# La participación Argentina en el relevamiento LSST y su relación con la Astroestadística

M.J. de L. Domínguez Romero[1,2,3]

[1] *Instituto de Astronomía Teórica y Experimental, CONICET–UNC, Argentina*
[2] *Observatorio Astronómico de Córdoba, UNC, Argentina*
[3] *Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina*

*Contacto* / mariano.dominguez@unc.edu.ar

**Resumen** / El Legacy Survey of Space and Time (LSST) es un relevamiento del cielo austral que llevará a cabo el Vera Rubin Observatory desde Cerro Pachón (Chile). Para ello se utilizará el telescopio Simonyi (8.4m) con una cámara y óptica especial de gran campo (aproximadamente 10 grados cuadrados). El relevamiento implica 10 años de observaciones. En esta charla discutiremos cómo más de 60 científicos argentinos ya se encuentran trabajando en ocho grandes colaboraciones científicas, cómo dicha participación coordinada por el Instituto de Astronomía Teórica y Experimental (IATE) es posible, y cómo los estudiantes interesados pueden sumarse a este proyecto. Revisaremos la línea de tiempo del proyecto y su importancia en áreas que van desde el sistema solar a la cosmología y las importantes sinergias que presenta con los proyectos observacionales desarrollados actualmente en Argentina. Exploraremos los desarrollos informáticos y herramientas de automatización que permiten reducir más de 20 TB por noche, cómo y dónde se analizarán dichos datos y catálogos y la importancia/consecuencias de tener un centro de análisis independiente en nuestro país. Revisaremos los métodos de aprendizaje automático que generarán y clasificarán más de 10 millones de alertas de variabilidad por noche y el impacto de la Astronomía en el dominio temporal en todas las áreas. Veremos qué herramientas Astro Estadísticas hemos desarrollado para realizar inferencia Bayesiana directa en cosmología y estudios del volumen local, así como también sus implicaciones en los estudios del sector oscuro.

**Abstract** / The Legacy Survey of Space and Time (LSST) is a survey of the southern hemisphere sky that will be carried out by the Vera Rubin Observatory from Cerro Pachon (Chile). It will be done using the Simonyi telescope (8.4m) with a special wide-field camera and optics (approximately 10 square degrees). This study implies 10 years of observations. In this talk, we will discuss how more than 60 Argentinian scientists are already working in eigth significant scientific collaborations, how this collective participation coordinated by the Instituto de Astronomía Teórica y Experimental (IATE) is possible, and how interested students can participate in this project. Furthermore, we will revise the project timeline, its importance in areas ranging from the solar system to cosmology, and the critical synergies it presents with the observational projects currently designed in Argentina. Additionally, we will explore the computer developments and automatization tools that allow us to reduce more than 20 TB per night, how and where such data and catalogs will be analyzed, and the importance/consequences of having an independent analysis center in our country. Moreover, we will revise the automatic learning methods that will generate and classify more than 10 million variability alerts per night and the impact of Astronomy in the temporal domain in all areas. We will show what Astro Statistics tools we have developed to achieve direct Bayesian inference in cosmology and local volume studies, as well as their implications in the studies of the dark sector.

*Keywords* / instrumentation: detectors — instrumentation: telescopes — astronomical databases: catalogs — virtual observatory tools

## 1. El estado de la Astronomía Observacional.

Durante las últimas décadas se ha desarrollado el paradigma que algunos llaman e-Science; esto es la unificación de los paradigmas empíricos, teóricos y de simulaciones utilizando las herramientas de ciencia de datos y el aprendizaje automático sobre los resultados experimentales, las predicciones teóricas y los resultados de las simulaciones numéricas. En particular la Astronomía ha liderado el desarrollo y aplicación de este paradigma en nuestro dominio de la ciencia.

