

Manejo de Datos Abiertos en la Agricultura y Nutrición

Este curso de aprendizaje digital (e-learning) es el resultado de una colaboración entre socios de GODAN Action, incluyendo a **Investigaciones Ambientales Wageininen (WUR)**, **AgroKnow**, **AidData**, la **Organización de las Naciones Unidas para la Alimentación y la Agricultura** (FAO por sus siglas en Ingles), **El Foro Global sobre Investigaciones de Agricultura** (GFAR), y el **Instituto de los Estudios del Desarrollo** (IDS), **The Land Portal**, **el Instituto de Datos Abiertos** (IDI) y el **Centro Técnico de Agricultura y cooperación Rural** (CTA).

GODAN Action es un proyecto de tres años [por] el Departamento del Desarrollo Internacional del Reino Unido para capacitar a los que usan, producen, e intermediarios de datos para conectarse efectivamente con datos abiertos y maximizar la potencial por su impacto en los sectores de agricultura y nutrición. En particular, trabajamos para mejorar la capacitación, promover estándares comunes y mejores prácticas para medir el impacto. [www.godan.info]

Este trabajo está registrado con una licencia **CC BY-SA**



Unidad 3: Creando datos abiertos

Lección 3.3: Creando y Gestionando a repositorios de datos abiertos



Foto por el Centro Alemán de Computación Climática licenciado bajo CC BY 2.0

Objetivos y metas de aprendizaje

Se puede utilizar a datos abiertos cuando

Esta lección tiene el objetivo de:

- Proveer una vista general de cómo crear un repositorio de datos abiertos y explicar los requisitos para la gestión de ello.
- Describe la importancia y los diferentes componentes de la curación de datos

Después de estudiar esta lección debes saber cómo:

- Identificar los pasos a considerar en crear un repositorio de datos abiertos
- Seleccionar el software adecuado para considerar en crear un repositorio según sus requisitos
- Entender los requisitos para gestionar de manera eficiente a un repositorio de datos abiertos
- Entender a los componentes diferentes en la curación de datos.

Contenido

Unidad 3: Creando datos abiertos

Lección 3.3: Creando y Gestionando a repositorios de datos abiertos

Objetivos y metas de aprendizaje

Lista de tablas

1. Introducción
2. Creando un repositorio de datos abiertos
 - 2.1. Estableciendo a las necesidades y requisitos de usuarios
 - 2.2. Eligiendo a software de repositorio
 - 2.2.1. Software de código abierto
 - 2.2.2. Alojamiento comercial
 - 2.2.3. Alojamiento gratis
3. La gestión de un repositorio de datos abiertos
 - 3.1. Interoperabilidad y el intercambio de datos
 - 3.1.1. Publicando a Metadatos
 - 3.2. Consideraciones para la gobernanza
 - 3.3. Curación y Preservación

Resumen

Lista de gráficas

Gráfica 1 El Modelo de Curación de Ciclo de Vida DCC

Lista de tablas

Tabla 1 Una vista general de repositorios de datos

Tabla 2 Una vista general de requisitos

Tabla 3 Los metadatos que los usuarios consideran útil para entender a los datos

Tabla 4 Acciones dentro del Modelo de Curación de Ciclo de Vida DCC

1. Introducción

En la lección 3.2, la Gestión de Conjuntos de Datos Dinámicos, miramos a las varias opciones de publicar datos dinámicos y describimos procesos automáticos para facilitar esa publicación. En esta lección exploraremos las varias opciones disponibles para el alojamiento de este repositorio de datos abiertos. Estos pueden ser o los que están alojados comercialmente o con soluciones gratis o de código abierto

2. Creando un repositorio de datos abiertos

Por lo general, un repositorio de datos sirve como una pieza central para un esfuerzo de datos abiertos y sirve como un lugar central para encontrar a los datos, un área donde establecer prácticas estándares, y algo para mostrar el uso de esos datos. En un sentido práctico, un repositorio sirve como un lugar central donde alguien puede buscar a datos. Algunos softwares de repositorios automáticamente convertirán a datos de un formato a otro, para que, aunque tú puedas proveer a los datos solamente en un formato (como CSV), puedan ser generados en otros formatos como XML, JSON, Excel, etc.

