CIMMYT
Proyecto Dataverse

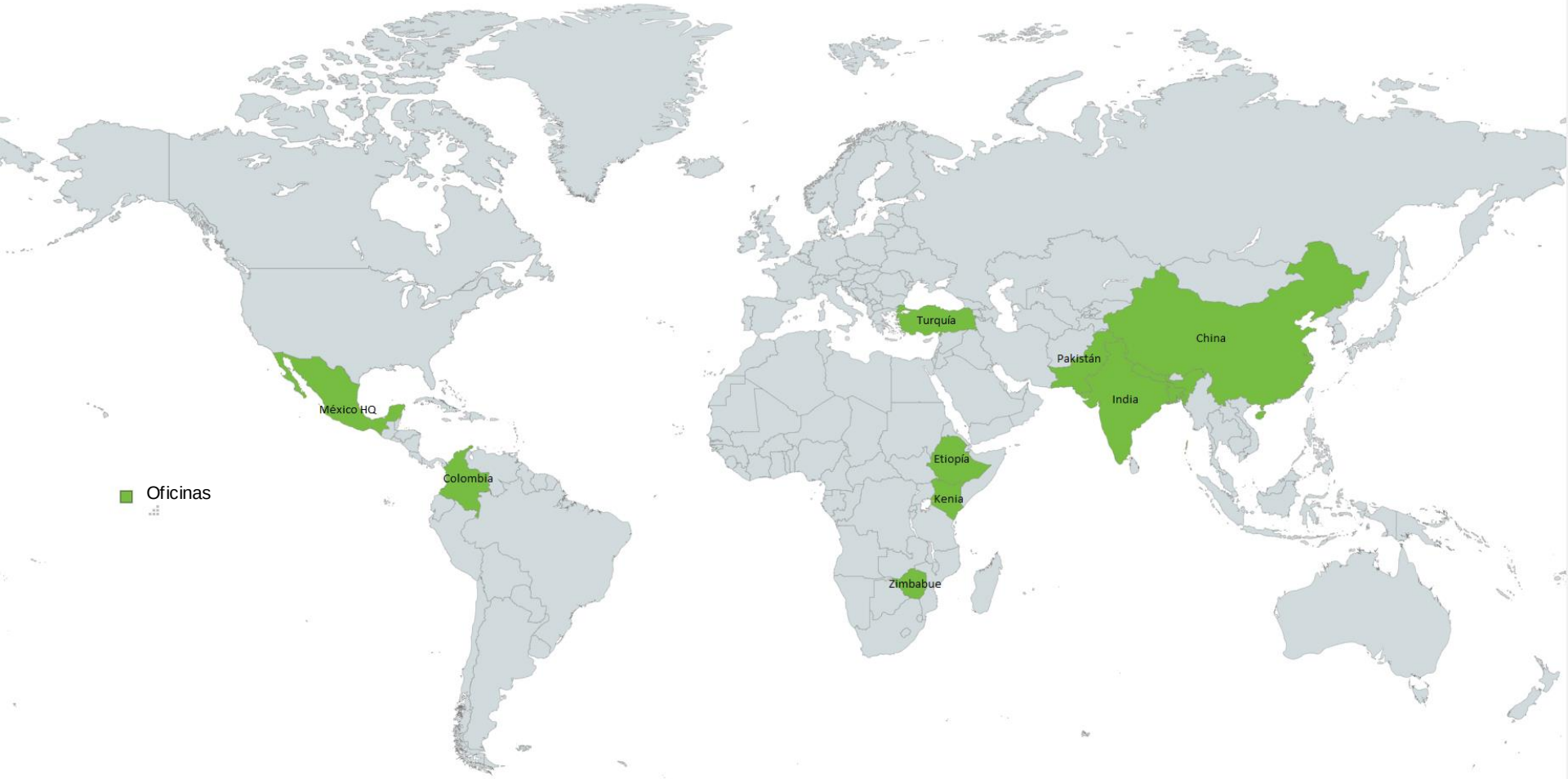
Alejandra Tenorio y Jesús Herrera

17 de abril de 2023, Buenos Aires, Argentina.