La exploración en las diferentes dimensiones accesibles de los datos astronómicos ha progresado notablemente con el desarrollo de multiples relevamientos masivos desarrollados durante las últimas dos décadas (Two degree Field 2dF Galactic Ring Survey GRS, Sloan Digital Sky Survey SDSS, Subaru Special Program HSC-SSP, Dark Energy Survey DES, etc.). En términos de profundidad el nuevo telescopio espacial James Webb se encuentra explorando la formación y evolución de las primeras estrellas y galaxias, que luego son estudiadas en detalle por los mayores telescopios actuales con base terrestre, a los cuales se sumarán la nueva generación de telescopios de 20-30 mts (Extremely Large Telescope ELT, Thirty Meter telescope TMT, Giant Magellan Telescope GMT) en desarrollo.

Informe invitado

En general para observar objetos débiles son necesarias grandes superficies colectoras.

En relación a los relevamientos, para muchos análisis estadísticos es relevante el campo de visión (FoV) y en este sentido son cruciales las dimensiones del campo focal y el detector. Dos intrumentos recientes, destacados en este aspecto son la cámara de energía oscura (DECAM) en el telescopio Blanco y la Hiper Suprime Camera (HSC) en Subaru con los cuales se han realizado importantes relevamientos de gran campo (DES y HSC-SSP respectivamente).

La dimensión o dominio temporal tiene una gran tradición en Astronomía sobre todo en estudios del sistema solar. Durante las últimas décadas se produjo un importante cambio en términos de industrialización de la misma. En ese sentido la reconversión del Observatorio de Monte Palomar en un sistema completo de detección y clasificación de variabilidad temporal ha tenido un rol dominante con la realización del Palomar Transient Factory (PTF), intermediate Transient Factory (iPTF), y culminando con el Zwicky Transient Factory (ZTF). La exploración de la variabilidad en grandes áreas del cielo y en magnitudes cada vez más débiles ha permitido identificar nuevos tipos de variabilidad estelar, encontrar las contrapartes visuales de las fuentes de ondas gravitacionales detectadas por la colaboración Laser Interferometer Gravitational-Wave Observatory y (ALIGO/VIRGO). Además se ha incrementado en órdenes de magnitud la cantidad de cuerpos menores del sistema solar y se han desarrollado redes globales de seguimiento de objetos variables con telescopios cada vez más grandes y mejor equipados como (Catalina Real-Time Transient Survey CRTS, y Las Cumbres Observatory LCOGN).

Finalmente, la dimensión del bajo brillo superficial continiúa produciendo sorpresas en términos de completar la población de galaxias mediante la introduccion de novedosos sistemas como Dragonfly y nuevos métodos de anális de imágenes en grandes telescopios como Subaru. La exploración de todas estas dimensiones en el cielo austral durante una década es el objetivo del Observatorio Vera C. Rubin, actualmente en construcción.

## 2. El relevamiento LSST del Observatorio Vera C. Rubin.

El observatorio lleva el nombre de la famosa astrónoma americana Vera C. Rubin, pionera de los estudios sobre la materia oscura mediante el análisis de las curvas de rotación de galaxias espirales. El mismo llevará a cabo un censo completo multicolor del cielo austral durante diez años, conocido como el Legacy Survey of Space and Time (LSST). A continuación se describen las principales características del mismo.

### 2.1. Sitio y Telescopio

Las instalaciones del observatorio Vera C. Rubin se encuentran en la cumbre del cerro Pachón en la parte central-norte de Chile. Este sitio se encuentra tierra adentro, a aproximadamente 100 km por carretera de la ciudad de La Serena, donde están ubicadas las instalaciones de soporte. Un video ilustrativo del mismo (video telescopio) se encuentra disponible. El telescopio de relevamiento Simonyi utiliza un diseño especial de tres espejos con el fin de crear un campo de visión excepcionalmente amplio. Esto permite realizar un relevamiento de todo el cielo en solo tres días. Se tomarán imágenes del cielo continuamente con una cadencia automática, y a lo largo del relevamiento se visitará cada ubicación del cielo más de 800 veces; mientras que regiones seleccionadas serán visitadas aún más.

