

Investigación colaborativa con R: Uso de la librería {pins} para equipos de investigación

II Congreso de Datos Abiertos y Metadatos

CODAM 2022

DATOS ABIERTOS EN LA CIENCIA Ecuador

26 al 28 de octubre de 2022 / Modalidad Híbrida



Alex Bajaña
Centro de Investigación Estadística
ERGOSTATS
Quito-Ecuador
alexvbr@ergostats.org

ORGANIZADORES

data-lat.org



AUSPICIANTES



Chan
Zuckerberg
Initiative

GORDON AND BETTY
MOORE
FOUNDATION

SUMARIO/CONTENIDO:

- Caso de estudio: Un vistazo al desempeño empresarial ecuatoriano
- ¿Por qué usar la librería {pins}?
- Instalación y preparación del ambiente de trabajo
- Datos, metadatos y más metadatos
- Referencias y documentación
- Agradecimientos

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

Chan
Zuckerberg
Initiative

GORDON AND BETTY
MOORE
FOUNDATION

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES



Un vistazo al desempeño empresarial ecuatoriano

Análisis exploratorio del Directorio de Empresas del INEC

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES



La investigación con registros administrativos



Ventajas

- Datos abiertos de bajo costo
- Mayor cobertura
- Reducción de errores de muestreo
- Reducen la carga administrativa

Desventajas

- Transformación a registros estadísticos previa
- Requieren de conciliación y revisión de coherencia
- Susceptibilidad
- Temporalidad

El directorio de empresas (DIEE)

Fuentes primarias:

- SRI
- IESS

Población:

- Personas naturales con RUC
- Sociedades (con o sin fines de lucro)
- RISE/AMIE
- Establecimientos de educación pública y fiscomisional

Cobertura

- Nacional con desagregación

Periodicidad:

- Anual con 1 año de rezago



El Directorio de Empresas y Establecimientos (DIEE) genera información estadística sobre la estructura empresarial ecuatoriana a partir de registros administrativos; proporciona información sobre el total de unidades económicas, que durante un período fiscal registraron movimientos económicos de ventas, personal ocupado – afiliado medido a través de plazas de empleo registrado en la seguridad social y/o realizaron una declaración al RISE, desde una perspectiva sectorial y territorial.



Resumen 2020

Una visión general de los resultados del periodo.

Variables económicas	DIEE 2020	
	2019	2020
Número de empresas	886.600	846.265
Ventas totales (millones de dólares constantes de 2007)	\$114.957	\$97.756
Plazas de empleo registrado	3.031.475	2.845.018
Masa salarial (millones de dólares constantes de 2007)	\$18.433	\$17.248



Principales resultados

Presentación de los principales resultados del DIEE.



Boletín técnico

Documento que contiene un análisis descriptivo y evolutivo del DIEE.

Pregunta de investigación: ¿Cuáles fueron los sectores económicos más afectados por la pandemia COVID-19?



Cual sea el objetivo responder esta pregunta implica:



Por que usar la librería {pins}

Control de versiones para una fácil integración de múltiples esfuerzos

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION



Octubre, 2022

ORGANIZADORES

datalat.org

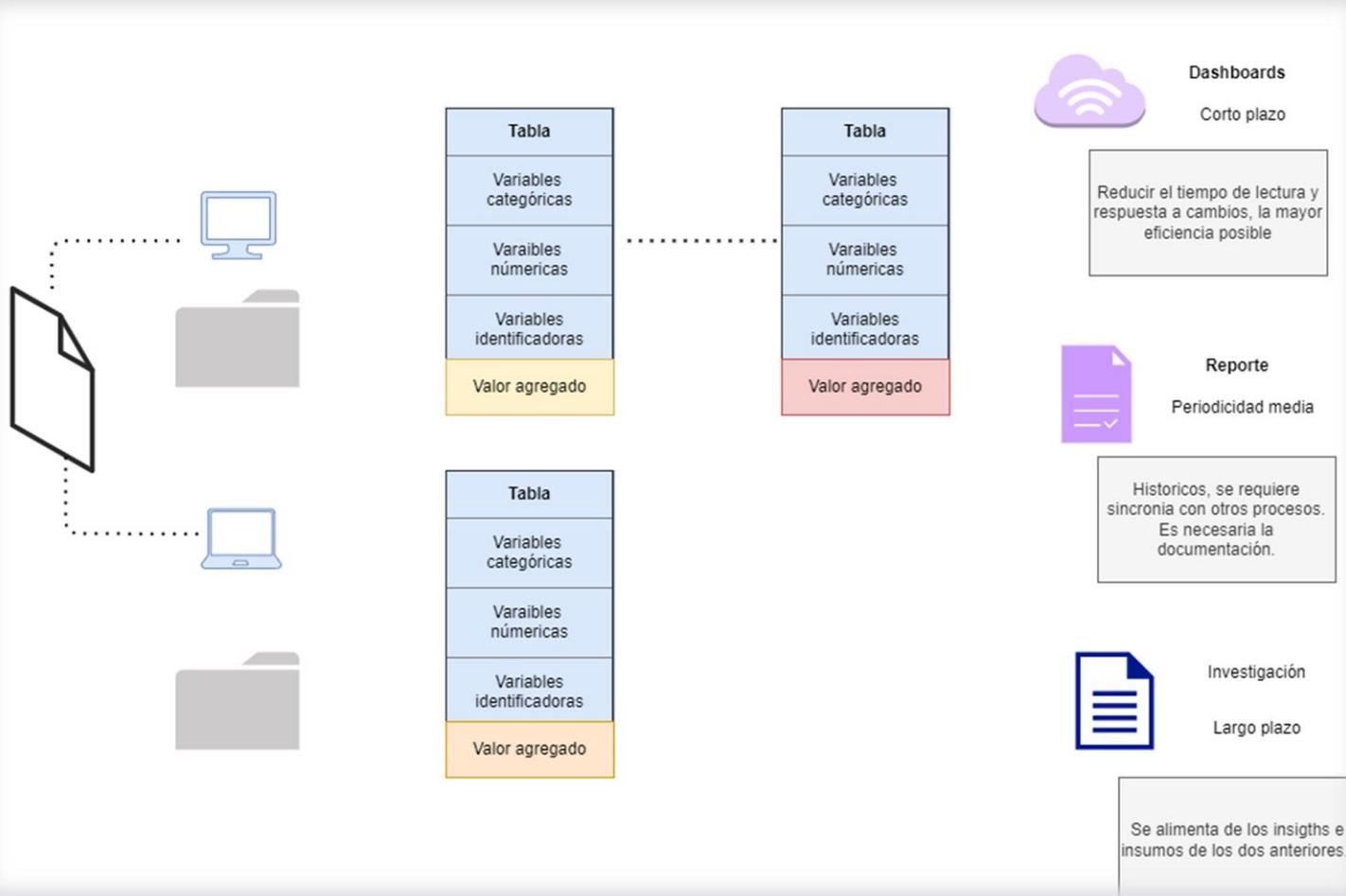
REDAM
Red Ecuatoriana De
Datos Abiertos Y Metadatos