Algunos softwares crean visualizaciones en el buscador de web, permitiendo que la gente crea, ordene, consulta, y combina conjuntos de datos, sin tener que saber cómo hacer programación.

Algunos softwares de repositorio ofrecen la sindicación, que permite que otras organizaciones automáticamente incorporen a sus datos (por ejemplo, una agencia pública de la transportación puede recopilar todos los datos de transportación local y re-publicarlos).

Por lo general, con los softwares de repositorio puede o subir archivos para estar almacenados allí, o conectando al repositorio a un sitio web donde existe ese archivo. La primera opción funciona para organizaciones más pequeñas (localidades, agencias pequeñas), mientras que la segunda funciona para organizaciones grandes o gobiernos, donde la centralización de productos no sería práctica.

Para una vista general, hay tres tipos de alojamiento disponibles: el alojamiento comercial, el auto-alojamiento, y alojamiento gratis. La Tabla 1 abajo enumera a los repositorios de datos disponibles.

Tabla 1: Una vista general de repositorios de datos

Nombre	Tipo	Notas
ArcGIS Open Data	Alojado	
CKAN	Código abierto	
DataHub	Gratis, alojado	Impulsado por CKAN
DKAN	Código abierto	
GitHub	Gratis, alojado	
JKAN	Código abierto	
Junar	Alojado	

NuData	Alojado	
OpenData.city	Gratis, alojado	Impulsado por CKAN
OpenDataSoft	Alojado	
Open Data Catalog	Código abierto	
Socrata	Alojado	

2.1. Estableciendo a las necesidades y requisitos de usuarios

Los estudios que investigaron las causas de fallas en proyectos han hallado que la “definición de requisitos” es la etapa más frecuentemente citada de donde falla un proyecto. Por lo tanto, es imprescindible entender que debe hacer un repositorio y asegurar que cumple con los requisitos de las partes interesadas para tener éxito en el proyecto.

Identificar a las necesidades

- Entender lo más posible sobre los usuarios, y también su trabajo y contexto de trabajo.
- El sistema bajo desarrollo debe apoyar a los usuarios en lograr sus metas
- Identificar a las necesidades es esencial para la próxima etapa

Establecer a las necesidades

- Edificando en las necesidades identificadas, produce un conjunto de requisitos
- Utiliza un enfoque al desarrollo centrado en el usuario
- El entender lo que debe hacer el producto y asegurar que cumple con las necesidades de los partes interesados es absolutamente necesario para el éxito del producto.

¿Qué son requisitos?

- Un requisito es una declaración que especifica lo que debe hacer un producto, propone hacer, o como debe ser su desempeño. Tradicionalmente, hay dos tipos de requisitos:
- Requisitos funcionales especifican que debe hacer el sistema
- Requisitos no-funcionales especifican los límites en el sistema o su desarrollo

La Tabla 2 provee una lista expandida de las categorías de requisitos.

Tabla 2: Una vista general de requisitos

Categorías de los Requisitos	Descripción
Requisitos funcionales	Que debe hacer el producto
Requisitos de datos	El tipo, volatilidad, tamaño/cantidad, persistencia, precisión, y valor de las cantidades de los datos requisitos
Requisitos del ambiente	O “contexto de uso” -circunstancias en donde el producto interactivo debe operar
Requisitos del usuario	Características del grupo de usuarios sujeto
Requisitos de utilidad	Las metas de utilidad y sus medidas asociadas

Basado en las necesidades y servicios del repositorio, las instituciones querrán evaluar a las plataformas de software disponible. Hay tres tipos de opciones disponibles:

- **Software de código abierto:** Este software es gratis para descargar, pero normalmente requiere un nivel de experiencia para implementar y mantener. Un cuerpo central de gobierno manejó al código bruto, pero está abierto para cambios o mejoras que provienen de la comunidad de desarrollo (por ejemplo, CKAN, DKAN, JKAN).
- **Software Comercial:** típicamente pagan para el software, y opciones para suscripciones adicionales o para consultorios. Eres el dueño del uso del software, y con una suscripción, puedes aumentar de nivel de servicio en el software. Con una interfaz de programación, o API, puedes personalizar al software, pero el vendedor del software es el dueño de, crea, y mantiene el código fuente.
- **Modelo de Software como Servicio:** un vendedor es el dueño y distribuye una plataforma de servicio, o también aloja y mantiene a los datos por ti. En este modelo el vendedor de software provee servicios adicionales por un costo, y también controla y actualiza al código fuente del software (por ejemplo, *EPrints Services*, *Open Repository and bepress*).

