

## **A. Título del Proyecto Propuesto**

**Informatización de colecciones del Museo Argentino de Ciencias Naturales, con énfasis en registros de Parques Nacionales de Argentina**

## **B. Información para contactar al (los) investigador(es) principal(es)**

- Nombre: Martín J. Ramírez
- Domicilio: Museo Argentino de Ciencias Naturales – CONICET, Av. Angel Gallardo 470, C1405DJR Buenos Aires
- País: Argentina
- Teléfono: +54 11 4982-8370 int. 169
- Fax: +54 11 4982-4494
- Correo electrónico: ramirez@macn.gov.ar
- Página Web: <http://aracnologia.macn.gov.ar/index.php>; <http://www.macn.secyt.gov.ar/>

## **C. Información de Contacto de la Institución Responsable**

- Nombre de la Institución: **Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)**
- Domicilio: Av. Rivadavia 1917, C1033AAJ Buenos Aires, Argentina
- Persona contacto en la Institución: Presidente
- Nombre: Eduardo Charreau
- Teléfono: +54 11 4953-7230/39
- Fax: +54 11 4953-7230/39
- Correo electrónico: presidencia@conicet.gov.ar
- Página Web: <http://www.conicet.gov.ar/>

## **D. Si procede, una Lista de las otras Instituciones participantes**

No corresponde

## **E. Resumen del Proyecto**

### **Resumen**

El Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" (MACN) alberga 21 Colecciones Nacionales biológicas (total estimado, 1,5 millones de especímenes), con crecimiento constante a un ritmo de 12000 especímenes por año. Las colecciones tienen cobertura regional, mayormente de Argentina y países limítrofes, con muy buenos muestreos en áreas protegidas. Se propone la conversión de los registros en papel (etiquetas, libros de entrada) a formato digital de acuerdo a estándares de interoperabilidad, sometidos a ciclos de validación en cuanto a formatos y contenidos (geográficos, taxonómicos, etc.), integrados a las bases del Museo, y compartidos a través de IABIN. Se enfatizará en conjuntos de registros provenientes de Parques Nacionales de Argentina y sus áreas adyacentes. El trabajo será realizado por pasantes universitarios, un coordinador de procesos, y consulta permanente con los curadores de las colecciones. Se incluirá un curso de capacitación para los participantes. Los resultados esperados son (a) producción de unos 42000 nuevos registros digitales, a un costo de US\$0.24 por registro, en un plazo de 7 meses; (b) validación de los contenidos geográficos y taxonómicos producidos; (c) integración de estos productos en el esquema general de datos de especímenes del MACN, incluyendo difusión gratuita y provisiones para actualización continua.

### **Summary**

The Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" (MACN) holds 21 biological National Collections (total estimates, 1.5 million specimens), with a constant grow rate of ca. 12000 specimens each year. The collections are of regional coverage, mainly Argentina and neighbor countries. This project aims to convert records in paper (labels, books) to a digital format according to standards of interoperability, submitted to cycles of validation of format and content (eg., geography, taxonomy). The records will be integrated to the museum database and shared through IABIN. This project will emphasize in records from Argentinean National Parks and adjacent areas. The work will employ undergrad trainees, a process coordinator, in close consultation with the curators of the collections. There will be a training course for participants. The expected results are (a) the production of 42000 new digital records, at US\$0.24 per record, in 7 months; (b) validation of the produced geographic and taxonomic content; (c) integration of these products in the general data schema of MACN, including free access and provisions for continued update.

## **F. Descripción del proyecto**

El Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" (MACN) alberga 21 Colecciones Nacionales biológicas, con un total estimado de 1,5 millones de especímenes, con crecimiento constante, a un ritmo estimado de unos 12000 especímenes por año (estimación 2002 – 2006). Cada colección cubre un grupo taxonómico o cronológico específico (aves, insectos, plantas vasculares, vertebrados fósiles, etc.) y está a cargo de un curador, usualmente investigador del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) especializado en el grupo. Dependiendo de la colección, puede haber además curadores asociados y técnicos curadores. Las colecciones son usualmente de cobertura regional, especialmente de Argentina y países limítrofes. El MACN mantiene una intensa actividad de préstamos y consultas de sus colecciones, que agregan permanentemente datos valiosos (determinación actualizada; el espécimen es designado voucher de estudios específicos, etc.).

