

# PISCIS: Plataforma de Búsqueda Interactiva y Ciencia Ciudadana

PISCIS: Platform for Interactive Search and Citizen Science

**GERMÁN ALFARO** 

Instituto de Astronomía Física y Experimental (IATE), Argentina

**INGRID VANESSA DAZA PERILLA** 

Instituto de Astronomía Física y Experimental (IATE), Argentina

**JOSÉ BENAVIDES BLANCO** 

Physics & Astronomy Department, University of California, Riverside, United States;  
Instituto de Astronomía Física y Experimental (IATE), Argentina

**MARCELO LARES** 

Observatorio Astronómico de Córdoba (OAC), Argentina; Universidad Nacional de Córdoba (UNC), Argentina

**VICTORIA SANTUCHO**

Instituto de Astronomía Física y Experimental (IATE), Argentina

**JUAN CABRAL** 

Comisión Nacional de Actividades Espaciales (CONAE), Instituto de Astronomía Física y Experimental (IATE), Argentina; Observatorio Astronómico de Córdoba (OAC), Argentina; Facultad de Matemática, Astronomía, Física y Computación (FaMaFyC), Universidad Nacional de Córdoba (UNC), Argentina

**ANA LAURA O'MILL** 

Instituto de Astronomía Física y Experimental (IATE) y Observatorio Astronómico de Córdoba (OAC), Argentina

**FACUNDO RODRÍGUEZ** 

Instituto de Astronomía Física y Experimental (IATE), Argentina

**MAURICIO KORAJ**

Programador y director de ingeniería

**Para correspondencia:** Ingrid Vanessa Daza Perilla [vanessa.daza@unc.edu.ar](mailto:vanessa.daza@unc.edu.ar)

## RESUMEN

Muchos temas de la Astronomía moderna se caracterizan por la identificación de características en imágenes. Si bien esta es una tarea fácil para un ojo entrenado, es difícil obtener la misma calidad al realizarse mediante modelos o métodos numéricos. Esto ha conducido a la cooperación entre equipos de investigación y la ciudadanía tiene una larga historia en diferentes disciplinas científicas, debido al gran interés que despierta en el público, la presencia de asociaciones de personas aficionadas y la necesidad constante de procesamiento de grandes volúmenes de datos. En este trabajo el objetivo es presentar la experiencia y el desarrollo de una plataforma web de ciencia ciudadana (PISCIS, Platform for Interactive Search and CIizen Science), plataforma que tiene como objetivo la generación de catálogos de valor agregado a partir de datos que comprenden un conjunto de imágenes haciendo uso del interés en la Astronomía del ciudadano, facilitando la recolección y el análisis de datos así como también el conocimiento al público de la metodología y temas de investigación en el área de la Astronomía. Para llevar a cabo se creó una aplicación de fácil uso para el encuestador y el público, esta permite crear encuestas concisas y acompañarlas de imágenes, junto con espacios para agregar información y ejemplos de las encuestas a realizar. Actualmente, la plataforma ya está disponible. Además, se está empleando para clasificar el primer conjunto de datos, éstos consisten en pares de galaxias, el objetivo es catalogarlos según el tipo de interacción y, para esto, se brindan ejemplos de pares que tienen una alta, media o baja interacción. La colaboración de la ciudadanía está siendo importante, tanto para evaluar la plataforma como para inspeccionar visualmente a cientos de pares de galaxias. PISCIS permite vincular a las personas con proyectos de investigación, lo que posibilita a un investigador recolectar una gran cantidad de datos. Para poder realizar las clasificaciones, es necesario un entrenamiento, por lo que, a su vez, quien participa también aprende sobre ciencia y cómo se trabaja en ella.