Los que implementan un repositorio querrán escoger un software que mejor cumple con las necesidades y recursos disponibles (presupuesto y personal). Por ejemplo, instituciones sin experiencia técnica significativa tal vez preferían considerar a los servicios comerciales disponibles. En términos de plataformas de software con código abierto, cada uno tiene sus propias ventajas.

2.2. Eligiendo a software de repositorio

2.2.1. *Software de código abierto*

Hay algunos programas de código abierto excelentes que son buenas opciones para organizaciones con experiencia técnica, con organizaciones comprometidas a usar software de código abierto, o para organizaciones con el presupuesto para contratar a un consultor para implementar el software.

CKAN: CKAN¹ es un acrónimo para “Comprehensive Knowledge Archive Network (*Red de Archivos de Conocimiento Comprensivo*),” pero casi siempre se conoce simplemente como CKAN. Una creación por Open Knowledge (Conocimiento Abierto)², basado en el Reino Unido, es el repositorio de datos abiertos de código abierto más usado. El código está escrito con Python, y es el estándar para software de repositorio. Lamentablemente, también se conoce por ser difícil de instalar, aunque las imágenes *Docker*³ lo han simplificado muchísimo.

Los usuarios de CKAN incluyen a Data.gov⁴, y la Administración Nacional de Océanos y la Atmósfera (NOAA)⁵, entre muchos más. Consultores de CKAN incluyen a Open Knowledge, Ontodia⁶, y Accela⁷, en conjunto a muchos consultores independientes. Open Knowledge y Ontodia están entre algunas organizaciones que cobran por el alojamiento. Por favor mire a un sitio de demostración de CKAN⁸.

DKAN: DKAN⁹ es un clon de CKAN, aunque no comparte códigos con CKAN – se ha escrito en código de HP, como un módulo de *Drupal*. Para una organización que usa al sistema de gestión de contenido Drupal y también quiere un repositorio de datos, DKAN es una opción especialmente buena. Los usuarios de DKAN incluyen al Departamento de Agricultura¹⁰ de los Estados Unidos, entre otros¹¹. Por favor mire a un sitio de demostración DKAN¹².

JKAN: JKAN¹³ es basado en CKAN por nombre, aunque no comparte código con CKAN. JKAN fue creado por Tim Wisniewski¹⁴, El Jefe Oficial de Datos de la ciudad de Filadelfia, como un catálogo de datos impulsado por Jekyll¹⁵. Nota que JKAN es un *catálogo* de datos, no un repositorio, que quiere decir que almacena enlaces de datos y metadatos sobre los datos, pero no son los datos en sí. Los datos pueden ser alojados en un servidor FTP en vez de estar en sitios web de agencias, en Amazon S3, Dropbox, o cualquier lugar donde alguien almacenaría datos para acceso público. Instalando toma solamente un par de minutos. Por favor mire un sitio de demostración de JKAN¹⁶.

2.2.2. Alojamiento comercial

Para algunas organizaciones, el alojamiento comercial es una opción viable. Pagar para que alguien aloje a sus datos requiere poco o ningún conocimiento técnico de parte de su organización, y la compañía que aloja te guiará por el proceso. Su organización no tendrá que proveer infraestructura técnica (por

¹ <https://ckan.org>

² <https://okfn.org>

³ <https://hub.docker.com/u/ckan/>

⁴ <https://www.data.gov>

⁵ <https://data.noaa.gov/dataset>

⁶ <https://opengov.com/open-data>

⁷ <https://www.accela.com>

⁸ https://demo.ckan.org/pt_BR/

⁹ <http://www.nucivic.com/dkan/>

¹⁰ <https://data.nal.usda.gov>

¹¹ <https://github.com/NuCivic/dkan-sites>

¹² <http://demo.getdkan.com>

¹³ <https://jkan.io>

¹⁴ <https://usopendata.org/2016/03/28/jkan/>

¹⁵ <https://jekyllrb.com>

¹⁶ <https://demo.jkan.io>

ejemplo, servidores) o saber cómo hacer programación. No obstante, es importante que consideres a los acuerdos al nivel de servicio.