### **Objetivos generales**

La conversión de los registros en papel a registros digitales permitirá:

1. Acceder, mantener, compartir y diseminar los datos de especímenes de manera eficiente. Facilitar las instancias de captura de datos adicionales (agregado de coordenadas geográficas; actualización de determinaciones).
2. Proveer una fuente de registros relevantes de alta calidad (datos respaldados por especímenes testigo, colectados de manera sistemática y profesional, identificados por expertos, actualizados en ciclos breves). El SIB<sup>4</sup> incorporará los datos a su sistema.

### **Tareas específicas**

1. Digitalizar un conjunto importante de datos de especímenes de las colecciones del MACN de acuerdo a estándares de interoperabilidad (Darwin Core 2, ver respuesta iii y Costo por Unidad), con énfasis en especímenes de Parques Nacionales y áreas adyacentes.
2. Transformar 13000 registros digitalizados según estándares de interoperabilidad
3. Integrar la digitalización de datos con ciclos de validación de contenidos (ver respuesta vii abajo). Integrar los datos digitalizados a las bases de datos del MACN para garantizar su mantenimiento en el largo plazo (ver respuesta ii).

### **Metodología**

1. Los registros serán digitalizados por pasantes universitarios de la Universidad de Buenos Aires, en una rutina de 4 horas diarias. Se conformará el siguiente equipo de trabajo: un investigador principal, un responsable informático, un coordinador de procesos, 4 pasantes universitarios, y los curadores de las colecciones actuando como consultores y facilitadores.
2. Los pasantes serán seleccionados de acuerdo a antecedentes y desempeño en un curso preparatorio de tres días. Candidatos naturales son los estudiantes que ya se encuentran realizando trabajos en las colecciones como voluntarios o tesisistas. Utilizarán computadoras que están siendo actualizadas (ver fondos de contrapartida)
3. El responsable informático del MACN preparará la aplicación de adquisición de datos, basada en MS Access, incluyendo validaciones básicas de los datos ingresados (ver respuesta vii). Esta herramienta trabajará sobre las mismas tablas y campos que el sistema de datos de las colecciones del Museo. Las pautas y protocolos de trabajo serán redactadas por el investigador principal y el responsable informático.
4. Las tareas de los pasantes serán organizadas y supervisadas por un coordinador de procesos, contratado para este proyecto. El coordinador de procesos elaborará (con herramientas provistas por el responsable informático) las tablas de nombres de taxones que necesitan ser validados por curadores especialistas, y las tablas de localidades a ser validadas y georeferenciadas (ver respuesta vii). También realizará las rutinas masivas de validación y consistencia de datos, y procesará los resultados de validación por especialistas y georeferenciadores.
5. El coordinador de procesos y el responsable informático reportarán al investigador principal. Estos tres participantes mantendrán reuniones semanales para seguimiento e implementación de ajustes. Las pautas, protocolos y la herramienta de adquisición de datos

se modificarán de acuerdo a los resultados de evaluaciones, informes programados (ver cronograma), y a los resultados de los controles de calidad (ver respuesta vii).

6. Las colecciones y conjuntos de datos serán seleccionados por el investigador principal en consulta con los curadores de las colecciones y un responsable del SIB de Parques Nacionales<sup>4</sup>, teniendo en cuenta entre otros aspectos (a) áreas geográficas de interés, (b) grupos taxonómicos de interés, y (c) vacancias de cobertura taxonómica.
7. El coordinador de procesos, en colaboración con el investigador principal transformarán unos 13000 registros de arácnidos previamente digitalizados a los estándares de interoperabilidad.