**Palabras clave:** clasificación; web; astronomía; ciencia ciudadana

## ABSTRACT

Many topics in modern astronomy are characterised by the identification of features in images. While this is an easy task for a trained eye, it is difficult to obtain the same quality when performed by models or numerical methods. This has led to cooperation between research teams and citizenship has a long history in different scientific disciplines, due to the great interest of the public, the presence of amateur associations and the constant need for processing large volumes of data. In this work the objective is to present the experience and development of a citizen science web platform (PISCIS, Platform for Interactive Search and CIizen Science), a platform that aims to generate value-added catalogues from data comprising a set of images making use of the citizen's interest in Astronomy, facilitating the collection and analysis of data as well as making the public aware of the methodology and research topics in the area of Astronomy. A user-friendly application was created for the pollster and the public to create concise surveys and accompany them with images, along with spaces to add information and examples of the surveys to be carried out. The platform is now available. In addition, it is being used to classify the first set of data, consisting of pairs of galaxies, the objective is to catalogue them according to the type of interaction and, for this, examples of pairs with high, medium or low interaction are provided. The collaboration of the public has been important, both in evaluating the platform and in visually inspecting hundreds of galaxy pairs.

PISCIS allows people to be linked to research projects, making it possible for a researcher to collect a large amount of data. In order to be able to perform the classifications, training is necessary, so in turn, the participant also learns about science and how it works.

**Keywords:** classification; web; astronomy; citizen science

# 1. Introducción

La astronomía moderna se enfrenta al desafío de analizar grandes volúmenes y complejidad de datos observacionales y teóricos que se generan en todo el mundo (Szalay et al., 2002). El análisis manual de estos datos resulta cada vez más ineficiente e inviable. Por ello, es necesario desarrollar habilidades en ciencia de datos y en el diseño de proyectos de software que faciliten la obtención y el procesamiento de la información.

En este contexto, presentamos una aplicación informática flexible y adaptable para problemas de clasificación visual o similares, denominada PISCIS: Platform for Interactive Search and CIizen Science. Esta herramienta puede ser descargada desde el repositorio de GitHub (<https://github.com/vanedaza/piscis>) y cuenta con documentación disponible en la plataforma ReadtheDocs (<https://piscis.readthedocs.io/en/latest/>). PISCIS está destinada a aprovechar el interés ciu-

dadano para facilitar la recolección y el análisis de datos, o bien para procesar o clasificar datos privados en proyectos de investigación mediante una interfaz conveniente. Esta aplicación no solo nos brinda competencias en el desarrollo de software, necesarias para el análisis de grandes volúmenes de datos astronómicos, sino que también ofrece a la comunidad una herramienta fácilmente adaptable para la recolección de información en sus propias investigaciones a través de la web. La aplicación básica de PISCIS consiste en una encuesta conformada por una sola pregunta relacionada con una imagen que se asocia con una pregunta de múltiples opciones preestablecidas.

En este artículo presentamos PISCIS, un software que permite el desarrollo de proyectos de investigación científica de forma colaborativa y social. PISCIS se basa en un framework que facilita la adaptación del software a las necesidades específicas de cada proyecto, sin requerir conocimientos avanzados de programación ni desarrollo de software. Con PISCIS, el usuario puede diseñar y ejecutar experimentos, analizar datos, compartir resultados y recibir retroalimentación de otros usuarios. Además, PISCIS integra herramientas de gamificación y aprendizaje automático para motivar la participación y mejorar la calidad de los datos.

Una diferencia crucial frente a otro tipo de aplicaciones (Zooniverse, 2021), es que este último tiene opciones preestablecidas de cómo realizar la investigación, mientras que con PISCIS el usuario cuenta con un framework base que se puede adaptar según la necesidad. Con conocimientos básicos de python, se espera que PISCIS permita desarrollar un proyecto en un lapso de tiempo más corto y sin la necesidad de tener un equipo de expertos en programación y desarrollo de software.

En este artículo reportamos la experiencia en el desarrollo y los resultados científicos obtenidos hasta el momento haciendo uso de PISCIS, además de posibles aplicaciones que puede tener para la recolección de información astronómica. Para ello, describimos el funcionamiento general del software, sus principales características y ventajas, así como algunos ejemplos de proyectos realizados con PISCIS en el ámbito de la astronomía.

## 2. Desarrollo

En la descripción de las herramientas desarrolladas es importante distinguir dos tipos de usuario, a saber, el usuario de PISCIS (**UP**), que descarga nuestra aplicación para desarrollar parte de su investigación y el usuario del producto (**UC**), que interactúa con el producto hecho por el investigador a través de la web o de una intranet.