ArcGIS Open Data: *ArcGIS Open Data*¹⁷ es un nuevo entrante en la industria, habiendo empezado en la última parte del año 2014. *ArcGIS Open Data* está incluido con un contrato con *ArcGIS Online* – por la universalidad de ese servicio entre municipalidades y estados; es en efecto gratis para esos clientes existentes. Esto lo hace una opción muy atractiva para los gobiernos con bajos niveles de compromiso para un programa de datos abiertos, porque elimina el costo de un catálogo de datos. *ArcGIS Open Data* esta solamente disponible como un software alojado – no es posible correr una versión del software en sus propios servidores.

Junar: *Junar*¹⁸ provee plataformas y paquetes para negocios, gobiernos, ONG, y la academia con un enfoque en la colección de datos y análisis. Junar es bilingüe, con apoyo para audiencias en inglés y español. Sus precios son enfocados en pequeñas y medianas empresas, empezando en alrededor de \$10,000. El sitio de demostración de Junar está disponible bajo pedido.

NuCivic Data: *NuCivic Data*¹⁹ es basado en DKA, que fue creado y es mantenido por nūcivic. Son proveedores de medio tamaño; en términos de precio – sus precios son mucho más bajos que Socrata, pero más caro que, por ejemplo, Junar.

CivicDashboards: La empresa de consejos en datos abiertos Ontodia provee alojamiento por CKAN bajo el nombre *CivicDashboards*²⁰. Ofrecen un nivel de servicios gratis para almacenar una cantidad pequeña de conjuntos de datos. Sus precios son similares a los de Junar.

OpenDataSoft: *OpenDataSoft*²¹ es una empresa francesa que recientemente ha entrado en el mercado Estadounidense. Ofrecen un nivel de servicio gratis (hasta 5 conjuntos de datos, cada uno hasta 20,000 registros).

Socrata Open Data: *Socrata*²² es el vendedor mayor en la industria de repositorios de datos abiertos, con su plataforma de datos abiertos *Socrata Open Data*, *Socrata* solamente ofrece opciones con alojamiento – no hay manera de correr el software de *Socrata* en sus propios servidores. Es la opción con más funciones y es la más cara, con planes que cuestan hasta cientos de miles de dólares al año.

2.2.3. Alojamiento gratis

Hay algunas opciones para el alojamiento gratis de repositorios de datos abiertos (Nota que las opciones de código abierto también son gratis, pero requieren instalación, un servidor, y tiempo de mantenimiento). Por lo general, esto es el nivel más básico de servicio provisto por organizaciones que cobran.

¹⁷ <https://hub.arcgis.com/pages/open-data>

¹⁸ <http://www.junar.com>

¹⁹ <https://getdkan.org>

²⁰ <http://www.civicdashboards.com>

²¹ <https://www.opendatasoft.com>

²² <https://socrata.com>

Datahub: la *Open Knowledge Foundation* (Fundación para Conocimiento Abierto) provee a *DataHub*²³, que es un servicio con alojamiento gratis basado en CKAN. Es un repositorio grande colectivo – los usuarios no tendrán su propia sitio web, aunque es posible subir sus propios datos y compartir el URL que solamente tiene a esos conjuntos de datos.

Github: Github²⁴ realmente no está hecho para ser un repositorio, pero puede servir como uno. No tiene las características buenas de un software específico para repositorios (como conversión de formatos, de poder conseguir datos de URL remotos, etc.), pero si ofrece un avance o vista previa para algunos tipos de datos, registra cambios públicamente, y es un lugar razonable donde almacenar conjuntos de datos.

Si ofrece una ventaja, que es que GitHub – diferente que otro software de repositorio – provee un mecanismo para que la gente proponga cambios a sus conjuntos de datos que puedes aceptar o rechazar, si encuentran errores o áreas para mejorar.

JKAN en Github: JKAN²⁵ está diseñado a ser implementado en GitHub, donde el catálogo de datos que resulta puede ser alojado gratis. Así, Github puede servir como alojamiento sin sacrificar a los buenos aspectos de un catálogo de datos.

