procesos formalizados que facilitan la comunicabilidad y extensibilidad de la estrategia. Como trabajo a futuro, se avanzará sobre la definición formal del proyecto de M&E conjuntamente con los especialistas del INTA, con el objetivo de establecer los umbrales de funcionamiento normales del RM para las variables bajo estudio, como así también pre-definir los ajustes meteorológicos y cursos de acción ante eventuales desvíos, los cuales serán incorporados posteriormente a EIPFDcMM.

**Reconocimientos.** Esta investigación está soportada por los proyectos PICTO 2011-0277 y CD 066/12 por la UNLPam y por el Proyecto PAMSL-1282101 de INTA Anguil, en el marco del convenio de cooperación técnica entre la Facultad de Ingeniería de la UNLPam y el INTA Anguil.

### Referencias

1. J. Gehrke, J. Balakrishnan, H. Namit, "Towards a Streaming SQL Standard," Proceedings of the VLDB Endowment, vol. 1, no. 2, pp. 1379-1390, August 2008.
2. M Diván, L Olsina, and S Gordillo, "Strategy for Data Stream Processing Based on Measurement Metadata: An Outpatient Monitoring Scenario," Journal of Software Engineering and Applications, vol. 4, no. 12, pp. 653-665, December 2011.
3. M Diván and L Olsina, "Process View for a Data Stream Processing Strategy based on Measurement Metadata," Electronic Journal of SADIO, vol. 13, no. 1, pp. 16-31, June 2014.
4. Olsina L. Molina H, "Towards the Support of Contextual Information to a Measurement and Evaluation Framework," in QUATIC, Lisboa, Portugal, 2007, pp. 154-163.
5. Papa F, Molina H. Olsina L, "How to Measure and Evaluate Web Applications in a Consistent Way," in Ch. 13 in Web Engineering.: Springer, 2007, pp. 385-420.
6. M. Aref, W. Bose, R. Elmagarmid, A. Helal, A. Kamel, I. Mokbel, M. Ali, "NILE-PDT: A Phenomenon Detection and Tracking Framework for Data Stream Management Systems," in VLDB, Trondheim, Norway, 2005, pp. 1295-1298.
7. S. Chandrasekaran, S. Cooper, O. Deshpande, A. Franklin, M. Hellerstein, J. Hong, W. Madden, s. Reiss, F. & Shah, M. Krishnamurthy, "TelegraphCQ: An Architectural Status Report," IEEE Data Engineering Bulletin, vol. 26, 2003.
8. D. Ahmad, Y. Balazinska, M. Cetintemel, U. Cherniack, M. Hwang, J. Lindner, W. Maskey, A. Rasin, A. Ryvkina, E. Tatbul, N. Xing, Y. & Zdonik, S. Abadi, "The Design of the Borealis Stream Processing Engine," in Conference on Innovative Data Systems Research (CIDR), Asilomar, CA, 2005, pp. 277-289.
9. M Diván, M Martín, and L Olsina, "Towards the feedback of the Data Stream Processing based on Organizational Memory (in spanish)," in Congreso Nacional de Ingeniería Informática/Sistemas de Información, Córdoba, Argentina, 2013.
10. R Rinehart, RADAR for Meteorologists, 5th ed. St. Joseph, USA, Misuri: Reinehart, 2007.
11. Gematronik, Rainbow® 5 Products & Algorithms. Neuss, Germany: Gematronik GmbH, 2005.
12. T. Hartmann, M. Tamburrino, and S Bareilles, "Análisis preliminar de datos obtenidos por la red de radares del INTA para el estudio de precipitaciones en la Región Pampeana," in 39º Jornadas Argentinas de Informáticas - 2º Congreso Argentino de Agroinformática, Buenos Aires, Septiembre 2010, p. 826.
13. Gematronik, Instruction Manual. Rainbow 5. Neuss, Germany: Gematronik GmbH, 2007.

14. Y. Bellini Saibene, M. Volpaccio, S. Banchemo, and R. Mezher, "Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteo-rológicos del INTA," in 43º Jornadas Argentinas de Informática - 6º Congreso Argentino de Agroinformática, Buenos Aires, 2014, pp. 74-86.
15. M Diván, "Enfoque Integrado de Procesamiento de Flujos de Datos centrado en Metadatos de Mediciones," UNLP, La Plata, PhD Thesis 2011.
16. M. Rundensteiner, W. Mani, M. Wei, "Utility-driven Load Shedding for XML Stream Processing," in International World Wide Web, Beijing, China, 2008, pp. 855-864.
17. C. Duin, R. Tortorella, F. Marrocco, "Maximizing the area under the ROC curve by pair-wise feature combination," ACM Pattern Recognition, pp. 1961-1974, 2008.
18. S. Widom, J. Babu, "Continuous Queries over Data Streams," ACM SIGMOD Record, pp. 109-120, 2001.
19. C. Bockermann and H. Blom, "Processing Data Streams with The RapidMiner Streams Plugin," Technical University of Dortmund, Dortmund, Germany, Report 2012.
20. National Weather Service - USA. (2015, April) National Weather Service Doppler Radar Images. [Online]. <http://radar.weather.gov/>. Último acceso abril 6 de 2015.
21. Agencia Estatal de Meteorología. Gobierno de España. (2015, Abril) El Tiempo. Radar - Composición Península y Baleares - Últimas 12 horas - Agencia Estatal de Meteorología - AEMET. Gobierno de España. [Online]. <http://www.aemet.es/es/eltiempo/observacion/radar>. Último acceso abril 4 de 2015.
22. Instituto Nacional de Investigación Espacial (Ministerio de Ciencia, Tecnología e Innovación) de Brasil. (2015, Abril) Precipitación por Radar. [Online]. <http://sigma.cptec.inpe.br/radar/>. Último acceso abril 4 de 2015.
23. G Rocca. (2015, Abril) Cazadores de Tormentas. Servicio de Información Científica de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires. [Online]. <http://neciencia.exactas.uba.ar/radares-meteorologicos-sinarame-tormentas-pronostico-paola-salio>. Último acceso abril 20 de 2015.
24. Servicio Meteorológico Nacional. (2015, Abril) Imágenes de Radar. [Online]. <http://www.smn.gov.ar/?mod=radar&id=4>. Último acceso abril 20 de 2015.
25. Dirección de Agricultura y Contingencias Climáticas. Mendoza, Argentina. (2015, Abril) Imágenes de Radar. [Online]. <http://www.contingencias.mendoza.gov.ar/web1/Radar/radar.html>. Último acceso abril 20 de 2015.
26. Instituto Nacional de Tecnología Agropecuaria. (2015, Abril) Red de Radares Meteorológicos. [Online]. <http://radar.inta.gob.ar>. Último acceso abril 20 de 2015
27. A. Rodríguez and H. Ciappesoni, "Sistema Nacional de Radares Meteorológicos. SINARAME.," in Simposio de Radarización y Sistemas de Alerta Hidrometeorológicas del Cono Sur, Buenos Aires, Argentina, 2012.