### 2.1. Alcances e Infraestructura

El desarrollo de nuestra aplicación informática app se basó principalmente en que cumpla con los siguientes requerimientos:

- que sea una herramienta útil y fácil de usar para la comunidad científica en el área de la Astronomía,
- que facilite la recolección de datos mediante una pregunta simple de opción múltiple junto a una imagen, de tal manera que permita al público general proporcionar una respuesta rápida,
- que posea capacidad administrativa de creación, actualización y eliminación de objetos de contenido, con un registro de todas las ac-

ciones realizadas a través de una interface de fácil uso, y

- que los datos recopilados por la misma se guarden de forma ordenada, fácilmente accesibles.

En base a estos requerimientos, desarrollamos PISCIS sobre un *framework web*, el cual cuenta con un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular, y sirve como referencia para enfrentar y resolver nuevos problemas de índole similar. En particular, el *framework web* que reúne todos los requerimientos necesarios para la fabricación de nuestra app y la distribución de la misma es Django. El mismo sigue el patrón de diseño **MVT**: Modelo - Vista - Template, y su lenguaje de programación principal es Python.

Para Django, un **Modelo** es un tipo especial de objeto en donde se define la información de salida que es generada en la app, y que puede ser visualizada y editada con herramientas como `{\tt DB Browser for SQLite}`. En cuanto a los patrones de diseño **Vista** y **Template** de Django, la Vista describe “qué” datos serán presentados, mientras que el Template determina “cómo” se verán los mismos.

En resumen, en el diseño MVT primero definimos los Modelos, elegimos qué campos de los mismos serán observables desde las Vistas y como serán presentados a los usuarios de la web mediante los Templates.

## 2.2 Desarrollo de la aplicación

El objetivo es el desarrollo de una aplicación que genere y recolecte datos a través de imágenes acompañadas de una encuesta, que son dos elementos fáciles de modificar por cualquier UP. El formato más simple para esto es el de un sitio

web de ciencia ciudadana, cuyo diseño lo basamos en el siguiente lineamiento:

Para iniciar, instalamos un entorno virtual con el fin de aislar la configuración de nuestro proyecto de otro que estemos desarrollando, si bien este paso no es necesario, es altamente recomendable para evitar posibles problemas de compatibilidad con librerías previamente instaladas. Luego, en el entorno virtual, instalamos las librerías necesarias para el desarrollo de PISCIS, Django, pytz, sphinx, django-nose y Pillow, siendo esta última una librería de Python que permite manipular y guardar diferentes formatos de archivo de imagen, las versiones de cada una de estas librerías se encuentran en el archivo *requirements.txt* dentro del repositorio de GitHub.

Posteriormente, configuramos las opciones generales de Django, como el idioma, la zona horaria, el tipo de base de datos a usar, etc. Manipulamos el patrón de diseño MVT, comenzando con el diseño de los modelos, es decir, definimos la cantidad de bases de datos que queremos y su contenido. Como cada página en Internet necesita su propia URL (*Uniform Resource Locator*), trabajamos paralelamente en las vistas y los templates, asociando una dirección URL a cada una de las primeras y un archivo formato.html a los segundos.

Como último paso lanzamos PISCIS, el UP puede hacer uso de servidores públicos, privados o incluso usar su ordenador como servidor local.

## 2.3 Prueba

El objetivo de las pruebas es mejorar el manejo, rendimiento y optimización del sitio web. Dado que trabajamos sobre un *framework web* que incluye sus propias pruebas la verificación del sitio se redujo a:

- Verificar el tiempo que se demora en cargar la página, las imágenes, vídeos, archivos multimedia, etc.
- Mejorar los problemas de codificación del lenguaje JavaScript, CSS, HTML que son parte de los templates.
- Detectar error 404, el cual se refiere a páginas no encontradas en el servidor.
- Detectar enlaces rotos que afectan el posicionamiento de una web.
- Realizar pruebas en estilo de código y funciones programadas.