3. La gestión de un repositorio de datos abiertos

3.1. Interoperabilidad y el intercambio de datos

Según los principios *FAIR*, los datos tienen que ser “Findable (fácil de encontrar), Accesible (accesible), Interoperable, y Reusable.” Los principios FAIR de datos sirven como una guía internacional para mayordomía de datos de alta calidad. Se recorre con más detalle el intercambio de datos y mejores prácticas en la Unidad 4: Intercambiando Datos Abiertos.

Para ser fáciles de encontrar:

- Se asignan un identificador único y persistente a (Meta)datos
- Se describe datos con metadatos detallados
- Los (meta)datos están registrados o catalogados en un índice en un recurso que se puede buscar
- Los metadatos especifican al identificador de los datos

Para ser accesibles:

- Se puede encontrar a (meta)datos por su identificador único utilizando un protocolo de comunicaciones estándar
- El protocolo es abierto, gratis, y puede ser implementado de manera universal.
- Los (meta)datos usan un vocabulario que sigue a los principios FAIR
- Los metadatos incluyen a referencias calificadas a otros (meta)datos

²³ <http://datahub.io>

²⁴ <https://github.com>

²⁵ <https://how-to.usopendata.org/en/latest/The-Basics-of-Open-Data/Data-Repositories/#jkan>

Para ser Interoperables:

- Los (meta)datos usan un lenguaje formal, accesible, compartido, y aplicable ampliamente para representación de conocimiento
- Los (meta)datos usan vocabularios que siguen los principios FAIR

Para ser reusables:

- Los (meta)datos tienen varios atributos correctos y relevantes
- Los (meta)datos están publicados con una licencia de uso clara y accesible
- Los (meta)datos están asociados con su procedencia
- Los (meta)datos cumplen con los estándares comunitarios relevantes al dominio

3.3.1. Publicando a Metadatos

El requisito para metadatos que son informativos, consistentes, y útiles no es nuevo, pero todavía es un reto para muchos conjuntos de datos. Tener metadatos precisos es vital no solamente para hallar fácilmente, sino también para catalogar – los metadatos de baja calidad pueden socavar a la utilidad del repositorio en sí. Koesten et al.²⁶ exploraron las necesidades de búsqueda y lógica de datos de 20 profesionales que trabajan con ellos, incluso aspectos de evaluar si un conjunto de datos sea relevante o no. Distinguen entre tres dimensiones: la relevancia (¿son datos que necesito?; utilidad (¿puede usarlo en la práctica?) y calidad (¿cuán buenos son los datos y cuan fácil será usarlos?). Los datos deben estar acompañados por descripciones de estos aspectos, o como metadatos estructurados, o también en la forma de comentarios, estudios de casos, reportes de experiencias, ejemplos de uso, etc. Vea a la Tabla 3 abajo.

Tabla 3: Los metadatos que a los usuarios consideran útil para entender a los datos

Evaluar	Información Necesario
Relevancia	Contexto, cobertura, propósito original, granularidad, resumen, tiempo que representan
Utilidad	Si están etiquetados, la documentación, licencia, acceso, legibilidad por máquina, lenguaje usado, formato, esquema, pueden ser compartidos
Calidad	Métodos de colección, procedencia, consistencia de formato/etiquetado, lo completo, que falta

²⁶ Koesten, L, Kacprzak, E, Tennison, J and Simperl, E (2017) Trials and Tribulations of Working with Structured Data - a Study on Information Seeking Behaviour *CHI '17 Proceedings of the 2017 ACM SIGCHI Conference on Human Factors in Computing Systems* ACM, New York, USA. <http://dx.doi.org/10.1145/3025453.3025838>

3.4. Consideraciones para la gobernanza

Mejor transparencia y responsabilidad de la gobernanza son claves en determinar el valor de datos abiertos para la agricultura y la nutrición. Es importante considerar cuáles datos abrir y cómo. ¿Cuán responsables son los dueños de los datos ambos en el lado de la audiencia y de los políticos responsables? ¿Como productores y actores pueden asegurar a la calidad de datos? ¿Quiénes son los mayordomos que están encargados a hacer que los datos sean abiertos?²⁷

Países u organizaciones diferentes tendrán diferentes modelos para gobernar y administrar a sus actividades – o diferentes modelos de gobernanza. Por ejemplo, encontrarás que algunos países no delegan a la toma de decisiones mientras que otros tienen una administración pública más centralizada. Estos modelos de gobernanza claramente pueden impactar a como se gobierna a los datos abiertos – haciendo una gran red de diferentes estilos de gobernanza de datos abiertos alrededor del mundo haciendo difícil identificar quienes son los que toman decisiones y los que guardan al acceso o son mayordomos dentro de un país específico.