CGIAR

CIMMYT

CIMMYT en el mundo



CGIAR

CIMMYT MR

Centros de investigación del CGIAR

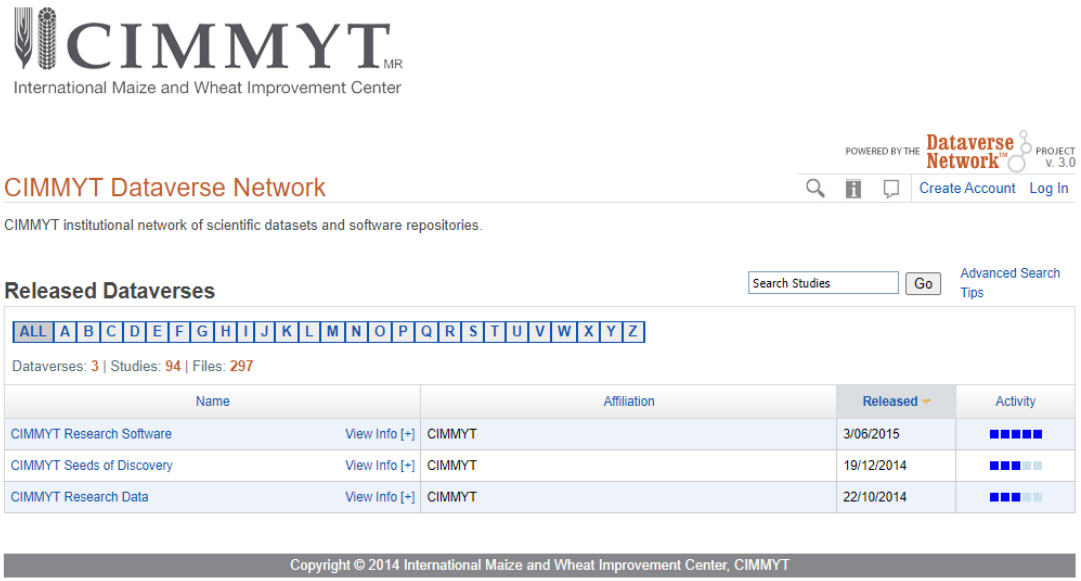
CIMMYT es uno de los miembros del CGIAR en 65 países

CGIAR

CIMMYT MR

CIMMYT Research Data & Software Repository Network

- Construir una plataforma de datos abiertos
- Inició en 2014
- Versión 3.0
- 3 subrepositorios