### 3. Ejemplos del uso de PISCIS

PISCIS es una herramienta que puede ser adaptada para cualquier encuesta basada en una imagen, por ejemplo, la clasificación de raza de gatos, tipos de mariposas, hongos, productos comerciales, etc. Sin embargo, hemos diseñado

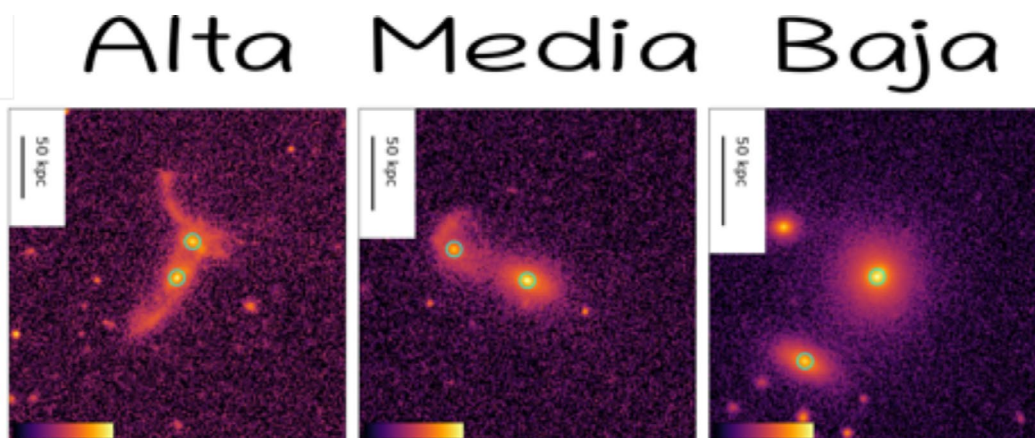
PISCIS con el propósito de recolectar información astronómica, como lo es, la clasificación de estrellas variables o exoplanetas a partir de series temporales visualizadas en imágenes, la morfología de galaxias y una infinidad de otros estudios, donde la observación detallada de características de un evento u objeto astronómico permite su clasificación.

#### 3.1 PISCIS usado en la clasificación de pares de galaxias

Como caso de estudio, PISCIS fue implementada para recolectar datos, presentando al usuario (UC) una imagen de un par de galaxias cercanas y solicitando seleccionar el tipo de interacción observada en base a las opciones ‘‘alta’’, ‘‘media’’ y ‘‘baja’’ (ver Figura 1)’. El UC vota qué tipo de interacción cree observar para el par de galaxias presentado, utilizando como referencia la documentación y ejemplos incluidos en la página web destinada para este producto del uso de PISCIS.

Figura 1

Ejemplo de imágenes a votar con el tipo de interacción en un par de galaxias.



### 3.1.1 Datos

Utilizamos un catálogo de pares de galaxias del Sloan Digital Sky Survey (SDSS, (SDSS, York et al. (2000)), provisto por O'Mill & Rodriguez, el cual cuenta con 8182 sistemas, con mediciones de corrimiento al rojo entre  $z=0.0105$  y  $z=1.9967$ , y una media de  $z=0.0913 + 0.0383$ . El mismo está definido a partir de las diferencias de velocidad, distancia y magnitud aparente de las galaxias que conforman el par y teniendo en cuenta criterios de aislamiento de este sistema con otras galaxias (Rodriguez et al., 2020). De este catálogo tomamos una muestra de 399 sistemas, y generamos las imágenes a través de SCORPIO (<https://github.com/josegit88/SCORPIO>). Este *software* permite a partir de información de las coordenadas Ascensión Recta, Declinación y corrimiento al rojo (**AR, DEG, z**) del par de galaxias generar una imagen con la resolución deseada en alguno de los relevamientos disponibles: SDSS, Wide Field Infrared Survey Explorer (WISE, Wright et al. (2010)) o Two Micron All Sky Survey (2MASS, Skrutskie et al. (2006)).

### 3.1.2 Implementación

El proceso para utilizar PISCIS para configurar una base de datos con una interface interactiva para clasificación de imágenes se puede resumir como sigue:

Luego de descargar PISCIS del repositorio en Github y de instalar las librerías requeridas, manipulamos dos de los valores por defecto de las opciones fundamentales de la app, en el archivo *settings.py* elegimos idioma español y la zona horaria de Argentina.