### 2.2. Cámara y Filtros

La cámara del Observatorio Vera C. Rubin esta diseñada para requerir un manteniento mínimo y tomar ventaja de la alta calidad de las imágenes sobre un campo tan extenso. La misma contiene más de tres mil millones de píxeles, convirtiéndose así en la mayor cámara digital del mundo. Un video del ensamblado de la cámara y sus primeras imágenes puede apreciarse en el siguiente enlace: https://youtu.be/Tb7nfJrNnKQ . El carrousel de filtros permite cambiarlos en minutos entre el conjunto de filtros u, g, r, i, z, y.

### 2.3. La Misión del Observatorio Vera C. Rubin

LSST producirá un conjunto completo de más de 500 Pb de imágenes y productos derivados que permitirán abordar varias de las preguntas más importantes de la Astronomía moderna, entre ellas podemos mencionar:

- Entender las componentes del sector oscuro.
- Realizar un inventario del sistema solar.
- Explorar el cielo en el dominio temporal.
- Realizar un mapeo completo de nuestra galaxia.

En síntesis, un mapeo de todo el cielo del sur, con visitas cada pocas noches sostenido durante diez años, permitirá abordar investigaciones en un amplio abanico de temas.
Los productos de datos que producirá la línea compleja de anális se encuentran detallados en el siguiente documento https://pipelines.lsst.io/. El volumen de datos adquiridos en una noche es de 20 Tb y el del conjunto total al final del relevamiento será de 15 Pb con datos de más de 20 mil millones de galaxias, 17 mil millones de estrellas resueltas, mas de 10 millones de supernovas y 6 millones de objetos del sistema solar. El sistema reduce y produce del orden de 10 millones de alertas públicas de variabilidad con una latencia de 1 minuto. Durante el desarrollo del relevamiento se realizarán 11 publicaciones de datos anuales (data releases DR).

### 2.4. El Cronograma de Operaciones

La construcción del observatorio se completó recientemente, al igual que la cámara. La misma será colocada y probada durante 2023. Los datos preliminares estarán disponibles para los científicos y el comienzo del relevamiento será en enero del 2024. Cabe destacar que un gran producto del observatorio son las aplicaciones de reducción y análisis, dado el gran volumen de datos involucrados.

### 2.5. Las Colaboraciones Científicas de Rubin

El observatorio se encarga de la toma, reducción de los productos de datos descriptos anteriormente, pero el análisis e investigación de los mismos se produce dentro de las ocho grandes colaboraciones científicas temáticas listadas abajo. Existen grupos mixtos donde confluyen colaboraciones con objetivos comunes.

- Galaxies
- Stars, Milky Way and Local Volume
- Solar System
- Dark Energy
- Active Galaxy Nuclei
- Transient and Variable stars
- Strong Lensing
- Informatics and Statistics

En los sitios de cada colaboración se puede encontrar información sobre membresía, gobernanza, objetivos, grupos de trabajo, recursos, becas etc. Los llamados planes de acción (roadmaps) de las colaboraciones son documentos muy útiles para orientarse y encontrar coordinadores y encargados. Estos servirán de guía hacia los grupos de trabajo y actividades comunes, seminarios, tutoriales, escuelas y reuniones, repositorios de códigos comunes y la colaboración en la publicación de artículos. A continuación dejo los links a los roadmaps de dos de las colaboraciones: DESC y TVS.

