

# Gestión automatizada de datos abiertos y portales de transparencia en las Universidades Públicas Argentinas.

## Fundamentación

La gestión de datos abiertos y la transparencia en la administración de la información requieren herramientas confiables, seguras y ampliamente adoptadas a nivel mundial.

En este contexto, para la gestión de información pública, la Universidad Nacional de Córdoba adoptó la utilización de **Apache Superset** y de **CKAN**, que se presentan como soluciones **Open Source** para la construcción de portales modernos y accesibles, asegurando control total sobre la información.

Al combinar ambas se logra una solución robusta, segura y reconocida internacionalmente para cualquier iniciativa de datos abiertos o portal de transparencia.

## Motivación

La decisión de avanzar en la creación de un Portal de Transparencia surge del compromiso asumido por la actual gestión rectoral con una política institucional basada en la apertura, la rendición de cuentas y la participación ciudadana.

Este compromiso se reforzó, además, por un contexto en el que la legitimidad de las universidades públicas ha sido persistentemente cuestionada por el gobierno nacional.

En este contexto, la elaboración de un índice de transparencia por parte de la Agencia de Acceso a la Información Pública (autoridad de aplicación de la Ley 27.275 de Acceso a la Información Pública) constituyó un incentivo adicional, entendiéndose como un dispositivo apropiado para desmontar discursos que buscaban minar la legitimidad de la Universidad en la opinión pública.

Los resultados obtenidos por la UNC en las primeras mediciones del índice evidenciaban la necesidad de fortalecer los mecanismos institucionales de sistematización de información que, si bien se encontraba disponible, estaba desorganizada.

El esfuerzo dio sus frutos, mediante una escalada constante de la institución en las mediciones sucesivas de este índice de transparencia.

En el plano interno a la organización, si bien existían esfuerzos dispersos y de bajo alcance en materia de gestión de datos, la información se encontraba fragmentada y carecía de un espacio centralizado de coordinación.

Esta situación puso de manifiesto la importancia de definir una política integral de manejo de datos que, además de servir a la toma de decisiones basadas en evidencia, pudiera articularse con la política de transparencia institucional como un valor agregado.

Finalmente, en un contexto de fuertes restricciones presupuestarias, se identificó la oportunidad de adoptar soluciones tecnológicas de código abierto que permitiera desarrollar un sistema sostenible, escalable y compatible con el ecosistema de herramientas SIU, garantizando así autonomía, eficiencia y soberanía tecnológica.

## Objetivos

- **Sistematizar los flujos de datos para abastecer un portal de transparencia que permita mejorar el desempeño de la UNC en el índice nacional de transparencia.**
- **Articular la política de transparencia con la gestión interna de datos para la generación de tableros de indicadores para la toma de decisiones basada en evidencia.**
- **Implementar soluciones tecnológicas mediante la utilización de plataformas Open Source sin restricciones de licencias.**
- **Dejar abierto a otras universidades la tecnología nueva que se genere para cumplir estos objetivos.**

## Etapas

1. Requerimientos
  - La primera etapa consistió en hacer un listado de necesidades, tanto en relación a la información a cubrir como en requerimientos tecnológicos..
  - Esta etapa se abordó de manera colaborativa, mediante la participación de diferentes actores relevantes en el uso de información para la toma de decisiones en un espacio centralizado, de manera iterativa, hasta que se logró saturar la discusión respecto de los usos potenciales de la información aplicada a la gestión interna de la institución. Asimismo, la articulación con la Agencia de Acceso a la Información Pública permitió identificar claramente las necesidades de información a incluir en el portal de transparencia.
2. Generación/Captura de datos - Investigación de herramientas open source
  - Con las necesidades explicitadas en un documento, se allanó el camino hacia la evaluación completa del proceso a implementar, tanto en términos de las herramientas tecnológicas necesarias, como en la identificación de los conjuntos de datos que serían necesarios sistematizar para su consumo. Si bien el diseño original de la estrategia se basó centralmente en Wichi y en los datos consolidados en el Datawarehouse, se identificaron otras necesidades de información planteadas por la AAIP (Agencia de Acceso a la Información Pública) que no estaban contempladas previamente en ninguna de las soluciones de Inteligencia de Negocios pre-existentes, lo que hizo necesario un replanteo de la estrategia original.
3. Instalación de Superset y CKAN (ckan.org)
  - Lo primero en instalar fue Superset. Contábamos con poco tiempo para la puesta en marcha, así que se hizo un listado de prioridades en la producción

de los tableros para que la creación fuera de lo más esencial a lo más deseable.

- Para la implementación de CKAN se contó con ayuda externa, especializada en el tema de datos abiertos mediante este software y ellos fueron los encargados de desarrollar los plugins necesarios para la toma de datos desde fuentes externas.
- Ambas instalaciones fueron realizadas con Docker y un ambiente de despliegue continuo utilizando Kubernetes. Esto agiliza las pruebas e implementación de los entornos de desarrollo, testing y producción.

