

Consortia Papyrus – Repositorio de Datos de Investigación en Colombia, una iniciativa de trabajo colaborativo

CONFERENCIA MAGISTRAL

6 DE OCTUBRE DE 2022

CÉSAR AUGUSTO RENDÓN VALENCIA

Consortia (Colombia)

RESUMEN EXTENDIDO

El experto presenta el proyecto Consortia Papyrus a través de una serie de preguntas, a partir de las cuales se plantean las respuestas.

¿Qué es un repositorio de datos?

Un repositorio de datos es una herramienta a través de la cual académicos e investigadores almacenan y realizan una amplia descripción de los datos producto de su ejercicio investigativo, con el fin de divulgarlos, gestionarlos y facilitar el acceso y uso por parte de la comunidad científica y el público en general.

¿Por qué una institución debe gestionar los datos de investigación?

Existen varias razones por las cuales realizar esta gestión, una de ellas es la exigencia de las revistas, hoy en día es obligatorio llegar a acuerdos sobre en dónde estarán publicados los datos que sustentan los estudios, ya no basta con que estén en los propios computadores de las personas docentes o investigadoras, sino que tienen que estar visibles; además para una institución de educación superior es muy relevante la posibilidad de tener visibilidad.

Del mismo modo, las agencias financiadoras están exigiendo que estos datos sean públicos y que los datos se puedan consultar para poder hacer desarrollos futuros. Y así mismo, las propias instituciones y las vicerrectorías de investigación están solicitando que se realice una gestión centralizada de los datos que se preserven en el tiempo, además de facilitar la recuperación y la reutilización de los mismos aporta transparencia a los procesos de investigación que estamos realizando.

¿Cuál es la principal motivación tras la creación de Consortia Papyrus – Repositorio de Datos de Investigación?

En primer lugar, está la necesidad de fortalecer el trabajo colaborativo, además de impulsar la ciencia abierta en el país y, de esta forma, visibilizar la producción nacional en escenarios internacionales.

Por otra parte, está la tarea de disminuir costos en todos los aspectos técnicos, considerando el contexto latinoamericano. En términos de un proyecto tecnológico cuando estamos hablando de desarrollos a la medida o de instalaciones, servidores, espacios en disco, etc., se hace evidente la necesidad de colaborar, y finalmente aprovechar infraestructura tecnológica y recurso humano calificado.

¿Qué es Consortia?

Se trata del operador del consorcio de Colombia que reúne 61 instituciones entre públicas y privadas, y en el marco de este proyecto de consorcio, se han generado un sinnúmero de actividades y de subproyectos. Aunado a esto, el consorcio ha apoyado el desarrollo de ciencia abierta y realizado seminarios, empezando por el seminario de acceso abierto.

El siguiente punto tiene que ver con la arquitectura, que es el modelo de funcionamiento de Consortia. Así es como se creó un diseño que tiene dos tipos de espacio, uno es el espacio de gestión conjunta y el otro un espacio de

recolección. Este nivel de arquitectura permite que, cuando una institución no tiene aún desarrollado su repositorio de datos y no está trabajando aún con Dataverse, de igual forma tiene la posibilidad de participar en el espacio de gestión conjunta en donde se le creó su propio Dataverse para poder desarrollarlo con las características requiera.

De esta forma, se comparte servidor, espacio en disco y una serie de elementos que se unifican en el repositorio. Pero también se respeta el espacio de recolección de aquellas instituciones que tienen desarrollos propios que ya tenían en el pasado instalado un Dataverse en funcionamiento, pero que querían participar en el consorcio. Para esto se creó un espacio de recolección en donde se recolectan los datasets que ya tienen a disposición y así pueden participar del proyecto conjunto.

A continuación, se describen algunas características propias de Consortia:

Definiciones institucionales: estas definiciones se toman en función de una serie de variables, por ejemplo, la distribución de los Dataverse. Es decir, es muy importante conservar la estructura que tiene cada institución en términos de la investigación, entonces cada institución puede decidir dentro de su Dataverse propio cómo se va a estructurar.

Tipos de datos: actualmente, la plataforma permite datos de tipo textual, multimedia, imágenes, tabulares, estructurados, código fuente, específico de programas y archivos netCDF.

Licencias de uso:

Abierto: contenido visible por todo usuario de internet. Su uso puede estar sujeto a distintas limitaciones según la licencia Creative Commons con la que se registre.

Cerrado: contenido visible sólo para el personal encargado de la gestión del repositorio institucional.

Restringido: contenido en principio cerrado, pero que puede ser puesto a disposición del público bajo ciertas condiciones y siempre bajo petición.

Embargado: contenido cerrado que pasa automáticamente a ser abierto, una vez transcurrido un período de tiempo determinado.

DOI en Papyrus (repositorios de datos): a cada organización miembro del consorcio se le asigna un prefijo diferente, que permitirá identificar a qué institución pertenece cada DOI asignado. Los DOI pueden ser asignados a sets de datos en el repositorio de datos de investigación o a otros resultados de investigación que estén almacenados en el repositorio institucional como tesis, trabajos de grado, preprints, entre otros. Cada organización tendrá hasta 10.000 DOI para asignar.

¿Cuáles han sido los alcances de Consortia Papyrus?

El proyecto, que se inició en enero 2022, desde su comienzo se dio a la tarea de adquirir un servidor, un dominio, y elaborar una interfaz gráfica para la instalación de Dataverse en el servidor, así como la creación de los espacios para cada institución miembro.

Hasta ese momento la Universidad del Rosario participó en el proceso de integración de las otras 12 instituciones, actualmente participan 13 instituciones que están en el espacio de gestión conjunta, cada una con su propio espacio. Por otra parte, se han realizado distintos Webinar en temas como: capacitación técnica, plan de gestión de datos, definición de metadata, variables transversales, y la gestión de datos en la investigación.

[Ver conferencia en YouTube](#)