## 3. La participación Argentina.

La participación Argentina se estableció entre el IATE y la Corporación LSST y esta consiste en la adquisición de derechos sobre datos mediante contribuciones en especies (in-kind) que comienzan a realizarse en 2023 luego de un proceso de evaluación por las agencias National Optical-Infrared Astronomy Research Laboratory (NOIRLAB), National Science Foundation (NSF) y otras quienes mantienen el funcionamiento del observatorio. Las contribuciones en especies incluyen contribuciones para el desarrollo de aplicaciones, tiempo de instrumentos y capacidades computacionales, en las cúales se incluyen contribuciones de otras instituciones del país. Esto permite la participación de investigadores y estudiantes argentinos dentro de las colaboraciones científicas y el uso de los datos, las simulaciones y los productos desarrollados para el análisis de los mismos durante el relevamiento LSST.
La participación se abrió a investigadores en Argentina y actualmente se cuenta con derecho de datos para 60 investigadores de todo el país, quienes ya participan de las ocho colaboraciones científicas del Observatorio Vera C Rubin. Esto permite aprovechar todo el potencial científico y de descubrimiento del proyecto. Existe la posibilidad de aumentar este número supeditado al soporte de las diferentes contribuciones propuestas.

### 3.1. El rol de los Agentes inteligentes para clasificación (brokers)

Uno de los resultados del observatorio más interesante son los resultados del análisis de las diferencias de imágenes, esto consiste en un flujo público de alertas de fuentes variables, transitorias y de movimiento. Estas alertas serán distribuidas a un conjunto de agentes inteligentes (brokers) comunitarios. Estos son complejos sistemas de aplicaciones que ingieren, procesan, clasifican y distribuyen globalmente las alertas a toda la comunidad científica. Algunas funcionalidades típicas consisten en cross-matching con datos de archivo, la identificación de objetos que requieren de seguimientos detallados (follow-ups) y clasificaciones según los datos fotométricos de Rubin, basados en la aplicación de técnicas de aprendizaje automático en las curvas de luz. A continuación se detallan los agentes inteligentes que se utilizan:

- Automatic Learning for the Rapid Classification of Events (ALeRCE
- Alert Management, Photometry, and Evaluation of Light curves (AMPEL)
- ANTARES
- Fink
- Lasair
- Pitt-Google

Se han realizado avances notables en las técnicas de clasificación automática durante la última década en una serie de desafíos motivados por LSST, como Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) 2018 y otros en desarrollo.

### 3.2. The Astronomical Event Observatory Network (AEON) y otras redes de seguimiento

Es previsible un gran impacto en todas las áreas de la Astronomía con el estudio de las alertas de variabilidad provistas por los los agentes inteligentes de clasificación en base a los datos de LSST, tal como ocurre actualmente en el hemisferio norte con los datos de ZTF y sus predecesores. Los eventos ya clasificados podrán ser monitoreados por redes de observatorios para seguimientos (follow-ups) donde los astrónomos podrán requerir observaciones utilizando interfaces estandarizadas que permitirán realizar estudios más detallados.

Una de ellas es AEON que proveerá acceso rápido, flexible y programable a instalaciones de primera clase para realizar seguimientos eficientes de transitorios astronómicos. En esta red confluyen el Observatorio de Las Cumbres (LCO) con NOIRLab y los telescopios SOAR,

el Observatorio Gemini y el telescopio Victor M. Blanco para apuntalar esta era de astronomía temporal con Rubin. Cabe destacar el rol la herramienta Target and Observation Manager (TOM toolkit) y el Observatory Control System desarrollados por LCO. Estas herramientas resultan indespensables para que un dado instrumento/observatorio pueda participar de estos seguimientos globales. Resulta interesante notar que prácticamente todas las facilidades astronómicas en Argentina pueden ser potencialmente utilizadas para realizar seguimientos y estudios de las alertas que producirá LSST durante la próxima década. Este posible uso sinérgico destaca la importancia de participar en LSST por parte de nuestros investigadores.