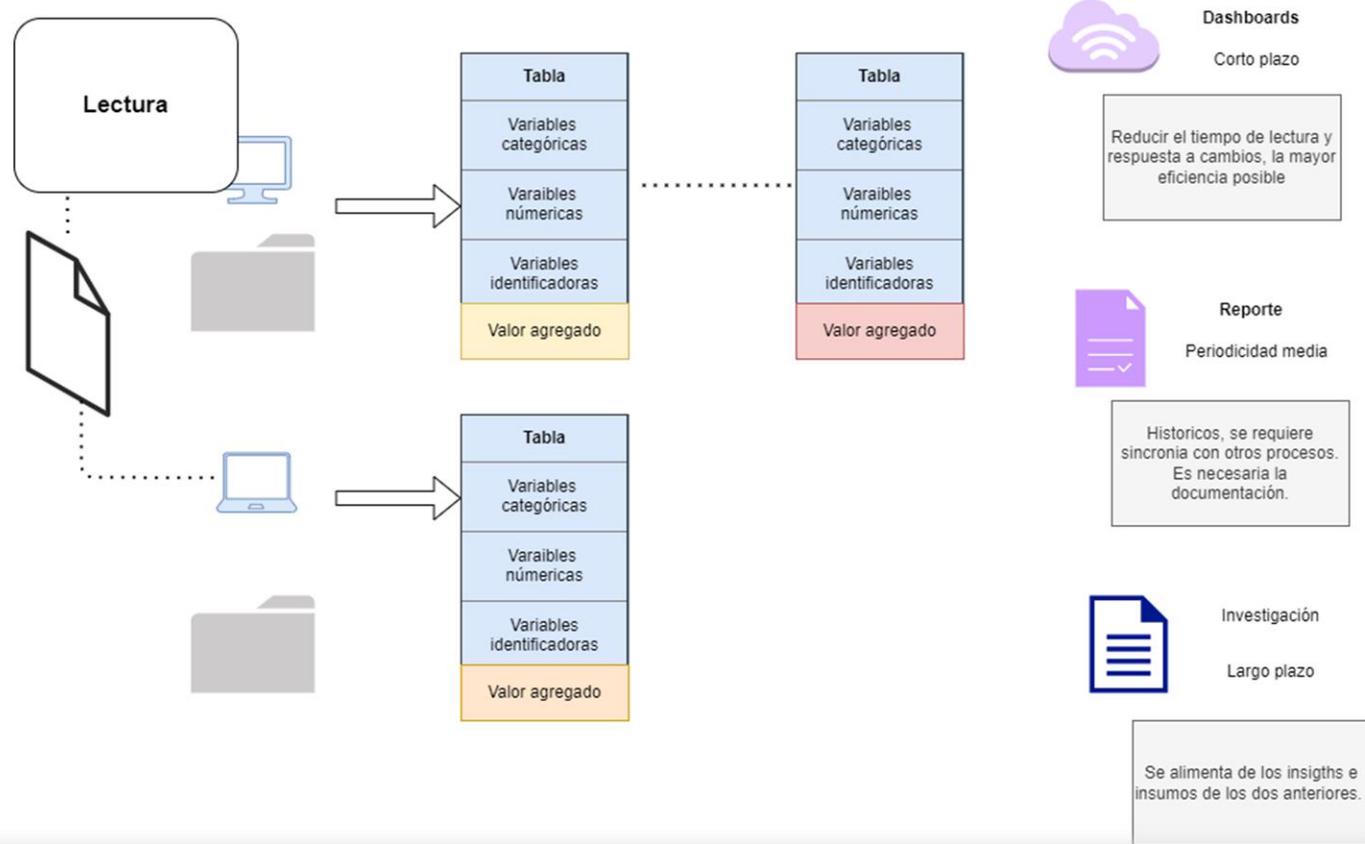
AUSPICIANTES

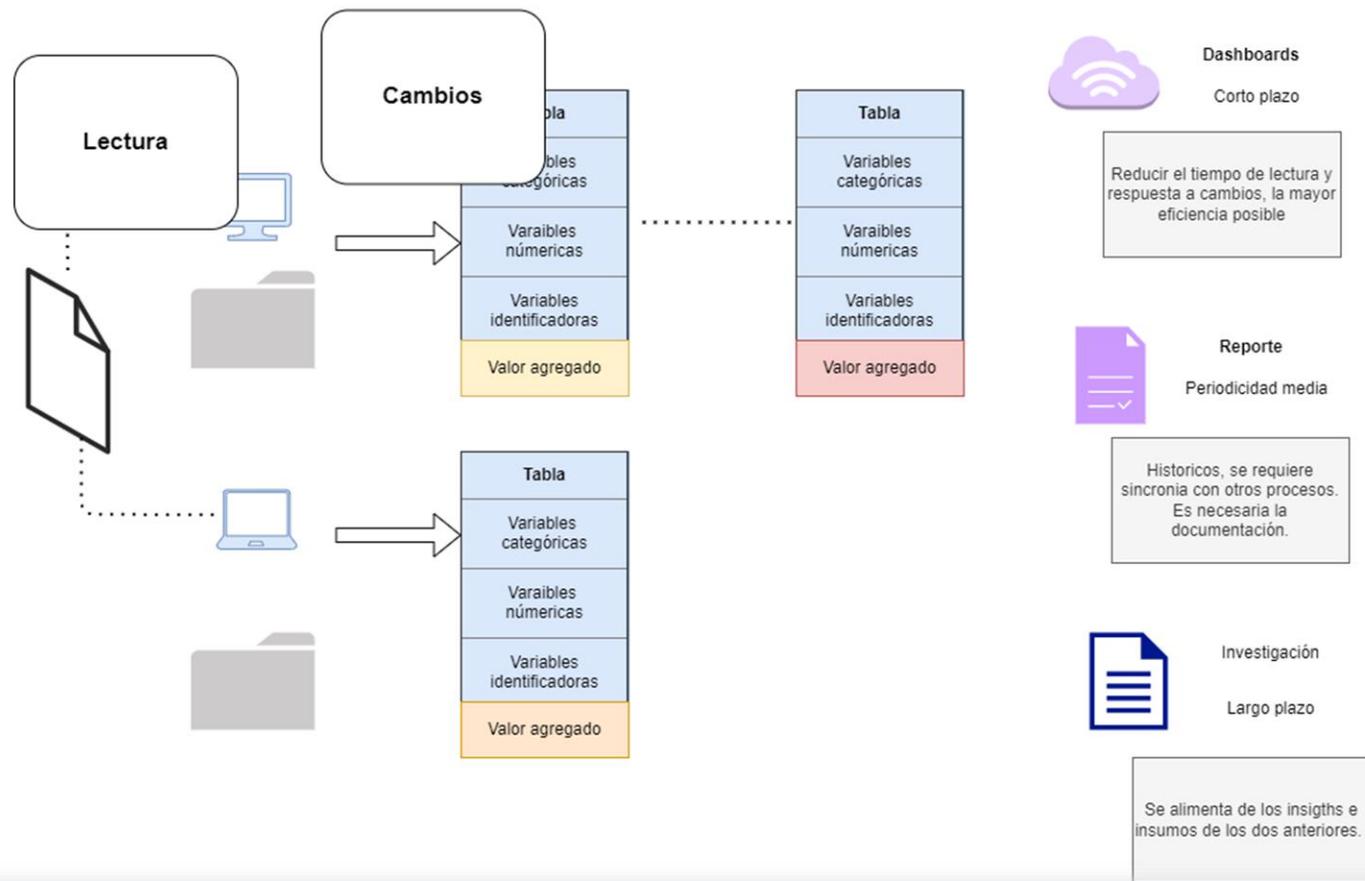
CS&S Code for
Science &
Society