Por ejemplo, si alguien quiere acelerar la apertura de datos en agricultura puede ser bajo la autoridad de un gobierno sub-nacional (como estados, provincias, territorios o hasta ciudades, mientras que en otras ciudades la agricultura es gobernada por un gobierno central o implementado por socios público-privados. Los datos de gobernanza pueden estar privatizados, mientras que en otros casos puede ser la responsabilidad de un gobierno municipal o regional. Por lo tanto, las responsabilidades están distribuidos por múltiples niveles administrativos y agencias que afectan cuan abiertos son los datos producidos y publicados.

Al considerar el modelo de gobernanza de datos abiertos para su iniciativa de datos abiertos, debes hacer un bosquejo del proceso de gobernanza en datos abiertos y del entorno por identificar a las siguientes partes interesadas, sus roles y responsabilidades en la administración de datos abiertos, y buscar cómo están conectados:

- *Los que toman decisiones* – quien toma/lidera la autoridad de datos abiertos en reuniones, protocolos, conducto, debates, votaciones, y otros asuntos?
- *Los dueños de datos* – cuáles organizaciones/cuerpos gubernamentales producen cuales datos agrícolas y nutricionales?
- *Actores de control de calidad* – ¿quiénes son los actores que aseguran que los datos producidos conforman a ciertos estándares de calidad?
- *Los que controlan acceso* – ¿Quién controla la publicación de datos abiertos?

3.5. Curación y Preservación

La curación y preservación combina a las políticas, estrategias y acciones para asegurar a la producción de datos de alta precisión tras el tiempo, sin importar a las dificultades con corrupción de archivos, fallos

²⁷ Brandusescu, A., Lämmerhirt, D. AND Verhulst, S. (2017) Mapping open data governance models: Who makes decisions about government data and how? <https://blog.okfn.org/2017/02/16/mapping-open-data-governance-models-who-decidesand-how/>

en los medios y cambios tecnológicos. La preservación digital aplica al contenido que nace o es convertido en formato digital.

Una vez que se elige al repositorio, la publicación de datos al repositorio puede ser un reto con varias dificultades. La primera dificultad para enfrentar es la selección de datos para publicar. Todos los datos tienen valor, o solamente debe hacer que una porción de datos esté disponible.

Dado que los recursos no son infinitos, se tiene que tomar una decisión de una forma, pero ¿cuáles criterios se deben usar para identificar o cuáles datos son adecuados para publicar? Todas estas consideraciones están articuladas en políticas y flujos de trabajos que en realidad siempre están evolucionando. Esto será elaborado en la lección 3.5. Por lo tanto, el objetivo no es establecer prácticas específicas y responsabilidades de manera permanente, sino documentar políticas nuevas y existentes con la intención de seguir a los cambios mientras que ocurren.

La gente que tienen un rol primario en gestionar o “cuidar de” datos vienen de muchas formas²⁸. Estos puestos incluyen a los:

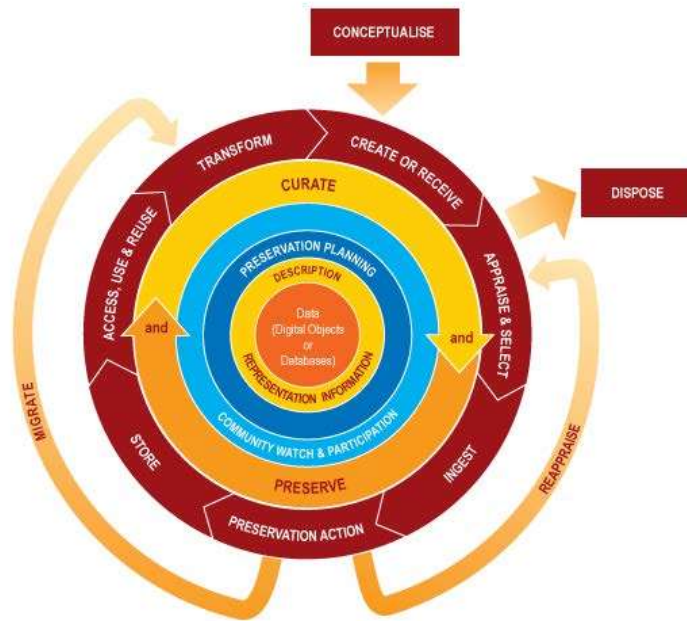
- Curadores de datos
- Archiveros
- Bibliotecarios
- Bibliotecarios de datos
- Anotadores

Como ejemplo, las responsabilidades de un curador de datos en un contexto de las biociencias puede incluir la gestión de datos continua, la descripción intensiva de datos, asegurando a la calidad de datos, el trabajo en la infraestructura de información colaborativa, y trabajo en estándares de metadatos. La gama de tareas y responsabilidades encapsuladas por la curación de datos puede ver algo así:

- Desarrollar e implementar políticas y servicios
- Analizar al contenido digital para determinar cuáles servicios pueden ser provistos de ello.
- Proveer consejos para creadores de datos para utilizadores/reutilizadores
- Asegurar que los datos están subidos al repositorio con éxito
- Negociando acuerdos
- Asegurar la calidad de datos
- Asegurar que los datos están estructurados de la mejor forma para proveer el acceso, actualización, almacenamiento, y mantenimiento
- Facilitar el uso y rehusos de datos
- Facilitar el hallazgo y obtención de datos
- Planificación de preservación e implementación (por ejemplo, asegurar el almacenamiento adecuado y rutina de respaldo, el monitoreo de obsolescencia)
- Asegurar que las políticas y servicios están establecidos para asegurar que los datos sean viables, se pueden generar correctamente, entendibles y auténticos.
- Promover la interoperabilidad

²⁸ DCC – Digital Curation 101: Curate and Preserve.

<http://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20Curation%20and%20Preservation.pdf>



Clockwise, outer circle

[Conceptualizar]

[Crear o recibir]

[evaluar y seleccionar] –[Desechar]

[ingerir]

[acción de preservación] <– [reevaluar] -> [migrar]

[almacenar]

[acceder, usar, reutilizar]

[transformar]

Inner circle [Curar] y [Preservar] y

Inner circle [Observaciones y Participación de la comunidad]

Inner circle [planificación de preservación]

Inner circle [descripción, representación, información]

Inner circle [datos, objetos o bases de datos]

Gráfica 1: El Modelo de Curación de Ciclo de Vida DCC

El Modelo de la Curación DCC²⁹ (vea gráfica 1) provee una vista general en forma gráfica de las etapas y requisitos para la curación exitosa y preservación de datos desde la conceptualización inicial. El modelo se puede usar para planificar actividades dentro de una organización o consorcio para asegurar que todas las etapas necesarias se toman cada uno en su secuencia correcta. El modelo permite que la funcionalidad granular sea representada; para definir roles y responsabilidades y construir un marco de estándares y tecnologías para implementar. Se puede ayudar con el proceso de identificar pasos adicionales que pueden ser requeridas, o acciones que no son requeridas por ciertas situaciones o disciplinas, y asegurando que las políticas y procesos estén documentados adecuadamente.

