

CONICET



Memoria del Taller

**Manejo y Almacenamiento de Datos en
Observatorios, Proyectos de Investigación
Orientados y Redes Temáticas**

**Taller Manejo y Almacenamiento de Datos en
Observatorios / Proyectos de Investigación Orientados / Redes Temáticas**

06 de marzo de 2015

Auditorio CONICET – Av. Rivadavia 1917 – Planta Baja.

El Taller Manejo y Almacenamiento de Datos en Observatorios / Proyectos de Investigación Orientados / Redes Temáticas, tuvo como objetivo principal reunir los representantes de los proyectos de Redes Temáticas, los Observatorios y los Proyectos de Investigación Orientados en pos de discutir sobre las formas de mejorar la calidad del manejo y almacenamiento de datos que realiza cada uno de ellos.

En el comienzo de la jornada, el Dr. Jorge Tezón y la Dra. Patricia Maccagno dieron una breve introducción donde se expusieron las problemáticas y necesidades principales en torno al tema de los datos en general y sobre la información ambiental en particular. Dichas problemáticas se centran en la dispersión de la información, la discontinuidad temporal, la no-correlación geográfica, el enfoque unidisciplinar, la heterogeneidad de metadatos y la falta de acceso o visibilidad. Lo cual trae como consecuencia la duplicidad de los esfuerzos institucionales, la visión restringida del problema a analizar y sus causas generando, a su vez investigaciones de menor impacto científico y la falta de contribución a la solución de las problemáticas a atender.

En este sentido es deseable que los Observatorios organicen la información primaria disponible asegurando su acceso, obtenida del monitoreo constante en diferentes puntos de relevamiento. Asimismo, se establecen los estándares de calidad de datos y las bases acordados por las instituciones integrantes del mismo de manera que exista y funcione la interoperabilidad de las bases a través de formatos compatibles y metadatos consensuados que incluyan el georreferenciamiento de la información procesada.

Con este tipo de organización de la información y el análisis integral aportado por diferentes disciplinas, recolectando información en diversos lugares de interés y manteniendo una misma estructura de producción de datos, se pueden obtener productos de mayor alcance en términos del impacto así como herramientas más abarcativas para los tomadores de decisiones. Entre ellos, se puede mencionar a los mapas de riesgo, escenarios de impacto climático,

estimación de costos y daños frente a amenazas, planificación de ordenamiento territorial, análisis costo-beneficio frente a determinadas situaciones de impacto así como estrategias de inversión.

Al finalizar la presentación de bienvenida, durante la sesión de la mañana tuvieron lugar las presentaciones correspondientes a las Redes Temáticas, Proyectos de Investigación Orientados y los Observatorios. Las mismas se refirieron a las características de la información que abordan y las necesidades identificadas sobre el manejo, acceso y almacenamiento de datos producidos en el marco de la iniciativa así como las dificultades para integrar y sistematizar datos históricos relevados pero aún no incorporados al corpus de información que cada proyecto maneja.

A continuación se sintetizan las presentaciones en su orden de ocurrencia:

1) Observatorio “Autoridad Cuenca Matanza-Riachuelo” (ACUMAR)

La Lic. Rosana Echarri describió las particularidades de este proyecto y las dificultades en términos de acceso e integración de la información relevada hasta el momento así como las dificultades para organizar nuevas tomas de datos.

Brindó un ejemplo que ilustra las dificultades afrontadas por esta iniciativa, que depende de la entrega de información realizada por numerosos distritos a lo largo de la cuenca, la cual presenta heterogeneidad en las variables, los soportes, los tiempos de relevamiento, la continuidad del dato, entre otros. Puntualmente, se explicó sobre la necesidad de caracterizar los barros de la cuenca y poder definir sus componentes, pero la extensión y el estudio intercalado de estos hizo de la información un cuerpo de datos sectorizados y heterogéneos. Asimismo, se planteó la dificultad al momento de captar la información de fuentes externas a las entidades que participan en ACUMAR.

2) Red Argentina para el Estudio de la Atmósfera Superior (RAPEAS)

La Prof. Marta Mosert y la Dra. Adriana Gulisano plantearon uno de los objetivos fundamentales de la Red que es optimizar el aprovechamiento de los recursos instrumentales instalados y proyectados para estudiar diferentes

fenómenos en la alta atmósfera. Para ello, esta red de investigadores hizo un relevamiento de los sistemas observacionales disponibles en la Argentina para poner su información a disposición de los integrantes del proyecto así como al público en general a través del sitio Web de la Red.