#### 4. Producción de tableros

- Analizamos los datos a usar internamente y los datos a publicar y creamos las consultas SQL pertinentes para cada caso.

#### 5. Construcción de los endpoints de comunicación entre CKAN y Superset

- Superset ya expone un API moderna que conseguimos abrir de manera segura a través de un proxy para que el portal de datos abiertos pueda consumir.

#### 6. Publicación de tableros y sincronización de datos en CKAN

- Decidimos luego del asesoramiento de la Open Knowledge Foundation ([okfn.org](https://okfn.org)) crear una nueva extensión libre de CKAN que permita extraer datos de una instancia de Superset y publicarlo en un portal de datos abiertos con CKAN

## Tecnología utilizada

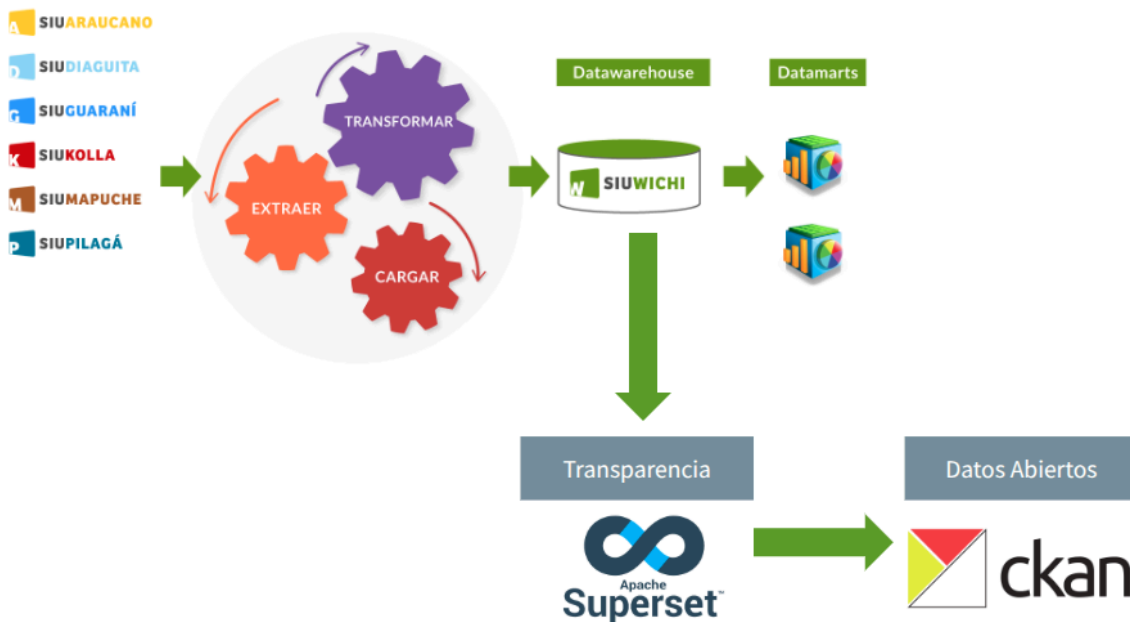
<p><b>¿Qué es Apache Superset?</b></p> <p>Apache Superset es una herramienta moderna de <b>visualización y análisis de datos</b>, que permite transformar grandes volúmenes de información en gráficos, tableros e informes comprensibles para todo público. Su fortaleza radica en que facilita la exploración de datos sin necesidad de conocimientos técnicos avanzados, lo que contribuye a mejorar la transparencia y la toma de decisiones basada en evidencia.</p>	<p><b>¿Qué es CKAN?</b></p> <p>CKAN es una de las plataformas más utilizadas en el mundo para la <b>publicación de datos abiertos</b>. Gobiernos, universidades y organizaciones internacionales confían en ella para poner a disposición datasets públicos de manera organizada, accesible y reutilizable. Con CKAN se pueden publicar, catalogar y administrar grandes volúmenes de información, facilitando que la ciudadanía, investigadores o periodistas puedan acceder a datos confiables y en formatos reutilizables.</p>
---	---