### **Relevancia y valor institucional**

Los curadores e investigadores serán beneficiados por el acceso y mantenimiento de sus propios datos de manera eficiente e interoperable, que terminarán reflejados en una mejor y mayor producción científica. El MACN se beneficiará por una mayor exposición de sus datos y por un mayor flujo de datos agregados a sus especímenes y una mejor asignación de recursos para la gestión de las colecciones y sus datos.

Al movilizar grandes conjuntos de datos e interactuar con un usuario inmediato interesado (Administración de Parques Nacionales, con fines de manejo, conservación y uso sustentable), el MACN mejorará su capacidad de impulsar el proceso de digitalización de datos en el país, de acuerdo con su vocación.

Se han detectado al menos dos grupos de investigación interesados en los datos que resulten de este proyecto (ver i.3 abajo). Se proveerán datos de calidad de grupos escasamente representados en las bases de Parques Nacionales (por ejemplo, de arácnidos, identificados como vacancia). El esfuerzo del MACN en digitalizar los datos de sus especímenes es complementario y sinérgico con procesos similares que están realizando otros institutos del país.

### **Costo por unidad**

El número de registros se medirá contando los registros producidos que cumplan (1) el mínimo de campos obligatorios; (2) haber pasado los ciclos de validación (ver respuesta vii). Se digitalizarán un estimado de 42000 registros, de acuerdo a la siguiente proyección: 4 pasantes a 20 horas semanales, durante 6 meses = 1920 horas hombre. Aplicando un promedio de 15 especímenes digitalizados por hora (según estimaciones previas), resultará en 28800 registros. A esto se agregan unos 13000 registros digitalizados que deben ser adaptados a estándares, totalizando unos 418000 registros. Esto produce un costo de US\$0.24 por cada nuevo registro.

Nuestra base de datos de especímenes tiene implementados 77 campos (más algunos en agenda), cubriendo la prácticamente la totalidad de Darwin Core 2, más campos adicionales de sus extensiones.

### **Colecciones seleccionadas y preseleccionadas**

La colección aracnológica (200.000 especímenes) cubre una vacancia de datos en el SIB<sup>4</sup> y será seleccionada para este trabajo. Se adaptarán los 13000 registros previamente informatizados, y se seleccionarán taxones y conjuntos de datos de especial interés para el SIB. Por ejemplo, de los 1862 especímenes estudiados en una monografía reciente<sup>3</sup>, una gran parte provienen de parques nacionales de Patagonia y áreas adyacentes. La taxonomía de arañas está completamente catalogada<sup>1</sup>. Otros candidatos de acuerdo a interés taxonómico son las colecciones de Herpetología y Mastozoología.

### **Cronograma de trabajo e informes**

**Mes 1.** Comienzo del proyecto. Compra de computadora para coordinador de procesos. Curso de formación (tres días) y selección de personal. Redacción de pautas y protocolos.

**Mes 2.** Comienzan los pasantes universitarios (data-entry). Digitalización de registros, validación de los datos. Ajustes iniciales a los protocolos.

**Mes 3.** Digitalización de registros, validación de los datos. Redacción de informes de avance: técnico y financiero. Ajustes a protocolos si fuera necesario.

**Mes 4-6.** Digitalización de registros, validación.

**Mes 7.** Digitalización de registros, validación. Informes finales: técnico y financiero.

### **Literatura relevante y notas de pie de página**

- <sup>1</sup> Platnick, N. I. 2007. The World Spider Catalog, Version 8.0. The American Museum of Natural History. <http://research.amnh.org/entomology/spiders/catalog/INTRO1.html>
- <sup>2</sup> Ramírez, M. J. 2003. A cladistic generic revision of the spider subfamily Amaurobioidea (Araneae, Anyphaenidae). Bulletin of the American Museum of Natural History 277: 1–262.
- <sup>3</sup> Ramírez, M. J., J. A. Coddington, W. P. Maddison, P. E. Midford, L. Prendini, J. Miller, C. E. Griswold, G. Hormiga, P. Sierwald, N. Scharff, S. P. Benjamin, W. C. Wheeler. 2007. Linking of digital images to phylogenetic data matrices using a morphological ontology. Systematic Biology, 56, 283–294.
- <sup>4</sup> Sistema de Información de Biodiversidad. Administración de Parques Nacionales. <http://www.sib.gov.ar/>