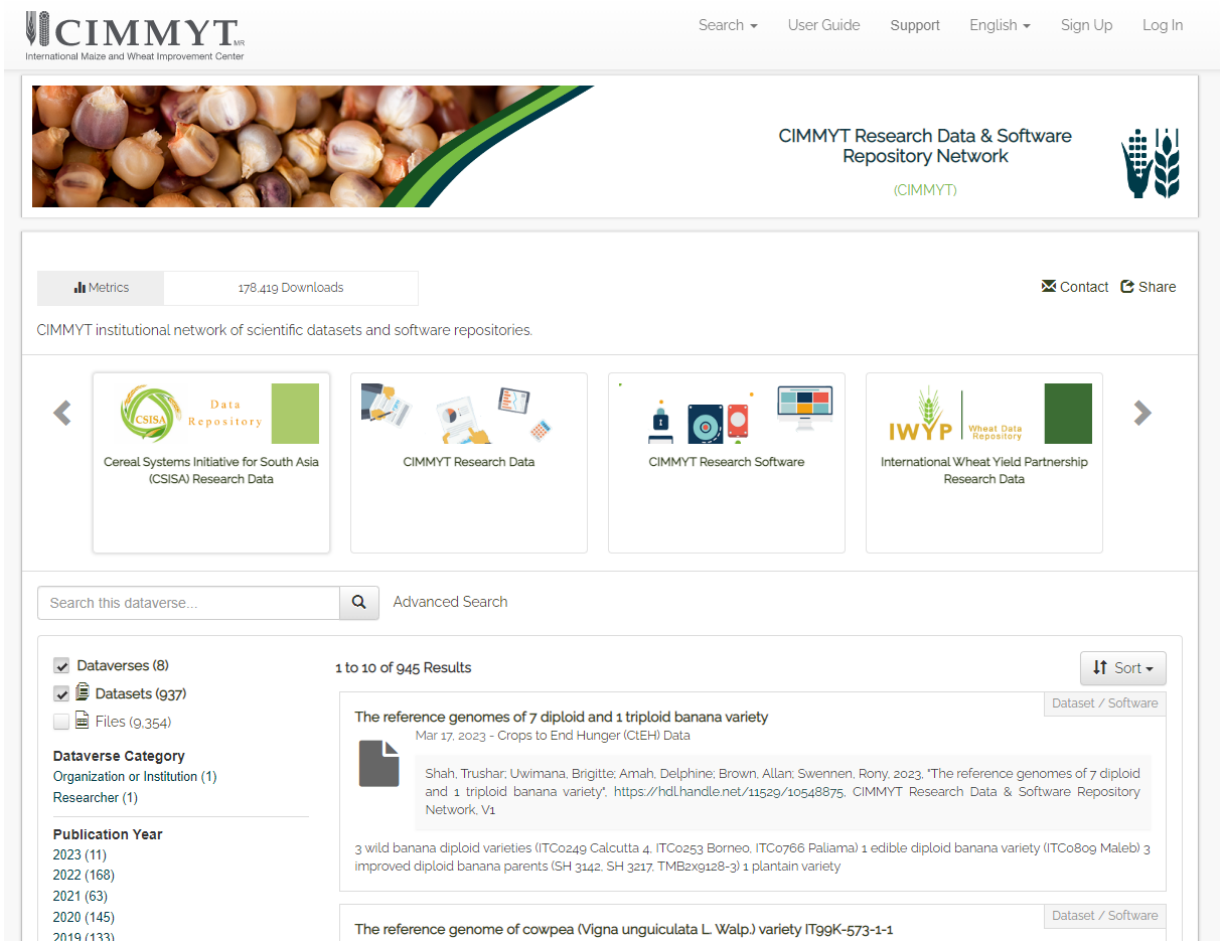


CGIAR

CIMMYT

CIMMYT Research Data & Software Repository Network

- Versión: 5.10.1



CGIAR

CIMMYT MR

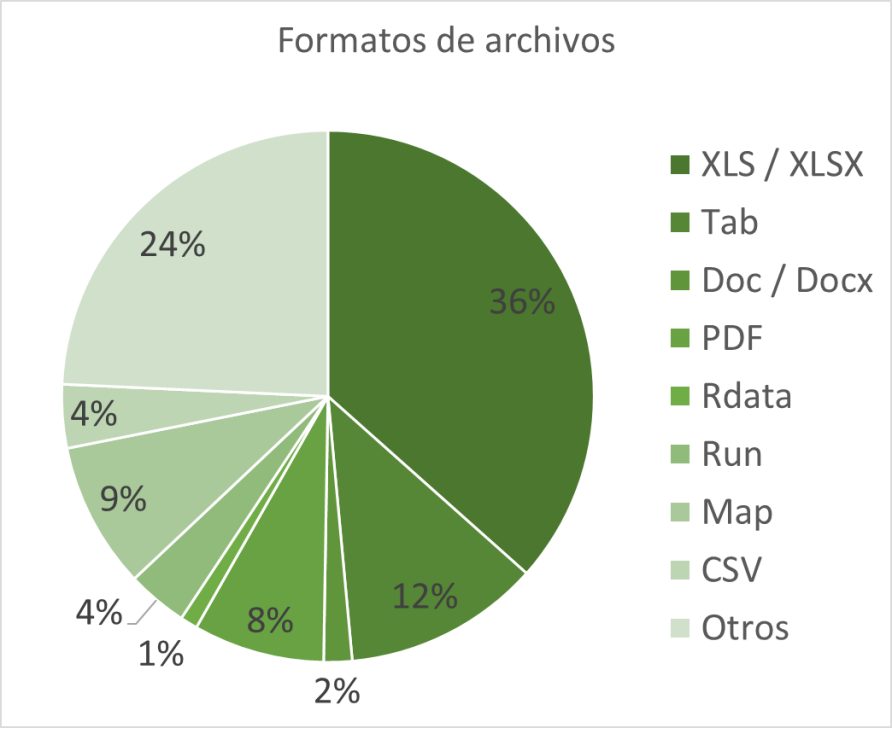
CIMMYT Research Data & Software Repository Network

Tipos de datos principales:

- Datos procesados
- Datos limpios
- Subconjuntos
- Encuestas
- Software

Almacenamiento:

- Sistema de archivos local



CGIAR

CIMMYT

Nuevos desafíos – Big Data

- Cada dataset podría llegar a ~2 TB
  - Datos de las secuencias
  - Datos crudos o borrosos