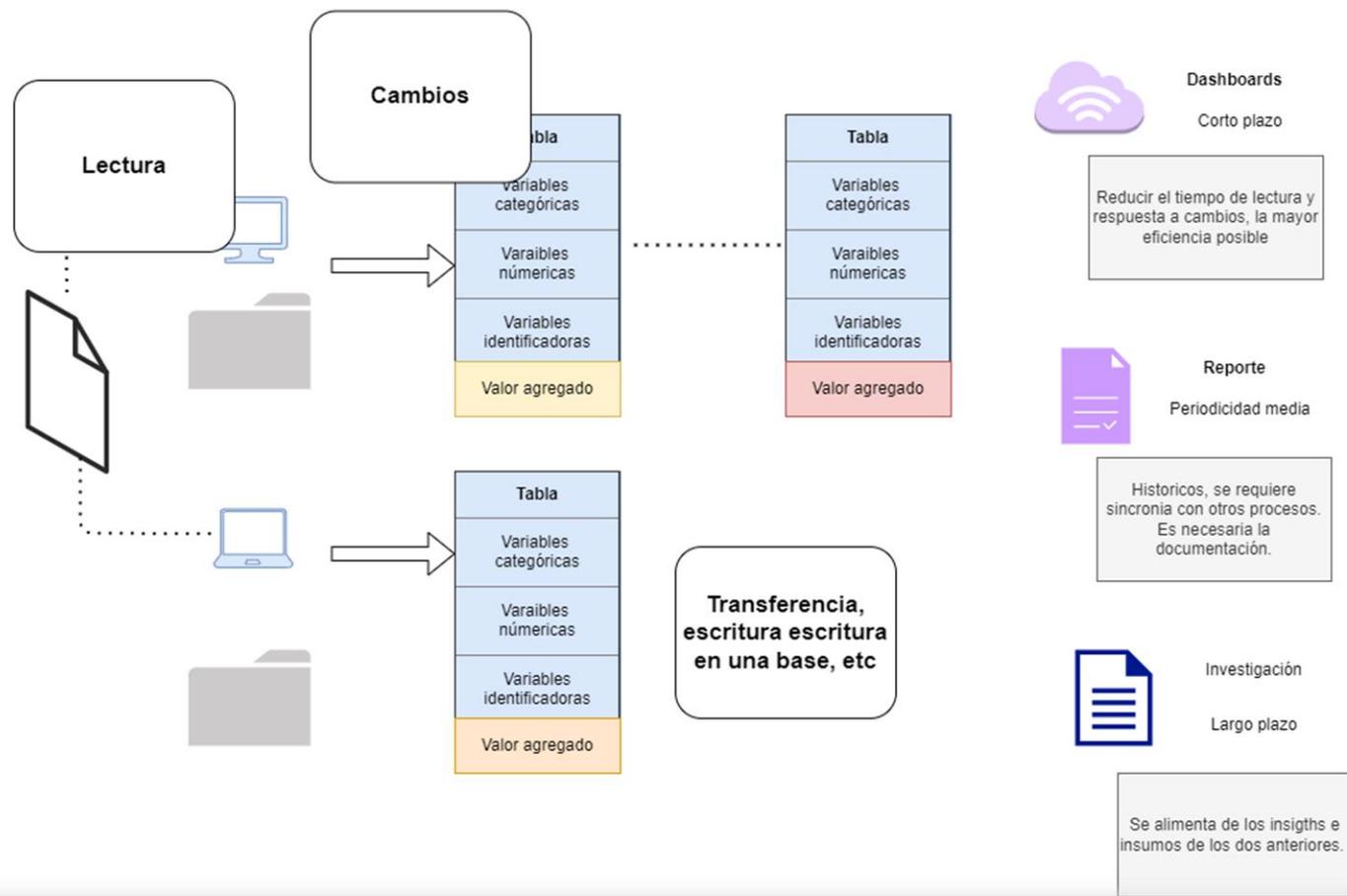
Chan
Zuckerberg
Initiative

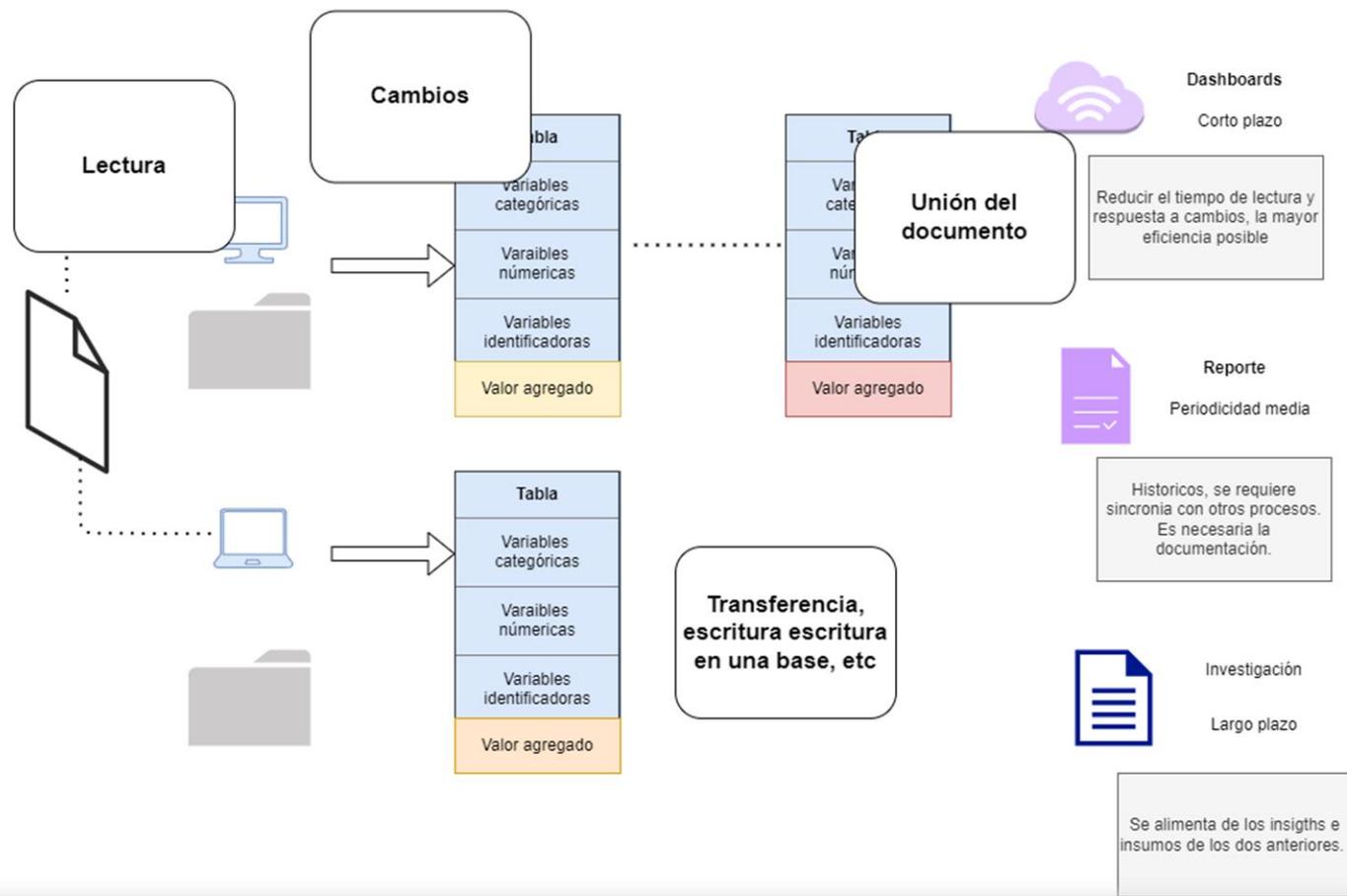
GORDON AND BETTY
MOORE
FOUNDATION

