

Gestión de Proyectos de Investigación y Ciencia Abierta

María Camila Caraballo¹
Laura Sarif Rivera²
Zaira Alejandra García³
Sergio Andrés Melo⁴
[Repositorio en GitHub](#)

Proyecto transversal - Primer Avance

El presente taller busca que se comprenda la importancia de la automatización en el proceso de investigación, identificar las ventajas y desafíos de la automatización en proyectos científicos, desarrollar habilidades prácticas en la automatización de flujos de trabajo y de implementar scripts para la reproducibilidad de análisis de datos.

Parte 1: Configuración entorno colaborativo

- Crear un repositorio en GitHub
- Implementar una estructura de directorios clara y documentada. Note que por ahora no es necesario que esté listo el código. Dado que ya está realizado el proceso original (el código y los datos los obtendrán de los insumos de replicación del artículo escogido) solamente deben escoger una estructura de directorios que tenga sentido según lo que decidan reproducir. Tengan en cuenta que no se puede almacenar un directorio vacío en GitHub, por lo que para ello pueden escoger simplemente subir códigos con la descripción de lo que planean hacer, algo similar al ejercicio del Taller 1.
- Incluir un README.md
- Configurar .gitignore apropiadamente
- Invitar como colaborador al profesor y compañeros de equipo
- Incluir el enlace al repositorio en el documento

¹Código: 201613424

²Código: 202421904

³Código: 202425452

⁴Código: 202021527

Parte 2: Análisis inicial

- **Justificación de la selección del paper**

Para efectos de la replicación, este trabajo se enfocará en el resultado que documenta la evolución temporal del uso de Chivo Wallet tras su lanzamiento, específicamente la evidencia de que la mayoría de las transacciones ocurrieron en los primeros meses (especialmente septiembre de 2021) y que posteriormente el uso cayó de forma marcada.

Este resultado es presentado en el artículo mediante el análisis de transacciones del blockchain asociadas a Chivo Wallet y constituye una pieza central del argumento del paper: a pesar del estatus legal del bitcoin y de fuertes incentivos iniciales, el uso real del sistema como medio de pago fue bajo y decreciente. Se selecciona este resultado porque depende exclusivamente de evidencia cuantitativa basada en registros públicos del blockchain, lo cual lo hace particularmente adecuado para un ejercicio de reproducción en la materia de ciencia abierta.

Este resultado fue elegido porque cumple con tres criterios fundamentales para una replicación académica: Primero, se basa en datos abiertos. Las transacciones del blockchain de Bitcoin son públicas y verificables, lo que permite reconstruir series temporales sin necesidad de acceso a bases privadas o encuestas propietarias. Segundo, el artículo documenta explícitamente que identifica direcciones asociadas a Chivo Wallet y utiliza estas para construir indicadores de actividad (número de transacciones, usuarios activos y montos), lo cual define un procedimiento cuantitativo claro. Tercero, el resultado indica que la adopción fue principalmente transitoria y motivada por incentivos iniciales, más que por un uso sostenido como medio de intercambio.

- **Identificación de componentes necesarios para la replicación**

La reproducción de este resultado requiere tres insumos principales. En primer lugar, los datos crudos con un diccionario de variables detallado y de ser posible los factores de expansión muestral que son esenciales para reconstruir el perfil transaccional de los usuarios y entender las barreras de adopción reportadas, infiriendo la desconfianza o la preferencia por el efectivo. El elemento crítico es contar con el listado de transacciones atribuidas a Chivo Wallet y la conversión respectiva a USD, los cuales se encuentran documentado en los materiales suplementarios del artículo y en repositorios asociados al proyecto.

En segundo lugar, se requiere el código utilizado para descargar, limpiar y agregar las transacciones. Los autores ponen a disposición scripts que permiten: extraer transacciones por dirección, construir series temporales, identificar usuarios activos, calcular volúmenes diarios o mensuales.

En tercer lugar, se requiere el entorno computacional basado específicamente en el software STATA 16, librerías para manejo de datos y acceso a blockchain, y una estructura reproducible de carpetas. Idealmente, el proyecto debe ejecutarse mediante scripts automatizados que permitan reconstruir las figuras asociadas al análisis de transacciones del blockchain como se evidencian en el artículo.

El producto final de la replicación sería una serie temporal que muestre el pico inicial de actividad y su posterior declive, replicando las gráficas presentadas en el paper.

- **Evaluación preliminar de desafíos potenciales**

El principal reto es la correcta identificación de las direcciones de Chivo Wallet. Aunque las transacciones son públicas, la atribución institucional de direcciones requiere confiar en la documentación provista por los autores. Cualquier error en este paso afecta directamente los resultados. Un segundo desafío es el procesamiento de grandes volúmenes de datos del blockchain, lo cual exige capacidades computacionales básicas y manejo de programación, lo que puede representar una barrera para equipos con menor experiencia técnica.

También existe dependencia del software externo (APIs del blockchain), cuya disponibilidad o formato puede haber cambiado desde la publicación del artículo, afectando la reproducibilidad exacta. Finalmente, aunque el resultado es reproducible, su interpretación depende de decisiones metodológicas específicas, como la definición de “usuario activo” o la agregación temporal, lo cual requiere una lectura cuidadosa del código original para evitar desviaciones.