La problemática principal es la heterogeneidad encontrada en el relevamiento de datos históricos y actuales, que son tomados por diferentes aparatos (Ionosondas, Información Satelital, Detector de radiación Espacial por efecto Cherenkov, Magnetómetros), que miden variables distintas y se materializan en diferentes soportes de acuerdo a su época de relevamiento y/o de las disponibilidades tecnológicas de los aparatos. Muchos de ellos se encuentran en film 7 mm y 16 mm, cintas de papel, discos magnéticos de 3,5” y 5 y ¼ “, data cartridges y planillas en papel, no se cuenta con los soportes para la lectura de los mismos o los elementos para la digitalización. La variedad de estos archivos hace muy complejo poder llegar a una uniformidad de datos y hacer inter-operables las bases que los alojen, para poder plantear un acceso uniforme.

3) Observatorio Nacional de Degradación de Tierras y Desertificación (ONDTYD)

Darío Soria y Rubén Soria describieron las características de la información que maneja el Observatorio, en su tarea de monitorear y evaluar la degradación de las tierras a escala nacional/ local. En primera instancia se recolecta información primaria vinculada con aspectos biofísicos y socioeconómicos en 17 sitios piloto a escala local en los cuales se monitorea la información periódicamente. Para manejar este amplio cuerpo de información es necesario acordar los estándares de calidad de los datos y de las bases que los alojen así como establecer los formatos de inter – operabilidad de dichas bases ajustando el sistema de metadatos e impartiendo en todos los registros su correspondencia con la georreferenciación de la toma del dato.

La necesidad identificada se vincula con la incorporación masiva de información proveniente de los sitios piloto del Observatorio a la base ya existente del repositorio de datos geoespacial que tiene su acceso público a través del sitio Web. Asimismo, otro desafío es el establecimiento de criterios para manejo e integración de los distintos tipos de soporte (informes, gráficos,

tablas, indicadores). Finalmente, se plantea la discusión sobre el acceso diferencial a la información que el Observatorio releva y produce vinculada con las características poblacionales de los sitios piloto debido a su carácter sensible.

4) RED Proyecto Argentino de Monitoreo y Prospección de Ambientes Acuáticos (PAMPA2)

El Dr. Horacio Zagarese, presentó las características de toma, recolección y sistematización de los datos que se corresponden con las 13 lagunas bajo monitoreo en el marco del proyecto. Todas ellas tienen muestreos de datos de baja frecuencia (mensuales o estacionales) que permiten monitorear información como los nutrientes, clorofila, fitoplancton, zooplancton, picoplancton y peces. Cinco de ellas tienen en funcionamiento los sensores automatizados (boyas) que permite hacer una colecta de alta frecuencia de información de distinta naturaleza como la temperatura del aire y del agua, profundidad, turbidez, conductividad, humedad, velocidad y dirección del viento, etc.

La principal necesidad identificada es la posibilidad de integrar las bases de datos ya existentes en formatos inter-operables, que permitan acceder a toda la información de manera de poder realizar un análisis integral de las variables medidas y establecer las series de larga duración para la caracterización de la región.

Finalizadas las presentaciones de las iniciativas, se dio paso a la exposición de las plataformas e iniciativas existentes que ya han brindado soluciones a algunas de las dificultades identificadas por los Proyectos.

5) Plataforma Interactiva de Investigación para las Ciencias Sociales (PLIICS)

La Mg. Laura Leff presentó la propuesta para la preservación, recopilación y uso integrado de datos primarios en las Ciencias Sociales cuyo objetivo es facilitar el acceso abierto a datos primarios en Cs. Sociales para fortalecer la investigación científica y ampliar el intercambio interdisciplinario.

Esta plataforma se conforma de datos primarios tanto sean de origen cualitativo (series de datos: censos, encuestas, relevamientos y registros administrativos) como cuantitativo (documentos, entrevistas, mapas, y objetos multimedia - audios, videos, fotografías). Cuenta, además, con criterios de integración, estandarización y almacenamientos de datos que responden a estándares internacionales, utiliza programas de acceso abierto y se basa en dos tipos de repositorios (DSpace y Dataverse).

Es un sistema de organización de la información que está dirigido principalmente a los miembros del sistema científico (no sólo investigadores) y es compartido con otros actores de la sociedad (educación, gobierno, periodismo, empresas, etc). Aún está en su fase de prueba piloto, implementándose en varios casos de prueba.

6) Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCYT)

La MSc Silvia Nakano, inició su presentación con un diagnóstico de base sobre las publicaciones y datos primarios de investigación científica, en el cual se detallaban las características del sistema local que impulsaron a diseñar una nueva manera de acceder a la producción científica. Luego, hizo un recorrido por los antecedentes del Ministerio en materia del acceso abierto, liderado por la gestión de los Sistemas Nacionales de Datos (Biológicos, del Mar y Genómicos) así como la creación del sistema Nacional de Repositorios Digitales. Asimismo, se hizo mención a la sanción y el proceso de reglamentación de la Ley 26.899 la cual establece la creación de los Repositorios Digitales de Acceso Abierto, indicando la creación de repositorios institucionales interoperables con la producción científico-tecnológica, la obligatoriedad del depósito de obras y datos por parte de los investigadores en dichos repositorios y la asignación del financiamiento futuro de investigaciones vinculado a la publicación en acceso abierto.