- Herramientas tradicionales como,
  - Dropbox
  - globus
  - SFTP
  - Discos Duros

Big Data

Integrar Dataverse con infraestructura tecnológica con las siguientes características:

- Escalable (almacenamiento)
- Alta disponibilidad
- Duradera
- Seguridad

CGIAR

CIMMYT MR

StorJ

- Plataforma de almacenamiento en la nube descentralizada
- Alta disponibilidad: codificación Reed Solomon
- Fragmentación de ficheros en 80 piezas
- Alta redundancia
- Alta durabilidad de los datos
- Compatible con el servicio de almacenamiento de objetos S3
- Ahorrar al menos 80% - 90% que con AWS S3

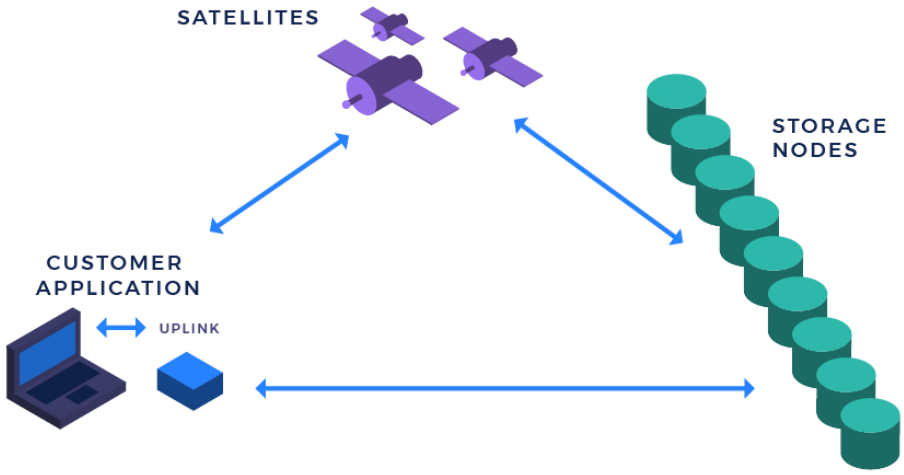
STORJ

CGIAR

CIMMYT

Integración de Dataverse con un almacenamiento S3 o compatible

- Al ser compatible con S3 se activó como almacén remoto para el repositorio de CIMMYT
- Se migró todo el almacenamiento del sistema de archivos local a StorJ