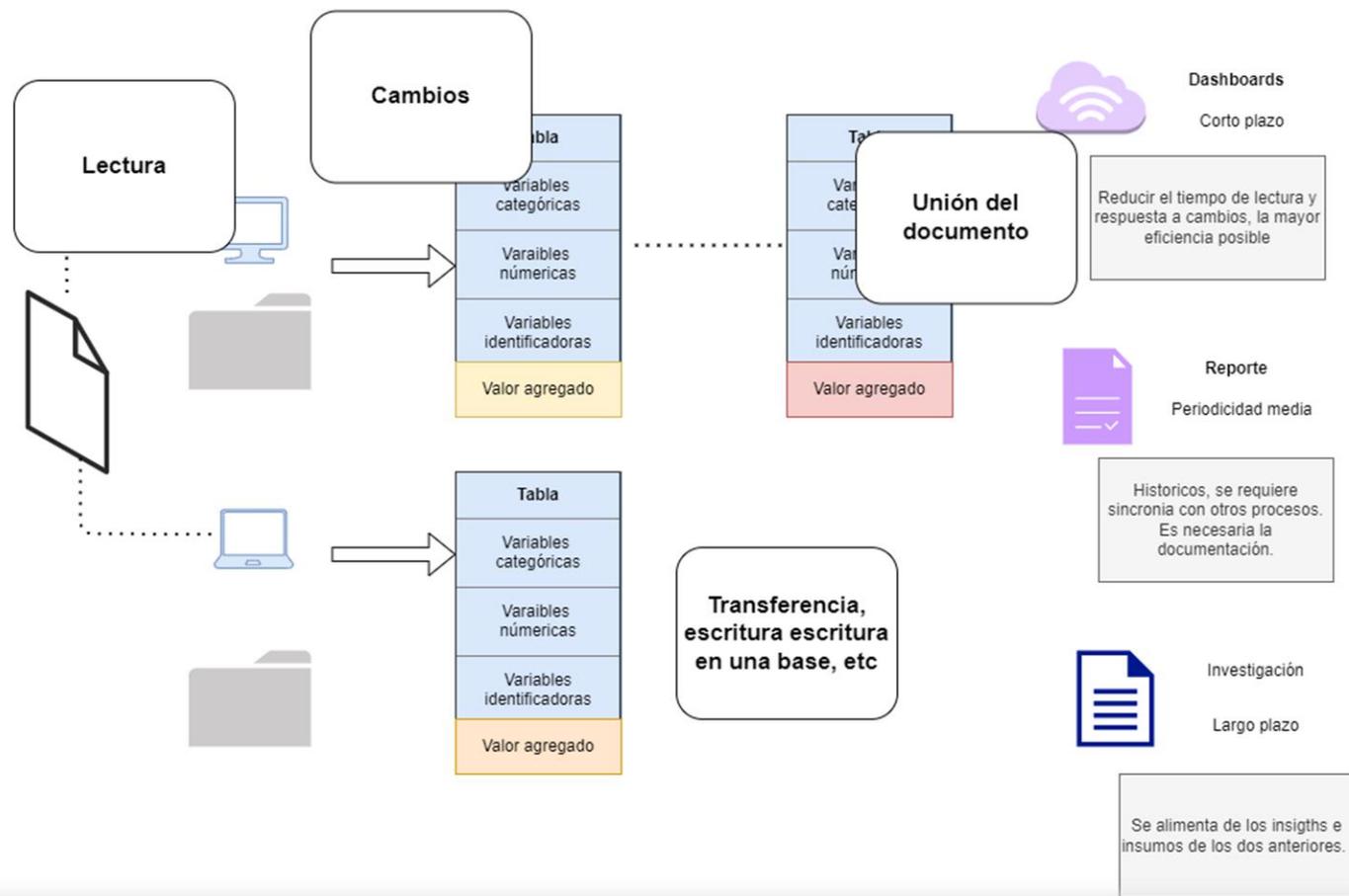


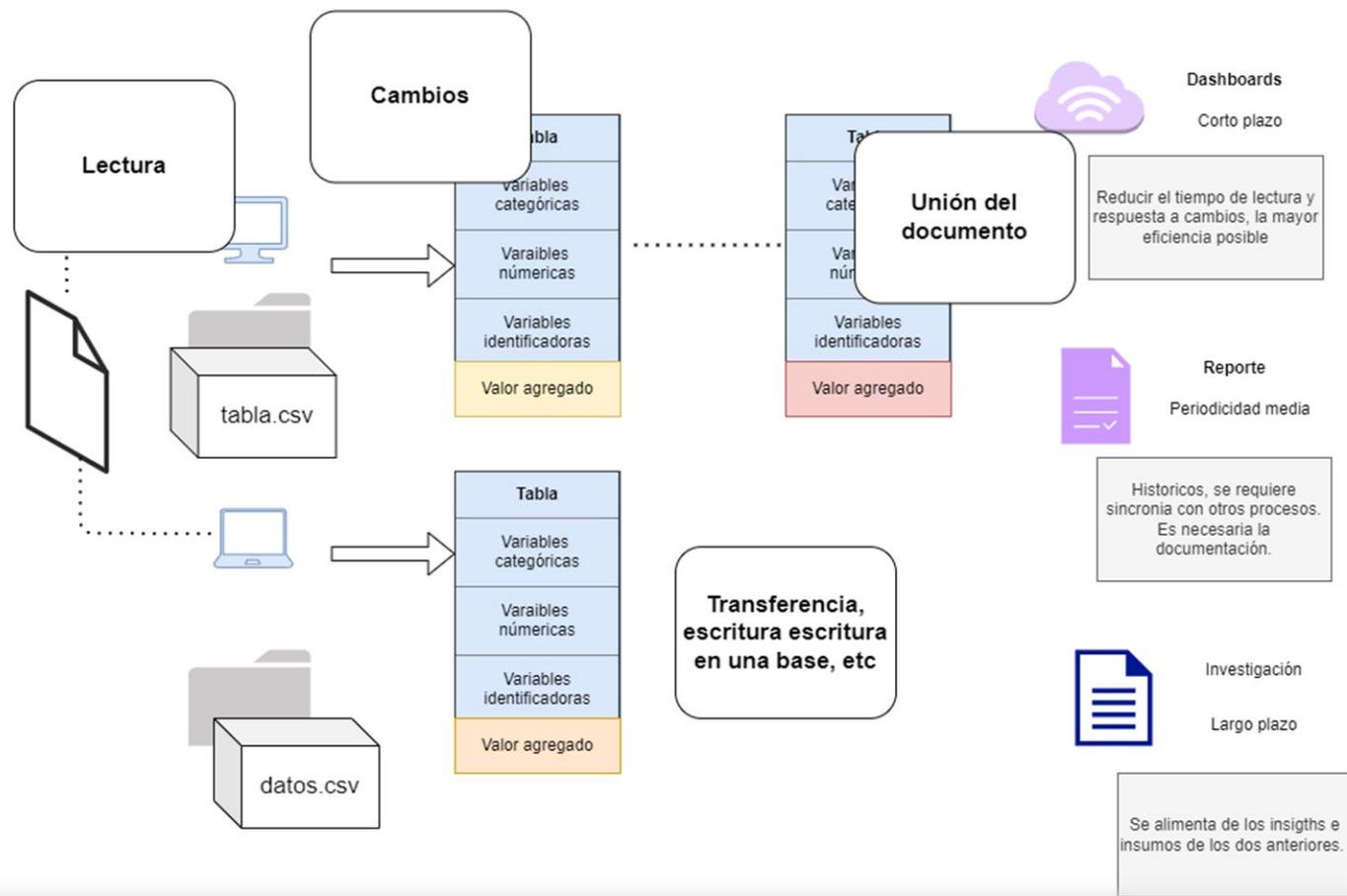


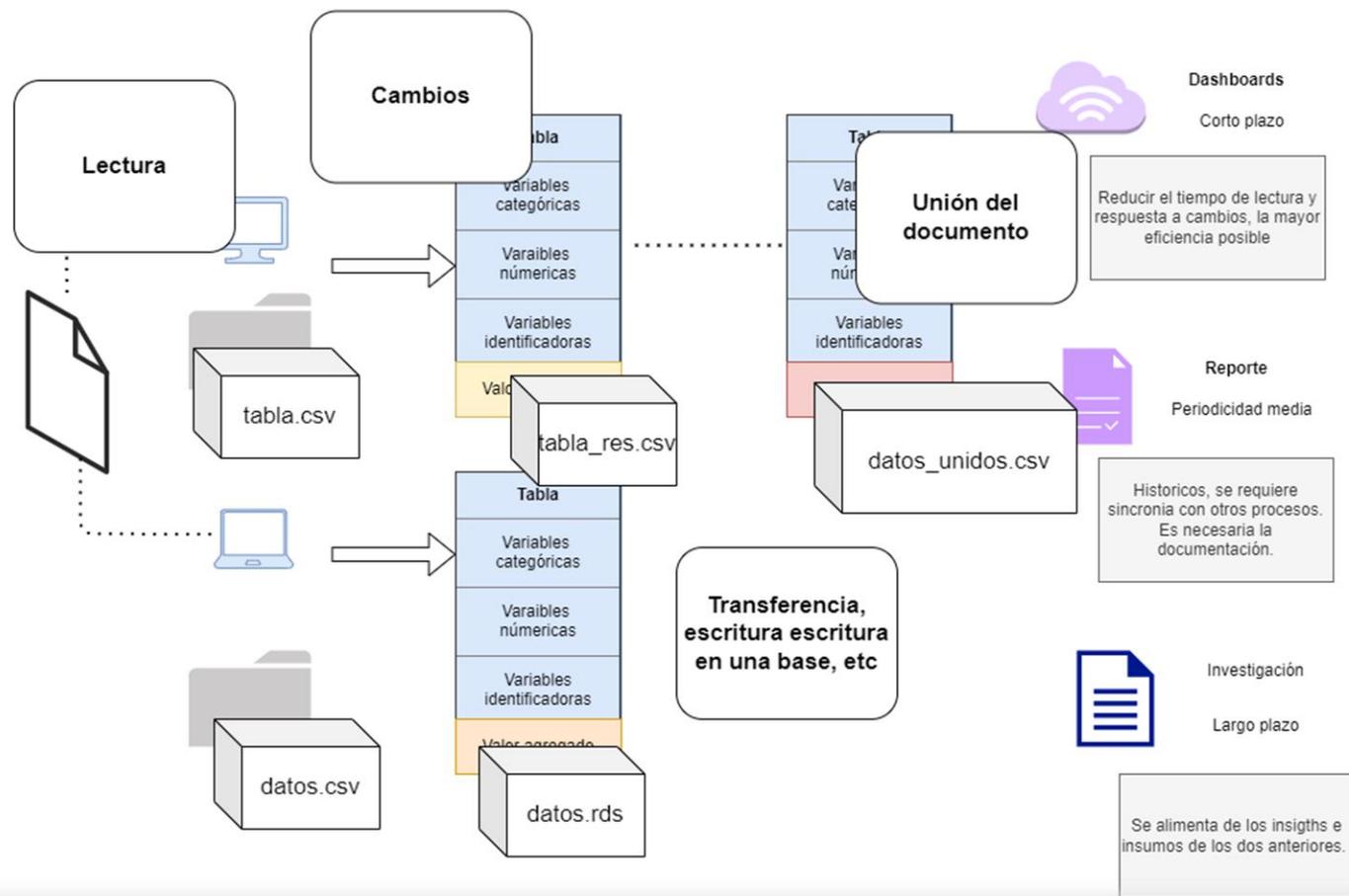


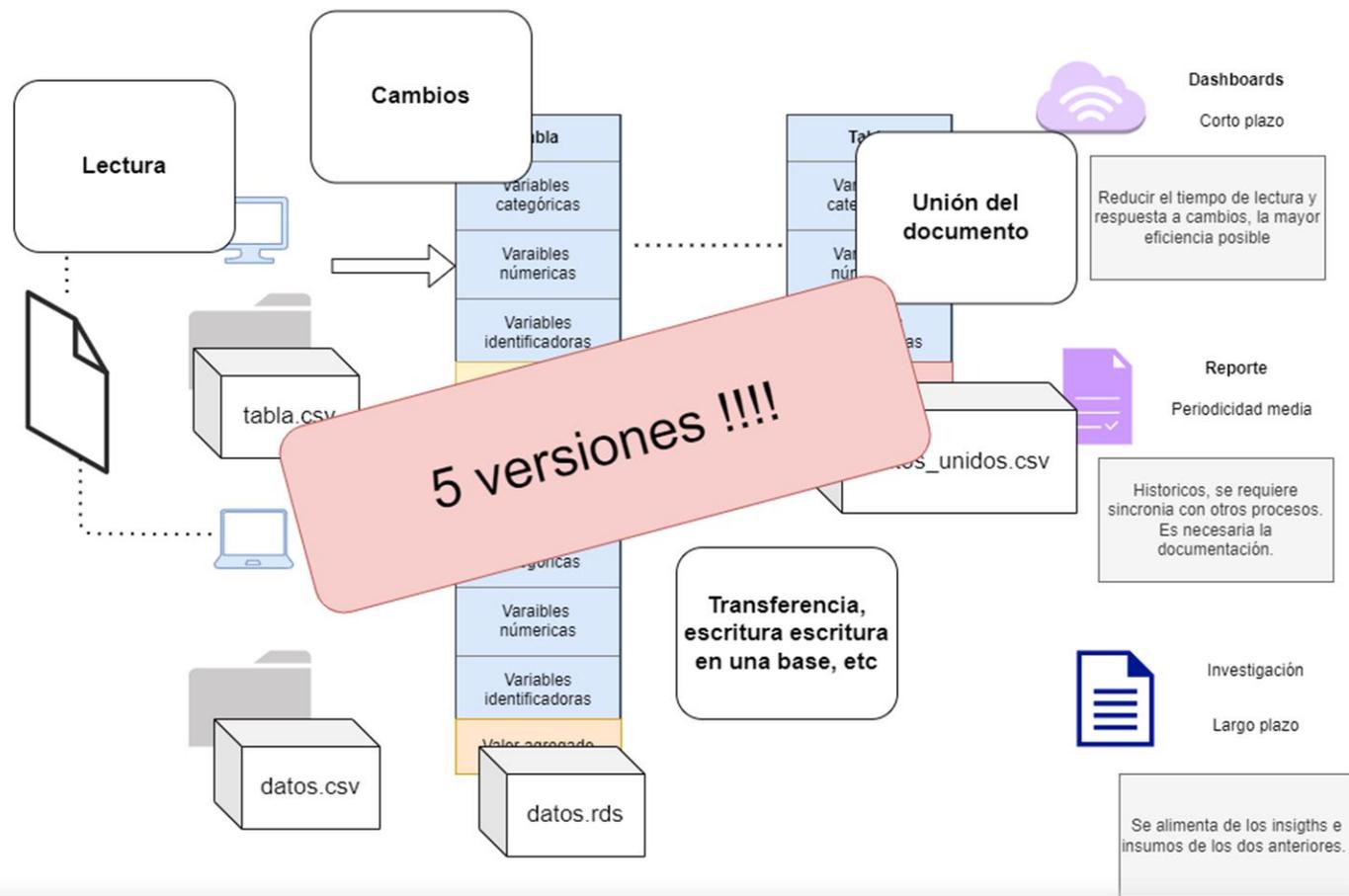