### 3.3. La protección del cielo por la Unión Astronómica Internacional (UAI)

La repentina industrialización y extensivo uso de las órbitas terrestres bajas (low Earth orbit, LEO) con la construcción de numerosas constelaciones de satélites artificiales, para proveer servicios de internet satelital y otros, complica ya el estudio del cielo. El número de satélites artificiales crece de forma exponencial y podría superar los 400000 hacia el final de la década. Esto afecta de forma notable las actividades de relevamiento y toda otra observación astronómica. Quiero destacar el reciente surgimiento de un centro de la UAI: el Centre for the Protection of the Dark and Quiet Sky from Satellite Constellation Interference (CPS) dedicado a la preservación del recurso natural que es el cielo, como resultado de la serie de talleres internacionales sobre el impacto de las comunicaciones satelitales SATCOM desarrollados durante la pandemia COVID-19. Quiero destacar que en este centro confluyen las partes interesadas, incluyendo las industrias aeroespaciales y que ya se han logrado avances, fruto de la coordinación de las actividades entre los astrónomos y la industria. Sin embargo, aún quedan muchos desafíos que resolver en relación al tema; me gustaría invitarlos a considerar nuestra participación, sea de forma institucional y/o individual. Es muy factible que la solución se oriente hacia la coordinación de las observaciones con los pasajes de los satélites en LEO.

## 4. El rol del aprendizaje automático en la Astronomía moderna.

El aprendizaje automático está cambiando todo aspecto de nuestra sociedad, basta ver las aplicaciones de los grandes modelos de lenguaje en generación y traducción de textos o de las arquitecturas multimodales en problemas de robótica y navegación autónoma. Asimismo, encontramos un gran crecimiento de aplicaciones en nuestra investigación astronómica diariamente; siendo esto un resultado natural de que nuestra ciencia es una de las que más datos produce (a la par con la Bioinformática).

### 4.1. El rol de los Centros de Análisis de Datos Independientes (IDACs) en la Astroestadística con LSST

En volúmenes de datos como los que involucra LSST, la adopción de métodos automáticos para las actividades de reducción y análisis resulta indispensable. Un punto importante a destacar es el desarrollo del cómputo que involucran estas actividades en el lugar donde residen los datos. Para el análisis de LSST se dispondrán de dos grandes centros de cómputo dedicados en Chile y Estados Unidos, que serán apoyados por centros de menor porte en la mayoría de los paises que participan de la colaboración. Estos centros resultan necesarios para descentralizar las actividades y proveerán acceso a otros conjuntos de datos y/o datos de valor añadido, producidos en ellos como catálogos especializados o ciertos cálculos especícos como identificadores de estructuras, etc. Muchos países realizan contribuciones en especies de este tipo (por ejemplo Brasil, Francia, Reino Unido, Dinamarca y muchos otros), compartiendo recursos que de todos modos son necesarios para realizar los estudios sobre los datos. En Argentina también hemos propuesto contribuir con un IDAC de menor porte, pero resta conseguir la financiación para el mismo. Una posibilidad es especializarlo para aplicaciones de inteligencia artificial, lo cual tiene el valor añadido de formar especialistas con una formación muy demandada por la industria.

### 4.2. Simulaciones de imágenes de LSST con la simulación cosmologica Second Data Challenge (DC2)

El desarollo de los algoritmos de análisis se ha realizado sobre imágenes detalladas simuladas del sistema óptico completo del telescopio y cámara, incluyendo efectos atmosféricos y de dispersión mediante el uso del paquete ImSim desarrollado por la Dark Energy Science Collaboration (DESC). Las imágenes producidas incluyen los efectos sistemáticos de la atmósfera, óptica y la física del sensor en la PSF. Las bibliotecas de software de LSST incluyen toda la información necesaria para simular la exposición como coordenadas en el plano focal, características de los filtros, brillo de cielo, etc., produciendo imágenes de un realismo notable.