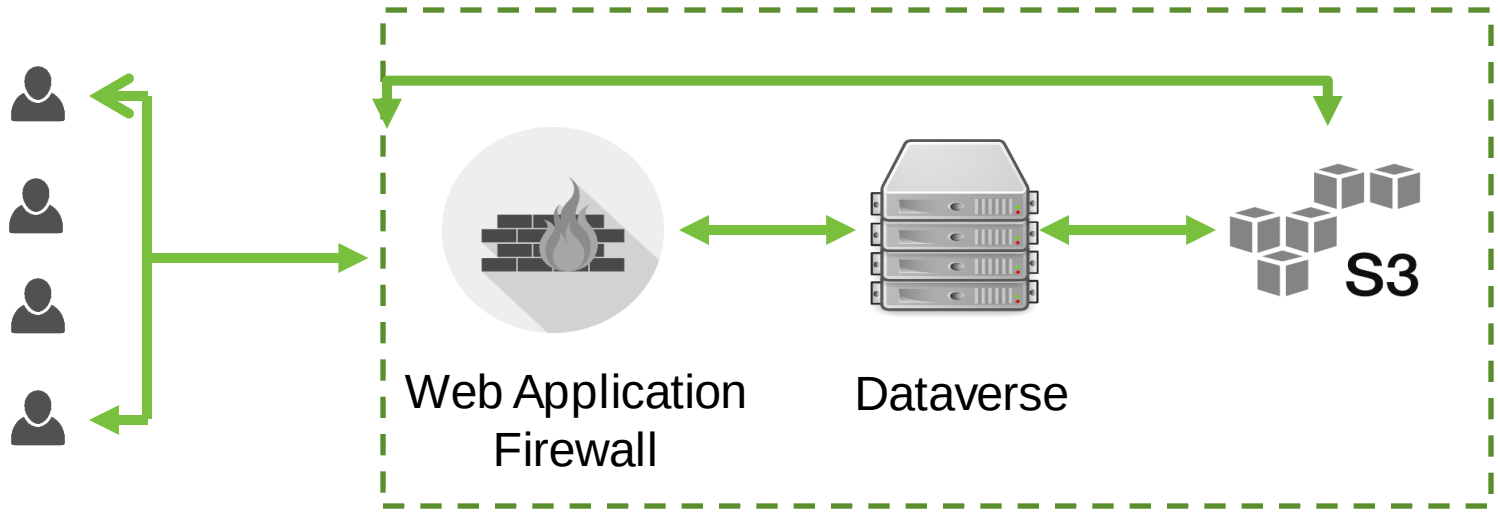


CGIAR

CIMMYT

Integración de Dataverse con StorJ

- Repositorio tiene la siguiente arquitectura

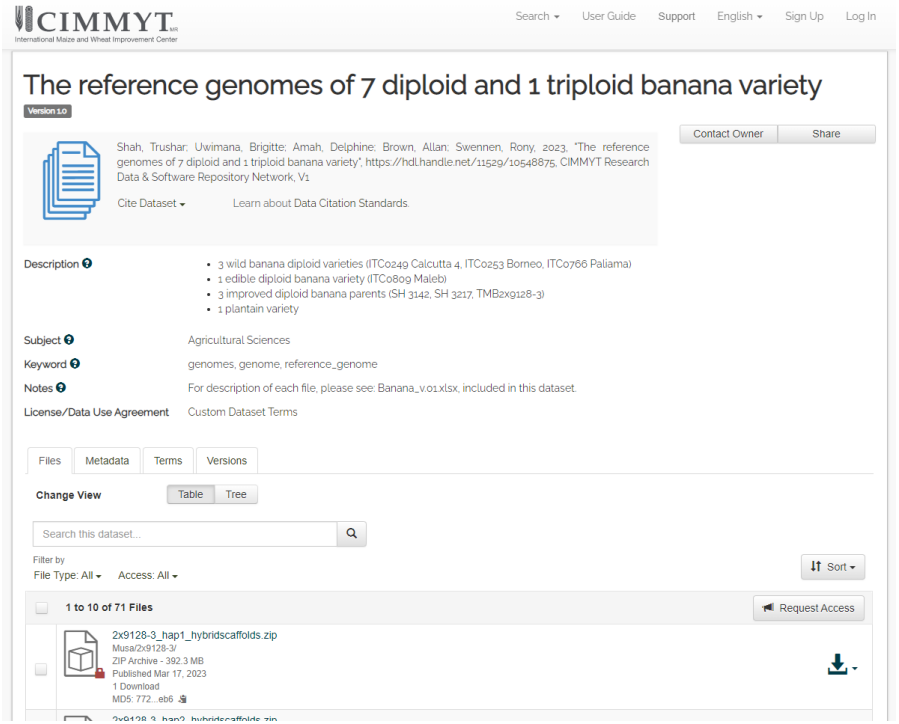


Herramientas para cargar archivos

- Desde la interfaz de usuario en el repositorio
- Directa desde DropBox
- DVUploader