Seguidamente, agregamos la información contenida en la encuesta, es decir, el tipo de interacción. Este paso se realiza desde el archivo */encuesta/models.py* donde se diseñan las bases de datos de las imágenes y la votación, el diseño de esta base de datos se hace a partir de Programación Orientada de Objetos en Python definiendo dos clases, "Images" y "Choice". En la Clase Choice se encuentran los campos a votar, identificación del usuario y la identificación de la imagen a votar. Con los campos de las bases de datos definidos, queda la carga de datos de las imágenes y de la encuesta, la primera por parte de UP en el proceso de desarrollo y la segunda por los UC quienes son los visitantes de la página web.

Con la encuesta definida la carga de información y control sobre el sitio web queda a cargo de un superusuario creado por el UP.

Con la parte funcional del proyecto finalizada, se realiza el diseño de la apariencia varios Templates en html, css y js contenidos en PISCIS, los cuales fácilmente se pueden modificar, desactivar o eliminar según se desee.

Para el desarrollo de esta encuesta, los Templates son usados para que la página web cuente con cinco pestañas, una principal que se conecta con una página de contacto, inicio de sesión, encuesta y otra con una descripción sobre la app que se usó para crear la encuesta, es decir, el proyecto PISCIS.

Finalizados los cambios en la parte funcional y estética de la encuesta de pares de galaxias, se configuró la app para ser desplegada en uno de los servidores de la Universidad Nacional de Córdoba.

La misma se puede acceder a través de la url del proyecto de clasificación de pares de galaxias (<https://paresdegalaxias.iate.conicet.unc.edu.ar/>).



Este proyecto en particular es de acceso abierto aunque requiere un registro previo.

### 3.1.3 Análisis y manejo de los datos recopilados

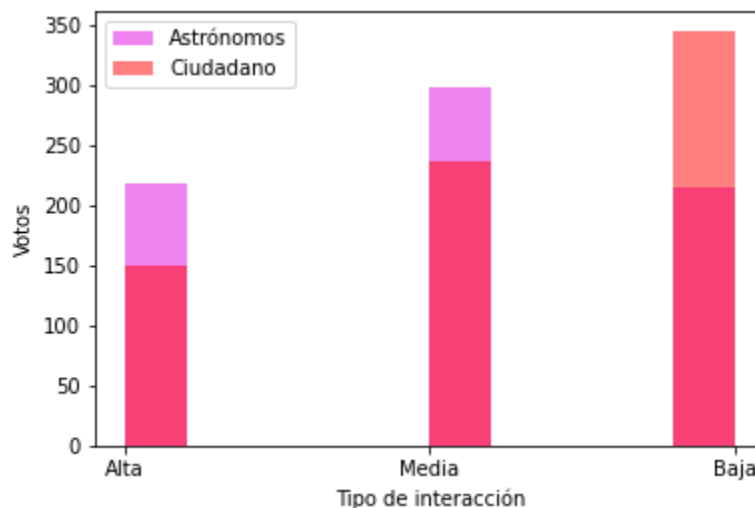
Desde el lanzamiento del proyecto de Clasificación de galaxias, PISCIS ha recolectado más de 700 votos del público general, permitiéndonos hacer algunos análisis estadísticos preliminares para testear el funcionamiento de la aplicación y la calidad de los datos recolectados.

Como ya mencionamos, un aspecto a tener en cuenta en nuestro proyecto es que los usuarios a quienes está destinada la encuesta no son personas especializadas en el área, lo que implica que los datos recolectados por la misma puedan

diferir de la clasificación real de los objetos. Para estimar cuánto varían los datos proporcionados por el público masivo respecto a la clasificación esperada, se le pidió a un grupo de astrónomos clasificar con el mismo criterio las imágenes presentadas. Se tomó una muestra al azar de 700 votos generados por astrónomos y comparamos la distribución de los mismos con la de los votos generados por los participantes de la encuesta, ambos resultados pueden verse en la Figura 2. Como se puede observar las distribuciones difieren considerablemente entre sí, y puede verse que el público general tiende a clasificar la interacción entre las galaxias como “Baja” por sobre “Alta” y “Media”. Por el momento no podemos realizar un análisis más profundo del fenómeno, hasta tener un mayor volumen de datos.

Figura 2

Diagrama de barras de votos hechos por astrónomos y ciudadanos.