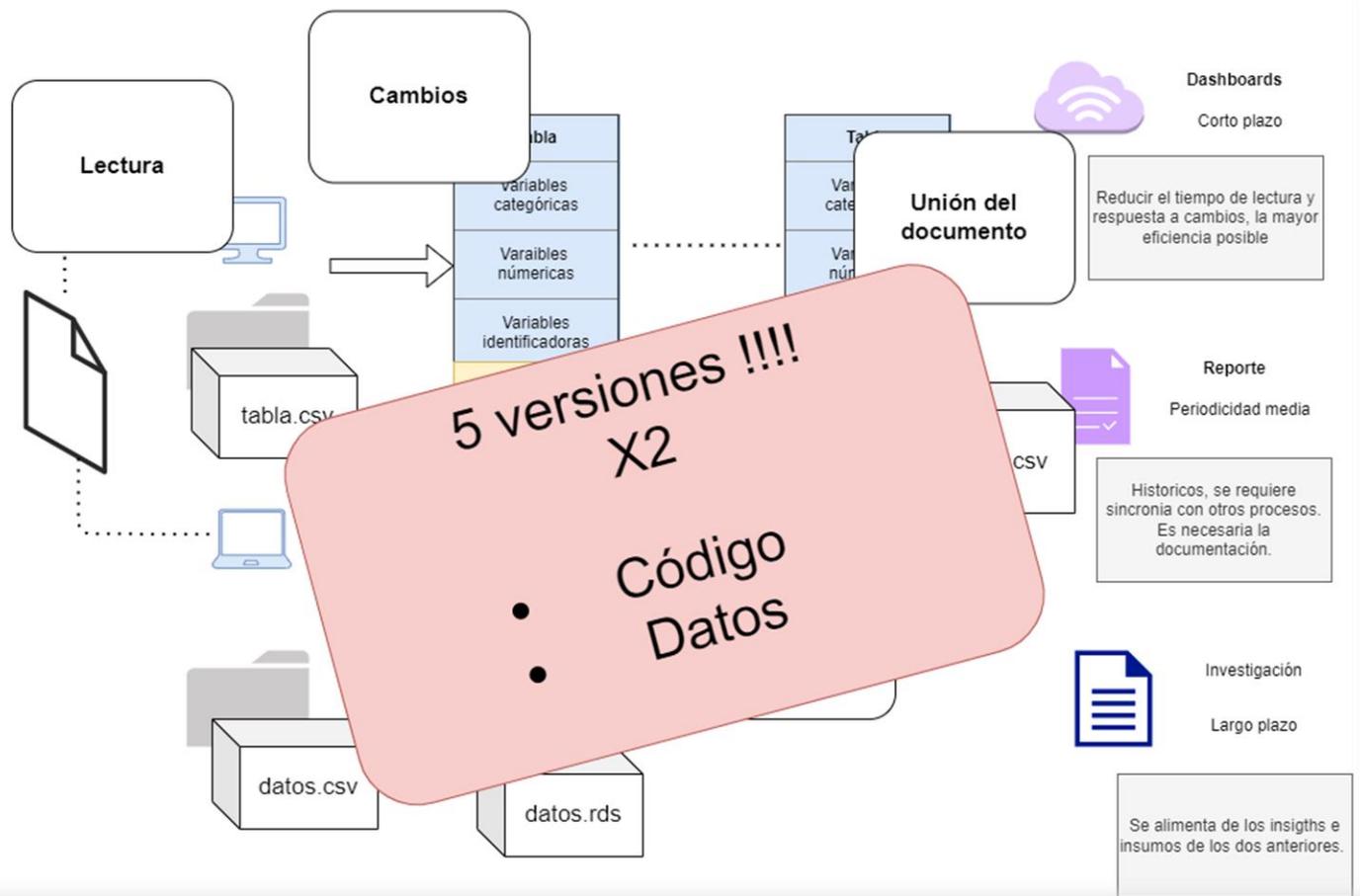














* Todos en el equipo de investigación

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES





Hora de reunir al equipo

Octubre, 2022

ORGANIZADORES

datalat.org



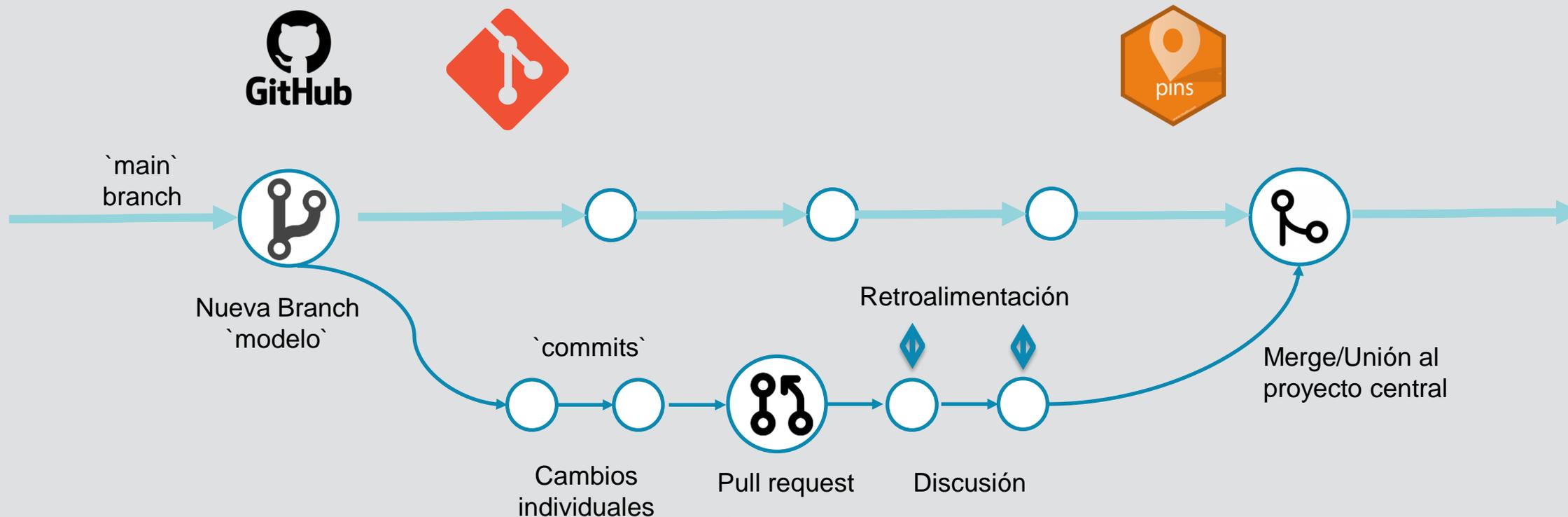
AUSPICIANTES



Dos claves importantes: *colaboración* y *reproducibilidad*

Para scripts de R, Python, Julia, CSS, Rmarkdown, Quarto, do
fiales, etc..