- **Documentación de requisitos técnicos**

La ejecución para la réplica cuantitativa demanda un entorno computacional basado en el software STATA 16, requiriendo la instalación previa de los paquetes `reghdfe`, `gtools` y `spmap` para asegurar la correcta ejecución de los scripts originales. El análisis se fundamenta en el procesamiento del archivo de datos `crystal.dta`, el cual contiene el registro de las 142,148 direcciones institucionales identificadas y sus 425,514 transacciones registradas hasta noviembre de 2022. Para ello, es necesario que en el flujo de trabajo se incorpore `BTC-USD_yahoofinance.csv` para asegurar la conversión de los valores de bitcoin a USD, siguiendo la recomendación metodológica de los autores para evitar que la volatilidad sustancial del criptoactivo distorsione los patrones de volumen real y asegurar la coherencia con los resultados obtenidos en la encuesta nacional.

Desde el punto de vista metodológico, la documentación técnica debe reconocer que Chivo Wallet opera como una billetera custodial, lo que implica que las transacciones directas entre sus usuarios (P2P) son procesadas de manera externa al blockchain (off-chain) debido a los altos costos y tiempos de verificación. En consecuencia, la reproducción de la evidencia se limitará a las transacciones identificables en el registro público, las cuales corresponden únicamente a movimientos entre direcciones de la propia entidad Chivo (transacciones internas) o interacciones con billeteras externas y exchanges. Este enfoque permitirá validar que los picos de actividad transaccional observados entre octubre y diciembre de 2021 responden a la dinámica de depósitos y retiros externos, confirmando el hallazgo de un uso real bajo y decreciente del sistema como medio de pago.

Parte 3: Plan de trabajo

- Cronograma detallado del proyecto

Idem.	Workstreams	2-feb	9-feb	16-feb	23-feb	2-mar	9-mar	16-mar
1 Infraestructura y Gobernanza								
1.1	Revisión del artículo y selección de resultado a replicar							
1.2	Creación del repositorio en GitHub							
1.3	Implementación de la estructura de directorios							
2 Adquisición y ETL								
2.1	Búsqueda de datos y repositorio asociado al artículo							
2.2	Descarga y validación de data para replicación							
2.3	Exploración de scripts e identificación de dependencias							
3 Replicación y QA								
3.1	Configuración del entorno de ejecución							
3.2	Ejecución de scripts							
3.3	Replicación y validación de resultados							
4 Documentación y Cierre								
4.1	Documentación del informe de replicación							
4.2	Limpieza final del repositorio y generación de metadatos							
4.3	Entrega final							

Figura 1: Cronograma de actividades y flujo de trabajo.

- Asignación de roles y responsabilidades

El equipo se organiza de acuerdo con las distintas etapas del proceso de replicación del resultado cuantitativo seleccionado del artículo, que documenta la evolución temporal del uso de Chivo Wallet a partir de datos del blockchain. Esta división busca reproducir la lógica del estudio original, separando las tareas de adquisición de datos, procesamiento, análisis y documentación reproducible.

- **Maria Camila Caraballo:** Coordinación general del repositorio del proyecto, organización de la estructura de carpetas, en específico lidera la revisión inicial del artículo y la validación de resultados (QA). Será responsable del seguimiento del avance del equipo y de consolidar los insumos producidos por cada integrante según los bloques de trabajo.
- **Laura Sarif Rivera Sanabria:** Limpieza de datos y estructuración del pipeline de código a replicar. Sus actividades incluyen la implementación de la estructura de directorios, la preparación de la data y ejecución del ejercicio. Lidera la configuración del entorno de ejecución para garantizar la replicabilidad del procesamiento.
- **Zaira Alejandra García Bernal:** Limpieza de datos y apoyo en la implementación del pipeline reproducible, asegurando la correcta transformación de los datos, desde la descarga y validación, hasta asegurar la coherencia técnica entre los scripts de procesamiento y los análisis finales.
- **Sergio Andrés Melo García:** Automatización del flujo de trabajo, control de versiones y reproducibilidad computacional. Estará a cargo de la configuración y mantenimiento del repositorio en GitHub, la documentación técnica del informe de replicación y generación de metadatos, implementando scripts que permitan ejecutar el proyecto de extremo a extremo bajo estándares de ciencia abierta.

- **Plan de comunicación del equipo**

El equipo utilizará un grupo de WhatsApp como canal principal de comunicación diaria para coordinación operativa y resolución rápida de dudas. Adicionalmente, se empleará un repositorio compartido en GitHub como espacio central de trabajo colaborativo, donde se gestionarán versiones del código, avances del análisis y documentación del proceso. Las decisiones técnicas relevantes se registrarán mediante commites en el repositorio, con el fin de mantener trazabilidad. Se realizarán reuniones virtuales semanales para revisar avances, redistribuir tareas y evaluar dificultades en la replicación del resultado seleccionado.

- **Establecimiento de hitos y entregables internos** El proyecto se estructura alrededor de hitos intermedios que permiten monitorear el avance del proceso de replicación. Estos incluyen: (i) identificación y descarga de los datos y código del artículo; (ii) configuración del entorno computacional y ejecución inicial de los scripts; (iii) reproducción de la figura o indicador seleccionado; (iv) validación de los resultados frente al paper original; y (v) elaboración del informe final documentando el proceso, las decisiones metodológicas y las limitaciones encontradas. Como entregables internos se contará con un repositorio reproducible, una tabla o figura replicada y un documento escrito que sintetiza el ejercicio, alineado con los principios de transparencia y ciencia abierta.