## Beneficios principales de utilizar Superset + CKAN

- **Automatización de procesos:** La combinación de estas herramientas permite la gestión de la información pública de manera automática, sin necesidad de destinar recursos a la carga y mantenimiento de datos en los portales de transparencia.
- **Open Source y estándar mundial:** Son **soluciones abiertas, de uso gratuito** y con una comunidad activa a nivel global que asegura innovación continua y soporte comunitario.
- **Instalación On-Premise:** Ambas herramientas pueden instalarse en servidores propios de la institución, evitando que la información se aloje en la nube de terceros y garantizando así mayor seguridad y **soberanía sobre los datos**.
- **Transparencia y acceso:** Superset toma los datos automáticamente (luego de un proceso de extracción de los datos desde los sistemas SIU) creando tableros y visualizaciones fáciles de interpretar y al mismo tiempo CKAN permite la integración desde Superset como fuente de datos, agilizando el proceso de publicación de la información en CKAN. Esto permite el acceso a la misma información de dos formas distintas (en formato de archivo y gráficamente).
- **Escalabilidad y flexibilidad:** Se adaptan tanto a proyectos pequeños como a portales nacionales de datos abiertos, asegurando que el crecimiento no limite su uso.
- **Confianza y legitimidad:** Al ser utilizadas por organismos de todo el mundo, aportan reconocimiento y credibilidad a cualquier iniciativa de apertura de datos.

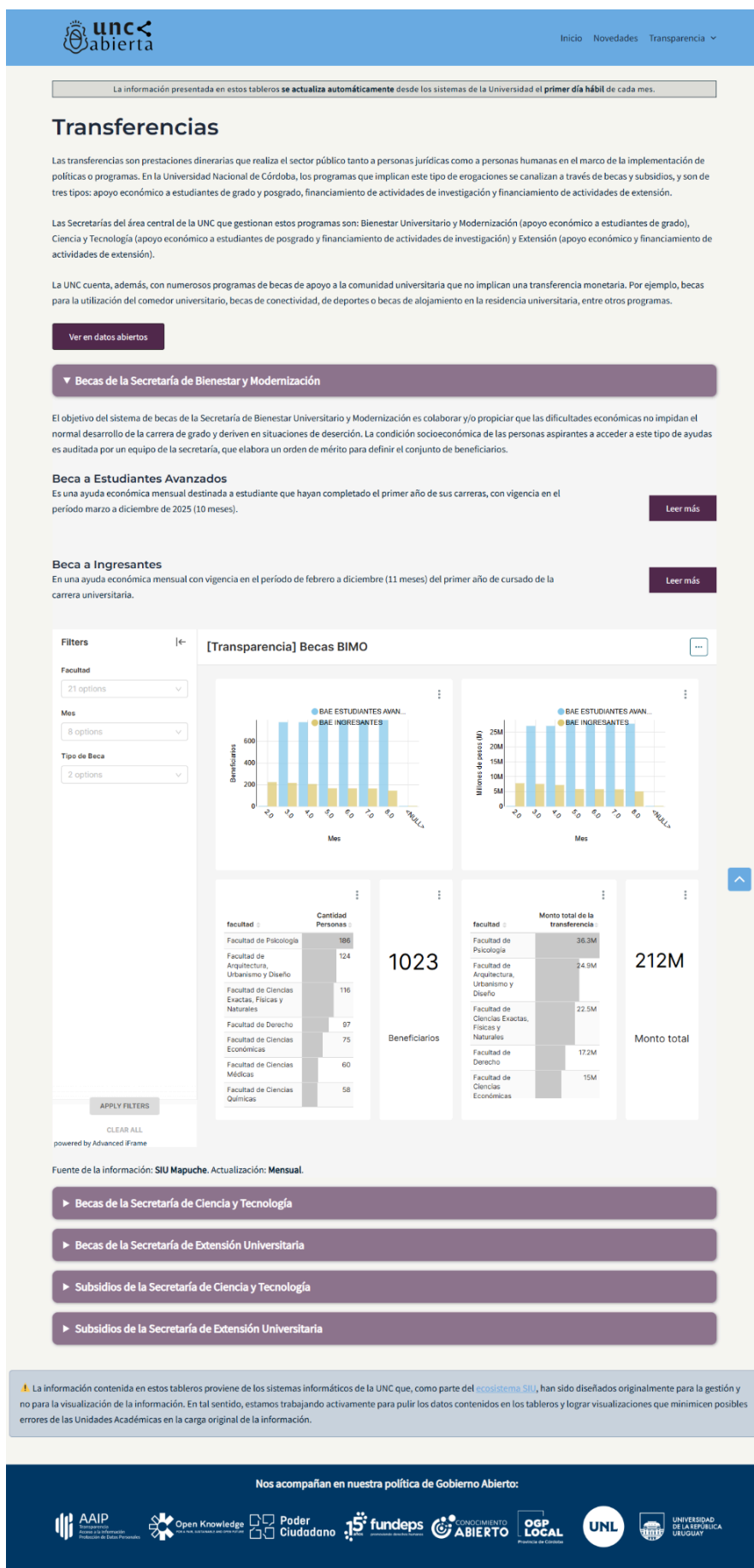
## Resultados obtenidos

Como resultado de este proceso de implementación, se logró obtener una **infraestructura que va automáticamente desde los sistemas de gestión hasta el portal de datos abiertos sin intermediarios** y sin generar dependencia con proveedores externos.