Big Data

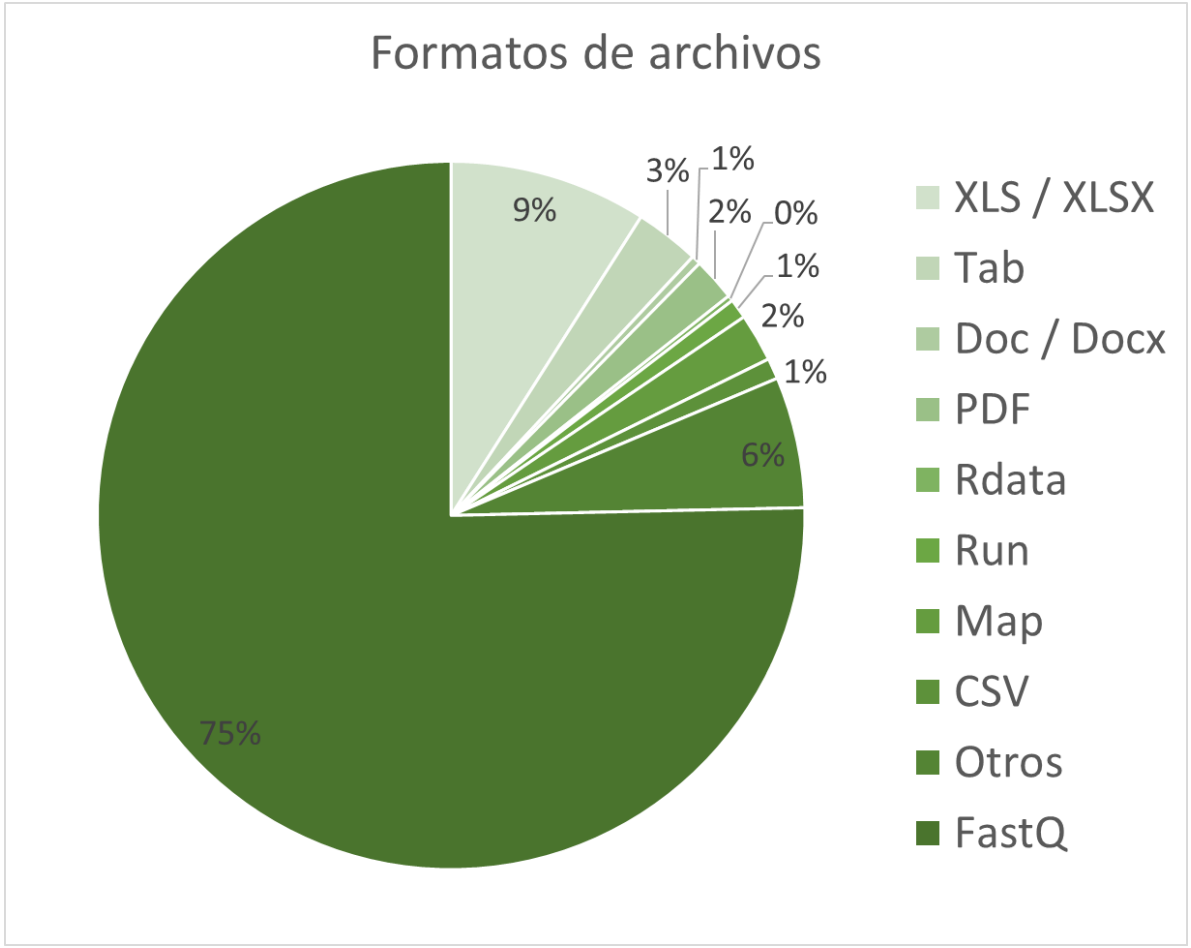
- Repositorio soporta aumentos en el volumen de datos y diversidad de formatos



CGIAR

CIMMYT MR

Tipos de archivos actuales



CGIAR

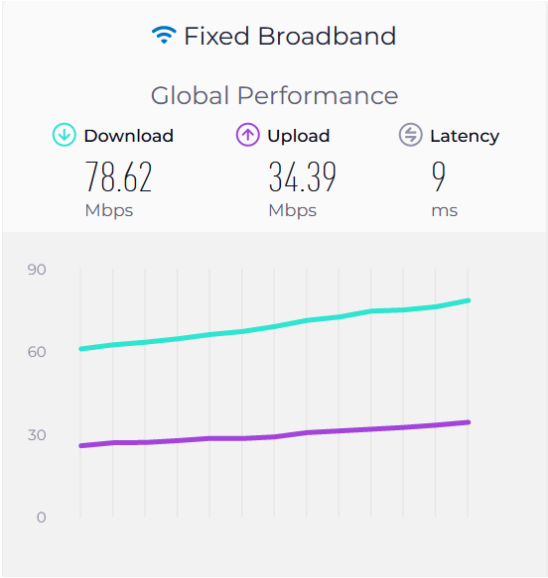
CIMMYT

Próximos pasos

Velocidad de Internet

- De acuerdo a los datos de Speedtest, países emergentes tienen una velocidad de Internet menor al promedio global

- Velocidad de descarga
  (Banda ancha fija hasta marzo de 2023)
  - Estados Unidos: 198.17 Mbps
  - **Promedio mundial: 78.62 Mbps**
  - Argentina: 54.05 Mbps
  - México: 50.15 Mbps
  - Etiopía: 6.13 Mbps



CGIAR

CIMMYT

Ámbitos de interés de CIMMYT

- Países emergentes
- Disminución de latencia
  - Red descentralizada de almacenamiento en la nube
  - Arquitectura distribuida
  - Enrutamiento inteligente