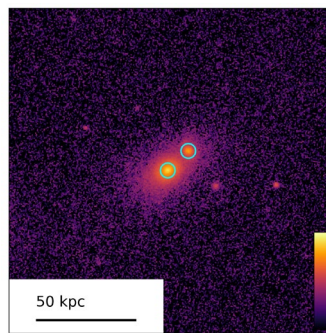


Otra función muy útil con la que contamos es el poder hacer un análisis individual de cada una de las imágenes. Por ejemplo, la **Figura 3** cuenta con 5 votos: 3 Alta, 1 Baja y 1 Media, en este caso en particular el valor más votado coincide con el asignado por los astrónomos para este par

de galaxias. Conocer la estadística individual de las imágenes resulta muy útil para poder estimar las características de las mismas que facilitan su correcta o incorrecta clasificación, y así aumentar la calidad de los datos recolectados.

**Figura 3**

Ejemplo de imágenes a votar con el tipo de interacción en un par de galaxias.



Al igual que con las imágenes, también es posible hacer estadística individual sobre cada uno de los usuarios que han usado la plataforma. De esta manera podemos eliminar de nuestros datos los votos de usuarios que produzcan algún tipo de sesgo en los resultados finales, por ejemplo: un usuario que clasifique todas las imágenes iguales.

## 4. Discusión

En este trabajo se reporta la experiencia del desarrollo de un proyecto de *software* para ayudar a un usuario en la clasificación visual de imágenes. Hemos adquirido habilidades para la elaboración de un producto que adquiera información astronómica, así como también la experiencia de trabajo interdisciplinaria con el fin de desarrollar una herramienta que puede ser usada para cualquier estudio astronómico en donde se necesite

realizar alguna clasificación a partir de un análisis visual. Además, generamos una muestra de pares de galaxias con descripción de su tipo de interacción al realizar una inspección visual. Esta herramienta será publicitada mediante las oficinas de prensa del OAC y del IATE para invitar al público en general a participar del proyecto. Como trabajo a futuro, analizaremos el impacto en la comunidad científica y la eficiencia de usar información provista por la comunidad no experta. Se analizarán además los datos recabados para determinar diferencias en la percepción de usuarios expertos y de usuarios del público en general, y se aprovechará el lanzamiento de la herramienta para trabajar en la difusión de conceptos de astronomía extragaláctica.

**Agradecimientos:** Este trabajo ha sido financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina



(CONICET) y la Secretaría de Ciencia y Tecnología (SeCyT) de la UNC.

## Referencias:

Facundo Rodríguez, Elizabeth J. González, Ana L. O'Mill, Enrique Gaztañaga, Pablo Fosalba, Diego García Lambas, Mar Mezcua, Małgorzata Siudek. Close galaxy pairs with accurate photometric redshifts. *Astronomy & Astrophysics*, Volume 634, id.A123, (2020): 123-133. [10.1051/0004-6361/201937215](https://doi.org/10.1051/0004-6361/201937215)

Skrutskie, MF *et al.* (1997). The Two Micron All Sky Survey (2MASS): descripción general y estado. En: Garzón, F., Epchtein, N., Omont, A., Burton, B., Persi, P. (eds) *The Impact of Large Scale Near-IR Sky Surveys*. Biblioteca de Astrofísica y Ciencias Espaciales, vol 210. Springer, Dordrecht. [https://doi.org/10.1007/978-94-011-5784-1\\_3](https://doi.org/10.1007/978-94-011-5784-1_3)

Alexander S. Szalay, Jim Gray, Jan vandenBerg. Petabyte Scale Data Mining: Dream or Reality?. *American Astronomical Society Meeting Abstracts, Proceedings of the SPIE*, Volume 4836, (2002): 333-338. <https://doi.org/10.1117/12.461427>

Wright E.L., et al. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *The Astronomical Journal*, Volume 140, Issue 6, (2010): 1868-1881. [10.1088/0004-6256/140/6/1868](https://doi.org/10.1088/0004-6256/140/6/1868)

York D.G., et al. «The Sloan Digital Sky Survey: Technical Summary». *The Astronomical Journal*, Volume 120, Issue 3, (2000): 1579-1587. <http://iopscience.iop.org/1538-3881/120/3/1579>

Zooniverse (2021). The world's largest and most popular platform for people-powered research. Recuperado de <https://www.zooniverse.org/>