Además de los datos almacenados actualmente en los sistemas, se logró centralizar reportes/información no sistematizada (archivos en pdf, informes, etc) y visibilizar las necesidades de capturar otra información que se llevaba de manera más descentralizada.

**Portal de Transparencia:** Sitio web hecho en Wordpress que tiene embebidos los diferentes tableros realizados en Superset. Se realizaron mini dashboards que contienen información muy específica y luego, los mismos se muestran organizados de manera más eficiente en el portal.





## Conclusiones

Para terminar, podríamos resumir los aspectos positivos y desafíos enfrentados en los siguientes ejes:

### 1. Mejora en la Gestión y la Toma de Decisiones

El proyecto ha permitido avanzar hacia una **gestión basada en información**, priorizando la ética de lo público. Al articular la política de transparencia con la gestión interna de datos, se facilita la generación de tableros de indicadores para la toma de decisiones basada en evidencia. Esta estrategia centralizada mejora la capacidad de análisis de la UNC, potenciando una **visión más acabada de sus actividades** y logrando una inteligencia institucional que permite la adaptación y el monitoreo continuo del impacto de sus acciones.

### 2. Coordinación y Optimización de Recursos

La implementación resolvió problemas de ineficiencia originados en la gestión descentralizada de datos, que ocasionaba duplicidad de iniciativas para la gestión de la información. Mediante la creación de espacios comunes para el planteo de necesidades, diferentes Unidades Académicas lograron identificar necesidades comunes, lo cual permitió establecer una agenda para avanzar de manera ordenada en la sistematización de datos, incrementando la **colaboración entre equipos interdisciplinarios**. Esta estrategia permitió, además, una sinergia con el origen de los datos, resolviéndose iniciativas que tienden a modificar procesos de carga para una mejor gestión de los datos y su proceso en información útil para la toma de decisiones. Estas modificaciones de procesos en la carga de los datos hubiese sido estéril si se realizaba de manera aislada, y toma dimensión cuando se define desde un espacio centralizado en donde están representadas todas las unidades académicas intervinientes en el proceso. Además, al unificarse el proceso de obtención y gestión de los datos, se logra garantizar una política de seguridad para la información crítica.

### 3. Transparencia, Gobernanza y Apertura

Impulsada por un compromiso institucional con la rendición de cuentas y la participación ciudadana, y reforzada por el contexto de cuestionamiento a las universidades públicas, la iniciativa logró una **mayor transparencia y responsabilidad en la gestión pública**. El esfuerzo por sistematizar la información desorganizada condujo a una **mejora constante de la institución en las mediciones sucesivas del índice nacional de transparencia**. Además, se definió la implementación de una política de apertura de la información generada en formato reutilizable, conforme a los estándares internacionales en datos abiertos.

### 4. Fortalezas Tecnológicas y Soberanía

El proyecto adoptó una estrategia tecnológica moderna y soberana, utilizando soluciones Open Source como **Apache Superset y CKAN**. Estas herramientas son gratuitas, de estándar mundial, y al permitir la **instalación On-Premise** (en servidores propios), garantizan la **soberanía sobre los datos**, ya que la información no se aloja en la nube de terceros. La combinación de Superset y CKAN permite la **gestión automática de la información pública**, creando una infraestructura que va automáticamente desde los sistemas de gestión hasta el portal de datos abiertos **sin intermediarios** y sin generar dependencia de proveedores externos.

## Proyección/Propuestas hacia el futuro

La **Universidad Nacional de Córdoba** tiene como propósito dejar claramente documentadas las tecnologías que conforman este modelo de trabajo, poniendo a disposición su experiencia y el conocimiento adquirido. Al mismo tiempo, busca acompañar y brindar soporte a otras universidades del sistema público que deseen implementar sus propios portales de datos abiertos basados en la información proveniente de los sistemas SIU. Se entiende que existe una oportunidad excepcional y estratégica: que todas las universidades nacionales del país puedan adoptar una misma plataforma tecnológica, fortaleciendo la interoperabilidad, la estandarización de procesos y la transparencia a nivel federal.

## Links de interés

Portal de transparencia: [transparencia.unc.edu.ar](http://transparencia.unc.edu.ar)

Portal de datos abiertos:

<https://datosabiertos.unc.edu.ar/organization/universidad-nacional-de-cordoba>

Ckan: <https://ckan.org/>

Open Knowledge Foundation: [okfn.org](http://okfn.org)

Superset: <https://superset.apache.org/>

Si te interesa saber más sobre este proyecto, comunicate con [ucumari@informatica.unc.edu.ar](mailto:ucumari@informatica.unc.edu.ar) - [melisa.caffaratti@unc.edu.ar](mailto:melisa.caffaratti@unc.edu.ar)