### **Respuestas a las preguntas**

#### **i. ¿Quiénes son las comunidades de usuarios en su país que estarían interesadas en los productos de este proyecto?**

1. Comunidad científica en general, a través de los portales de datos (IABIN, GBIF).
2. Administración de Parques Nacionales, para sus proyecto de sistema de información de biodiversidad SIB<sup>4</sup>.
3. Instituciones nacionales que hacen uso de datos de biodiversidad (INTA, estaciones agronómicas de control, secretarías de medio ambiente nacional y provinciales, Dirección Nacional de Fauna, etc.)
4. Proyectos científicos específicos que movilizan gran volumen de datos; dos ejemplos nacionales son los estudios de áreas de endemismo (Instituto Miguel Lillo) y distribuciones predictivas de organismos para recomendar nuevas áreas protegidas (IADIZA).
5. Especialistas en sistemática y taxonomía que consultan los especímenes del MACN.

#### **ii. ¿Cómo serán actualizados y mantenidos los datos en el futuro? Si se hacen cambios a la información, ¿cómo se harán estos cambios en los registros de acceso público?**

1. El personal del MACN a cargo de las colecciones (curadores y técnicos) ingresarán las actualizaciones en datos de especímenes que retornan de préstamos, o que son determinados por científicos visitantes. Los investigadores del MACN actualizarán los datos de especímenes personalmente desde sus puestos de trabajo, a medida que se generan actualizaciones como resultado de sus propias investigaciones.
2. Los datos de especímenes compartidos a través de IABIN y GBIF son un producto derivado de los datos propios de las colecciones del MACN, y por lo tanto reflejarán las actualizaciones periódicas internas del MACN en ciclos breves de actualización.

#### **iii. ¿Cómo cumplirá con los estándares de SSTN?**

Se coleccionarán datos de especímenes cubiertos por Darwin Core y extensiones.

#### **iv. ¿Qué experticia en informática y especímenes tiene su institución?**

1. El MACN cuenta con un responsable de informática (Ing. Daniel Rodríguez), quien diseñó sistemas de gestión de datos de especímenes para el Instituto Darwinion, y lo está haciendo ahora para el MACN (sistema Aurora).
2. El Investigador Principal (IP) es Administrador del Nodo Argentina de GBIF, ha participado y organizado de actividades de formación en recursos informáticos para colecciones (georeferenciación, curación y conservación, etc.), y ha publicado contribuciones utilizando herramientas informáticas sofisticadas<sup>3</sup>. El IP y los curadores del MACN han publicado numerosos trabajos revisando especímenes de colecciones de muchos museos en el mundo (ver por ejemplo <http://aracnologia.macn.gov.ar/publicaciones.php>).

#### **v. ¿Cómo interactúa su institución con otras instituciones en su país que manejan datos sobre especímenes y/o temas relacionados?**

1. El MACN es pionero y proactivo en la búsqueda de mecanismos para intercambiar datos de especímenes por medios digitales. Alberga al nodo nacional de GBIF y coordina la Red Nacional de Colecciones Biológicas de Argentina. Junto con CONICET está preparando proyectos para la creación de una oficina nacional de datos de biodiversidad. Ha organizado varios cursos y talleres sobre conservación y manejo de datos de colecciones.
2. El MACN tiene excelentes relaciones con usuarios directos en el país (Administración de Parques Nacionales, IADIZA, Instituto Lillo, por nombrar unos pocos).

**vi. ¿Qué rol ha jugado Usted y/o su institución en las actividades de IABIN? ¿Qué relación tiene con el Punto Focal de IABIN en su país? (Carta de aval del Punto Focal de IABIN enviada aparte).**

El MACN ha participado de numerosas actividades organizadas por IABIN y tiene reuniones periódicas de discusión y planificación general con el Punto Focal de IABIN y otros actores en la temática de biodiversidad. Ha preparado propuestas de proyectos en colaboración con el punto focal, que han sido elevados al Banco Mundial.