En preparación para los análisis con LSST, DESC ha creado un relevamiento simulado sobre un área de 300 grados cuadrados basado en una simulación cosmológica conocida como DC2, ver https://data.lsstdesc.org/. El relevamiento simulado incluye observaciones en las seis bandas con una cadencia realista y las imágenes simuladas (6 Tb) fueron procesadas con la línea de aplicaciones de ciencia de LSST, produciendo catálogos realistas (200 Gb) ligados de los objetos reales y observados que permiten realizar estudios y análisis estadísticos de los mismos. Estos catálogos han sido validados en multiples test (Kovacs et al. (2022)) y constituyen la base de numerosos estudios en curso.

### 4.3. El impacto de la IA en el análisis de datos y las simulaciones

El análisis de los datos simulados descriptos y en breve de los datos reales serán realizados en los centros de cómputo e IDACS mediante una plataforma ya desarrollada, Rubin Science Platform, donde el usuario dispone de toda las aplicaciones. Los usuarios se suman paulatinamente y son capacitados en el uso de las mismas. Es notable el amplio uso de diferentes algoritmos de inteligencia artificial para problemas que van desde la detección de objetos y clasificación (Hausen & Robertson (2020) y Li et al. (2021)), hasta la inferencia en modelos paramétricos en modelos cosmológicos o de formación de galaxias, estructuras etc., (Li et al. (2021)). Se han logrado avances importantes en inferencia Bayesiana en los ultimos años mediante la aplicación de las técnicas desarrolladas para programar los algoritmos de aprendizaje profundo. Estos desarrollos impactan de igual manera en el paradigma de simulaciones mencionado al principio, por ejemplo para lograr mejor resolución (Schaurecker et al. (2021)), la incorporación de la física bariónica a simulaciones de N cuerpos (Bernardini et al. (2022)) o el uso de redes informadas de la física (Karniadakis et al. (2021)) para simular diferentes sistemas.

## 5. Conclusiones

La participación de astrónomos argentinos coordinada por el IATE en LSST es una realidad en la actualidad. Ya hay 12 investigadores principales y 48 investigadores jóvenes distribuidos por todo el país que cuentan con derechos a los datos y participan de las ocho colaboraciones científicas del observatorio Rubin. Existe la posibilidad de hacer crecer este número, esto depende fuertemente del funcionamiento de la contrapartida de nuestras facilidades observacionales y computacionales, ofrecidas en calidad de contribuciónes voluntarias. El apoyo que posean las mismas por las instituciones de Ciencia y Técnica del país resulta entonces fundamental.

Cabe destacar que el sistema del Observatorio Vera C. Rubin es revolucionario y con un gran potencial de descubrimiento en diferentes áreas de la astronomía (que abarcan desde el sistema solar hasta la cosmología, pasando por la astronomía estelar y la estructura de las galaxias, incluida la nuestra) que deberímos pensar en términos de sinergia y no de competencia con las facilidades observacionales argentinas. Esto nos permitirá realizar el seguimiento de una selecta muestra de las más de 10 millones de alertas de variabilidad por noche que producirá LSST. Quiero cerrar este artículo invitando a los estudiantes y jóvenes investigadores interesados a contactarse con nosotros por ofertas de investigación en todas las áreas mencionadas. El entorno de investigación inclusivo del Observatorio Vera Rubin es ideal para el desarrollo de carreras en la intersección entre la Ciencia de Datos y la Astronomía conocida como Astroestadística.

*Agradecimientos:* Se agradecen las importantes contribuciones y cuidadosa revisión de Federico DG y la cordial invitación y apoyo por parte del comité organizador de la reunión anual 2022 de la AAA.

## Referencias

Bernardini M., et al., 2022, MNRAS, 509, 1323
Hausen R., Robertson B.E., 2020, ApJS, 248, 20
Karniadakis G.E., et al., 2021, Nature Reviews Physics, 3, 422
Kovacs E., et al., 2022, The Open Journal of Astrophysics, 5, 1
Li X.J., et al., 2021, ApJS, 252, 16
Schaurecker D., et al., 2021, arXiv e-prints, arXiv:2111.06393