En este sentido, se presentaron los 31 Repositorios adheridos (el CONICET está entre ellos) y la importancia de la organización y centralización de la información dispersa sobre los proyectos de investigación y la comunidad científica para la toma de decisiones tanto en lo referido a los lineamientos científico-técnico del Ministerio así como de la cosecha de datos de las

producciones científicas que son realizadas con fondos públicos en todas las disciplinas y a escala nacional.

7) Gerencia de Sistemas de CONICET – Sistema Nacional de Repositorio de Datos

El Ing. Diego Asensio y la Mg. Paula Nahirñak presentaron algunas de las características del cambio que tuvo el SIGEVA (Sistema Integral de Gestión y Evaluación) desde su creación hasta la fecha, incorporando distintos módulos para la ampliación de la información requerida de cada persona que tiene su perfil en el sistema, así como el incremento en el detalle de la producción que se auto-archiva en él. Estas necesidades fueron surgiendo a partir de su implementación como base de la información persona no sólo en CONICET, sino que fue requerida por otros organismos y entidades académicas y educativas. El avance en el carácter de interoperable con otras bases de datos (CVar, Categorización Docente, numerosas universidades nacionales, entre otras) ha sido fundamental para consolidar los criterios de inclusión de información y módulos específicos. Actualmente, el SIGEVA es requerido por 49 instituciones en las cuales será implementado.

Asimismo, se presentó las bases del CONICET Digital, el cual conforma el Repositorio Institucional indicado por la Ley mencionada a partir del 2013. El CONICET resuelve conformar un Comité Asesor que organiza los lineamientos a seguir para cumplimentar las nuevas necesidades del acceso abierto a los datos producidos por los investigadores y becarios con financiamiento público. En el marco de esta legislación, el CONICET esta a punto de lanzar su prueba de Repositorio con el material de las publicaciones que fueron auto-archivadas en el SIGEVA por el personal científico. En línea con esta iniciativa, se planea generar espacios temáticos en los cuales se pueda acceder a través de los temas específicos a las producciones.

Finalmente, en relación con los Datos Primarios aún no se ha avanzado en los desarrollos orientados a organizar los datos primarios producto de la recolección de información del personal científico del CONICET. Se plantea un desafío a corto plazo para establecer los criterios necesarios que permitan desarrollar bases de datos que contengan esta valiosa información para poder ser re-utilizada por otros especialistas.

9) Centro Argentino de Información Científica y Tecnológica (CAICYT)

La Prof. Mela Bosch, Directora del Centro, el Lic. Fernando Ariel López, Jefe Sectores Comunicación Institucional y Formación y Mgr. Diego Ferreyra Responsable de Tecnología Documental y la Ing. Mirna Prieto, realizaron una presentación en la cual se indicó el marco general de la problemática de datos abiertos en el mundo y las propuestas CAICYT al respecto. En tal sentido, se adelantaron los aspectos del Proyecto Institucional Metodologías en Herramientas Digitales en la Investigación. Además presentaron las bases de un Marco de Verificación de Calidad de Metadatos de datos científicos. Dicho marco permite valorar los niveles de estructuración formal para cualquier producción de datos y se divide en tres módulos: a) Denominación e Identificación de los Datos; b) Modelos de Acceso y Explotación; c) Derechos y Licencias. El cumplimiento de los mismos permite crear condiciones de comparabilidad, análisis, uso y explotación de los datos, con independencia de los marcos metodológicos, marcos teóricos y actores.

Conclusiones

1. **Próxima reunión: noviembre 2015.** Se acordó para esa fecha un nuevo Taller a fin de que presentes los avances alcanzados.
2. **Conformación de Grupo de trabajo sobre datos primarios:** Que usará una plataforma de intercambio durante el estudio proporcionada por CAICYT y como repositorio de resultados un espacio que será proporcionada por la Dirección de Informática de CONICET. Objetivo: oportunidades de intercambio de experiencias. Estudio y asesorías sobre aspectos puntuales. Para esta actividad se enviará un mail a los participantes, a fin de conformar el grupo fecha límite 30/05. este grupo de trabajo realizará:
 - a. Revisión de documentación y enlaces: experiencia internacional Plan de Gestión de Datos, Ciclo de gestión y sobre modalidades de exposición de datos. Se solicitará a los participantes que aporten bibliografía para ponerlos a disposición de todos a través de un espacio común
 - b. Talleres de intercambio (con obtención de productos) El primero a mediados de año