**vii. ¿Qué medidas serán tomadas para asegurar la calidad de los datos?**

Se implementarán ciclos de validación de datos, para lo cual se contratará un Coordinador de Procesos con experiencia en datos de especímenes, taxonomía y geografía.

Los pasos de control y validación de datos ocurrirán en diversas instancias:

1. La aplicación utilizada para ingresar los datos minimizará errores de tipeo y formato mediante las técnicas usuales (selección de elementos en menú, máscaras).
2. Todas las localidades geográficas serán consistidas en un proceso paralelo y enviadas a georeferenciación (ver fondos de contrapartida). Los registros problemáticos serán aprobados o marcados para revisión.
3. Todos los elementos taxonómicos serán consistidos (en lo posible automáticamente) con servidores de nombres (p.ej., Species 2000) o con catálogos sistemáticos, de acuerdo a cada colección. Los curadores de cada colección revisarán listas consistidas de nombres de taxones, según su especialidad. Los nombres se aprobarán o marcarán para revisión.
4. El MACN cuenta con investigadores activos en taxonomía de muchos grupos. La selección de las colecciones sobre las que se trabajará garantizará la disponibilidad de especialistas que asistan en el proceso de digitalización y verificación.
5. Se seleccionarán registros al azar y se examinarán en detalle, como seguimiento del desempeño de los data-entry del proceso de verificación.

**viii. ¿Cómo se llevará a cabo el monitoreo y la evaluación? ¿Qué indicadores de cumplimiento y de impacto medirán?**

Esquema de indicadores de cumplimiento

<b>Actividad</b>	<b>Mes 3</b>	<b>Mes 5</b>	<b>Mes 7</b>
Ingreso de nuevos registros validados	7200 registros (25%)	15840 registros (55%)	28800 registros (100%)
Transformación según estándares de 13000 registros de arácnidos	3250 registros (25%)	7150 registros (55%)	13000 registros (100%)
Validación de registros	<i>Backlog &lt; 3 semanas</i>	<i>Backlog &lt; 2 semanas</i>	<i>Sin backlog</i>

Esquema de indicadores de impacto

<b>Objetivos</b>	<b>Indicadores de impacto</b>	<b>Indicadores de resultado</b>
Digitalizar datos de las Colecciones Nacionales del MACN		Datos disponibles públicamente en IABIN durante 2008
Intercambiar datos con usuarios extrainstitucionales, identificados y potenciales	Relevancia e importancia de los nuevos conjuntos de datos en los resultados de usuarios	Uso de los datos generados por estos actores
Proveer datos de Parques Nacionales y áreas circundantes	Cobertura de vacancias en el SIB	Porcentaje y cantidad significativos de especímenes de dichas áreas

### G. Presupuesto del proyecto

Todos los costos están presentados en dólares americanos.

Rubro	IABIN	Fondos de Contrapartida	Total
1. Operativos: Pasantes universitarios, estipendio más seguro	5,981.07		5,981.07
2. Operativos: Coordinador de procesos, estipendio más seguro	2,536.28		2,536.28
3. Operativos: Insumos, material de oficina	500.00		500.00
4. Mercancía: Computadora para coordinador	982.65		982.65
5. Servidor de datos		5,425.87	5,425.87
6. Actualización de computadoras para pasantes		2,965.30	2,965.30
7. Salario de investigadores: investigador principal y curadores, en tareas de conducción, consulta, y curso (totalizando una dedicación al 100% entre todos).		8,832.81	8,832.81
8. Salario de informático (dedicación 20%)		1,715.77	1,715.77
9. Pasante universitario, georeferenciador (dedicación 50%)		662.46	662.46
<b>Total</b>	<b>10,000.00</b>	<b>19,602.21</b>	<b>29,602.21</b>

Los fondos de contrapartida surgen de CONICET (ítems 7, 8) y del "Proyecto de apoyo a la informatización de las colecciones del Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" financiado por la Fundación Williams (ítems 5, 6, 9).