Para los datos y metadatos



Octubre, 2022

ORGANIZADORES

data-lat.org

REDAM
Red Ecuatoriana De
Datos Abiertos Y Metadatos

AUSPICIANTES

CS&S Code for
Science &
Society

Chan
Zuckerberg
Initiative

GORDON AND BETTY
MOORE
FOUNDATION



Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES





Manos a la obra

Octubre, 2022

ORGANIZADORES

datalat.org



AUSPICIANTES



Instalación y preparación del ambiente de trabajo

Todos nuestros esfuerzos se centran en una sola herramienta

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Software y requisitos

Para los siguientes ejercicios requerimos de las siguientes herramientas:

- R ($\geq 4.2.0$)
- RStudio
- Una cuenta de OneDrive
- GIT instalado
- Cuenta de GitHub

Pero más importante:

- Curiosidad y amor por la investigación

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

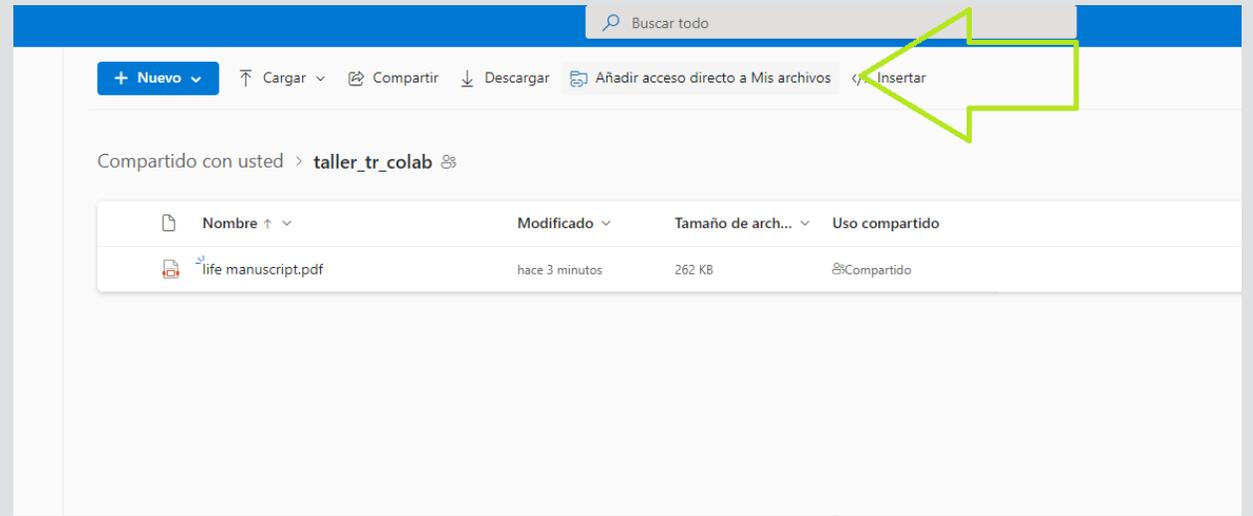
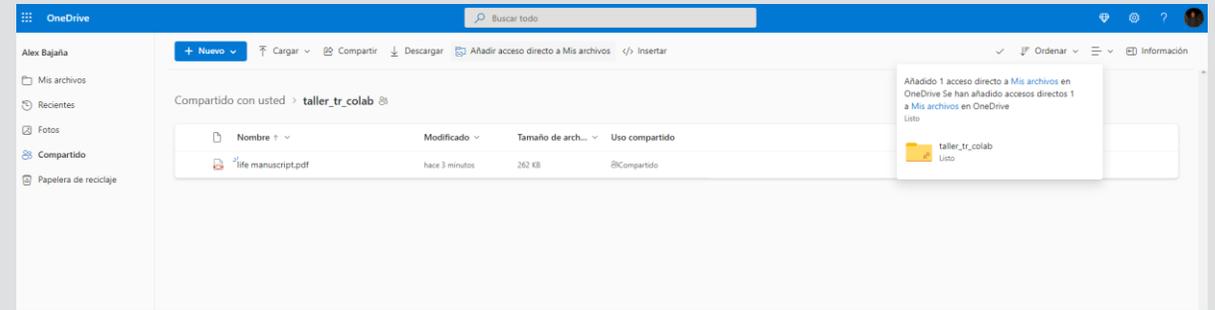
CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Configuración inicial

- La librería {pins} funciona tanto de forma local, como con la nube.
- Para este taller hemos creado la carpeta compartida [taller_tr_collab/](#) en OneDrive
- Revisa el correo con el que te inscribiste al taller, una invitación te espera.
- Abrimos OneDrive en el navegador de tu preferencia y será necesario que sincronices tu carpeta local con la carpeta remota.



Antes de ir a nuestra sesión de Rstudio vamos con unas instalaciones:

```
1 install.packages(c(  
2  
3   # La estrella de la tarde  
4   "pins",  
5  
6   # El aliado versatil  
7   "tidyverse",  
8  
9   # El pana cuadrado  
10  "gt",  
11  
12  # El artista  
13  "pathwork"  
14  
15 ))
```

Datos, metadatos y más metadatos

! ¡Todos a la sesión de Rstudio!

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES



Referencias y documentación

Material de apoyo para entender como emplear {pins} en tu equipo de trabajo

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES



Autores y mantenimiento

Silge J, Wickham H, Luraschi J (2022). *pins: Pin, Discover and Share Resources*.
<https://pins.rstudio.com/>, <https://github.com/rstudio/pins-r>.

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES

CS&S Code for Science & Society

**Chan
Zuckerberg
Initiative**

GORDON AND BETTY
MOORE
FOUNDATION

Enlaces:

- Directorio de empresas del Instituto Nacional de Estadísticas y Censos:
<https://www.ecuadorencifras.gob.ec/directoriodeempresas/>
- Bike Predict End-to-End Machine Learning Pipeline:
https://github.com/sol-eng/bike_predict
- Productionizing Shiny and Plumber with Pins
<https://www.r-bloggers.com/2019/10/productionizing-shiny-and-plumber-with-pins/>
- Librería {pins} en el CRAN
<http://cloud.r-project.org/web/packages/pins/index.html>

Agradecimientos

Una herramienta así solo se construye en colaboración

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES



- A toda la comunidad de usuarios de R, Rstudio, GitHub y demás herramientas, que día a día aportan con un grano de arena a la consolidación de la ciencia de datos responsable y ética
- A quienes conforman o conformaron el Centro de Investigación Estadística ERGOSTATS: Nathy, Esteban, Andrea, Carlos, Gustavo, Aracely, Evelyn, entre otros amigos y colegas que se han prestado para ayudarme a desarrollar este taller.
- A los equipos de trabajo con los que he compartido mi experiencia. En especial a Fausto, Juan Carlos, James, Jose, Fernando y más personas que no nombro por honor al tiempo, pero que fueron de gran ayuda y motivación.

Octubre, 2022

ORGANIZADORES

data-lat.org



AUSPICIANTES



CO-ORGANIZADORES



ALIADOS DE DIFUSIÓN



ORGANIZADORES



AUSPICIANTES



Octubre, 2022