- c. Estudio y comparación de las experiencias internacionales en Plan de gestión de datos
 - d. Estudio de las guías y recomendaciones sobre distintos aspectos de los marcos y plataformas para Planes de gestión de datos
 - e. Estudio de los Protocolos de manejo por tipología y temática de los datos
 - f. Estudio de los lineamientos para construir bases de exposición de datos de proyectos y redes
 - g. Acordar un ciclo de etapas de producción de datos primarios etapas desde colecta a disponibilidad de información
3. Aplicación del proceso en sitios Pilotos. Se seleccionará algunos de los Observatorios/REDES a fin de poner en funcionamiento todos los acuerdos logrados

A partir de la presentación de cada una de las iniciativas y de los avances en las distintas áreas especialistas en manejo de datos, se estableció un marco de discusión sobre las necesidades de tipo particular y otras de tipo institucional.

A continuación se presentan los acuerdos alcanzados para poner en marcha tanto al interior de cada una de las Redes Temáticas, Proyectos de Investigación Orientados y Observatorios, así como a nivel institucional de apoyo:

ANEXO 1: Participantes del Taller

ACUMAR	Arturi Diego	darturi@acumar.gov.ar
	Echarri Rosana	recharri@acumar.gov.ar
	Villegas Romina	rvillegas@acumar.gov.ar
	Iglesias Cecilia	
RAPEAS	Gulisano Adriana	adrianagulisano@gmail.com
	Mosert Marta	mmosert@icate-conicet.gob.ar
ONDTYD	Perinetti Mariano	mperinetti@mendoza-conciet.gov.ar
	Soria Darío	ndsoria@mendoza-conicet.gov.ar
	Soria Rubén	rsoria@mendoza-conicet.gov.ar
	Therburg Almut	atherburg@mendoza-conicet.gov.ar
PAMPA2	Zagarese Horacio	zagarese@intech.gov.ar
FYPF	Bellomo Santiago	santiago.bellomo@ypf.com
	Orio Lucila	maria.l.orio@ypf.com.ar
CAICYT	Bosch Mela	mbosch@caicyt.gov.ar
	Diego Ferreyra	dferreyra@caicyt.gov.ar
	Prieto Mirna	mprieto@caicyt.gov.ar
	López Fernando	flopez@caicyt.gov.ar
MINCYT	Nakano Silvia	snakano@mincyt.gob.ar
	Fischberg Sabrina	sfischberg@mincyt.gob.ar
	Gorjon Gabriela	ggorjon@mincyt.gob.ar
	Apollaro Alberto	santiago.bellomo@ypf.com
CONICET	Asensio Diego	dasensio@conicet.gov.ar
	Nahirñal Paula	paulan@conicet.gov.ar
	Carlino Lorena	lcarlino@conicet.gov.ar
	Leff Laura	laleff@conicet.gov.ar
	Tezón Jorge	jtezon@conicet.gov.ar
	Maccagno Patricia	pmaccagno@conicet.gov.ar
	Castro Mora	mora.castro@conicet.gov.ar
	Zenocrati Alejandra	anzenocrati@conicet.gov.ar
	Gorkin Micaela	mgorkin@conicet.gov.ar
	Fernández Cristina	mcfernan@conicet.gov.ar

ANEXO II: Agenda

Taller Manejo y Almacenamiento de Datos en Observatorios / Proyectos de Investigación Orientados / Redes Temáticas

06 de marzo de 2015

Auditorio CONICET – Av. Rivadavia 1917 – Planta Baja.

Agenda

10:00 a 10:15	Presentación del Taller: objetivos y metodología de trabajo
10:15 a 10:30	Observatorio Autoridad de Cuenca Matanza Riachuelo (ACUMAR)
10:30 a 10:45	Red Argentina para el Estudio de la Atmósfera Superior (RAPEAS)
10:45 a 11:00	Observatorio Nacional de Degradación de Tierras y Desertificación
11:00 a 11:15	Café
11:15 a 11:30	Proyecto Argentino de Monitoreo y Prospección de Ambientes Acuáticos (PAMPA2)
11:30 a 11:45	Plataforma Interactiva de Investigación en Ciencias Sociales (PLIICS)
11:45 a 12:30	Repositorio de datos MINCyT
12:30 a 13:30	Perspectivas en la integración de la información: CONICET Digital, Producción Edita y Datos Primarios.
13:30 a 14:30	Almuerzo
14:30 a 15:00	Estado del Arte de gestión y construcción de datos. CAICyT

-Mesa de discusión sobre aspectos técnicos de la gestión de datos-