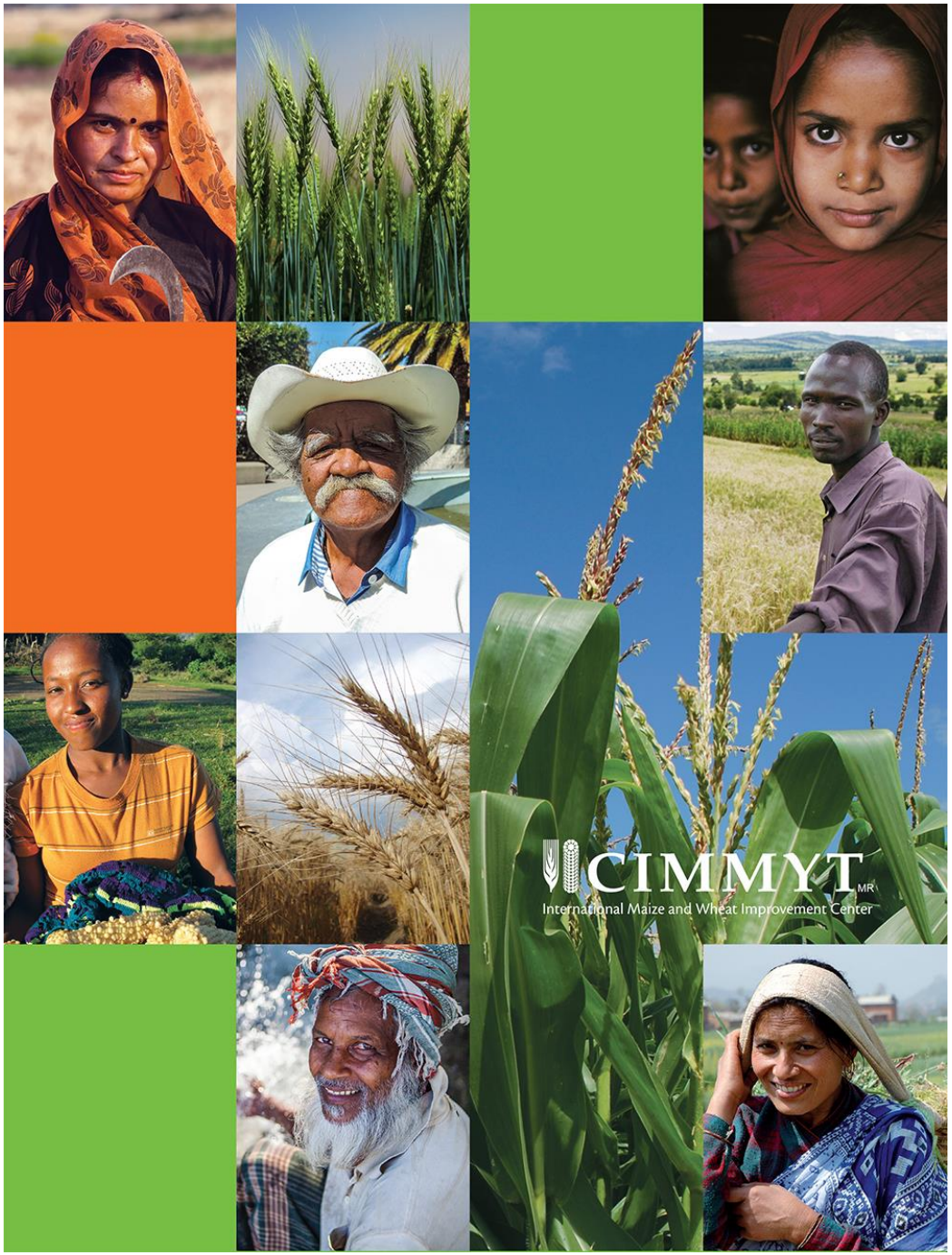
Adiciones a Dataverse

- Herramientas para mejorar el soporte a BigData
  - DVWebloader: investigadores tengan la facilidad subir archivos desde una carpeta local y subcarpetas
  - Rsync
  - Futuras herramientas que se adicionen

Contacto

- Alejandra Tenorio Robles
  a.tenorio@cgiar.org

- Jesús Herrera de la Cruz
  j.herrera@cgiar.org



Gracias