Tabla 4: Acciones dentro del Modelo de Curación de Ciclo de Vida DCC

Acciones de un Ciclo de Vida Completo	
Información de Descripción y representación	Asignar metadatos administrativos, descriptivos, técnicos, estructurales y de preservación, utilizando estándares adecuados, para asegurar descripción y control adecuado en el largo plazo. Colecciona y asigna información de representación necesario para entender y generar ambos el material digital y los metadatos asociados.
Planificación de preservación	Planifica para la preservación por todo el ciclo de vida de curación de materiales digitales. Esto incluye a planes para la gestión y administración de todas las acciones en el ciclo de vida de curación.
Observación y participación de la comunidad	Observa actividades relevantes de la comunidad, y participa en el desarrollo de estándares comunes, herramientas, y software adecuado.
Curar y preservar	Estar consciente de realizar acciones de gestión y administrativas planificados para promover la curación y preservación por todo el ciclo de vida de curación.
Acciones Secuenciales	
Conceptualizar	Conceptualizar y planificar para la creación de datos, incluso al método de capturar y almacenarlos.
Crear o recibir	Crea datos incluso a los que son administrativos, descriptivos, estructurales, y metadatos técnicos.
Evaluar y Seleccionar	Evaluar datos y seleccionar para la curación y preservación de largo plazo. Sigue guías documentadas, políticas, o requisitos legales.
Ingerir	Transferir datos a un archivo, repositorio, centro de datos, u otro custodio. Sigue guías documentadas, políticas, o requisitos legales.

²⁹ <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>

Acción de preservación	Realiza acciones para asegurar la preservación y retención de largo plazo de la naturaleza autoritaria de datos. Acciones de preservación deben asegurar que los datos siguen siendo auténticos, confiables, y útiles mientras que mantienen su integridad. Acciones incluyen la limpieza, validación, y asignación de metadatos de preservación, asignando información de representación y asegurando estructuras o formatos de archivos aceptables.
Almacenar	Almacenar a los datos de manera segura siguiendo estándares relevantes.
Acceso, utilización y reutilización	Asegurar que datos están accesibles ambos para usuarios designados y reutilizadores, de forma diaria. Esto puede ser en la forma de información público. Controles robustos de acceso y protocolos de autenticación pueden ser aplicables.
Transformar	Crea nuevos datos de los originales, por ejemplo, por migrarlos a otro formato, por crear un subconjunto de datos, por selección o consulta, por crear resultados nuevamente derivados, tal vez para publicación
Acciones Ocasionales	
Desechar	Desechar de datos que no han sido seleccionados para la curación y preservación de largo plazo según políticas documentadas, guías o requisitos legales. Típicamente los datos pueden ser transferidos a otro archivo, repositorio, centro de datos o custodio. En algunos casos se destruye los datos. La naturaleza de los datos puede, para razones legales, requerir su destrucción.
Reevaluar	Regresar datos que no aprueban los protocolos de validación para reevaluación y reelección
Migrar	Migrar datos a otro formato. Esto puede ser para acordar con el formato de almacenamiento o para asegurar que no se vuelvan obsoleto por causa de Hardware o Software

Resumen

Un repositorio sirve como un punto central para un esfuerzo de datos abiertos. Sirve como un lugar central para encontrar a datos, un lugar para establecer estándares de prácticas, y algo para demostrar el uso de esos datos. En un sentido práctico, un repositorio sirve como un lugar central donde la gente puede buscar y hallar a datos.

Antes de elegir una solución para el repositorio, primero hay que considerar a las necesidades de su uso y establecer a los requisitos:

- Entender lo más posible a los usuarios, igual que su trabajo y el contexto de su trabajo
- El sistema bajo desarrollo debe ayudar que los usuarios logran sus metas
- Construir sobre las necesidades identificadas y producir un conjunto de requisitos
- Utilizar un enfoque centrado en los usuarios para el desarrollo
- Entender que debe hacer un repositorio y asegurar que cumple con las necesidades de los partes interesados.

Cuando consideramos utilizar el modelo de gobernanza de datos abiertos en su iniciativa de datos abiertos, hay que implementar el proceso de gobernanza de datos y el ecosistema, identificando y siguiendo a los partes interesados claves, sus roles y responsabilidades en la administración de datos abiertos, y buscando entender cómo se conectan:

- Los que toman decisiones
- Los que son dueños de los datos
- Los que producen datos
- Los actores de aseguranza de calidad de datos
- Los que controlan a los datos

La curación y preservación combina políticas, estrategias y acciones para asegurar la generación más precisa de datos en el tiempo, sin importar los retos de la corrupción de archivos, el fallo de los medios, y cambio tecnológico.

El modelo del Ciclo de Vida de Curación DCC provee una vista general de las etapas requeridas para la curación exitosa de datos desde la concepción inicial. El modelo se puede usar para planificar actividades dentro de una organización o consorcio para asegurar que las etapas necesarias se cumplen cada una en la secuencia correcta.