



Gestión de datos de investigación en repositorios

Seminario Internacional de
Repositorios Digitales y Acceso Abierto



Joel Torres
Octubre, 2021

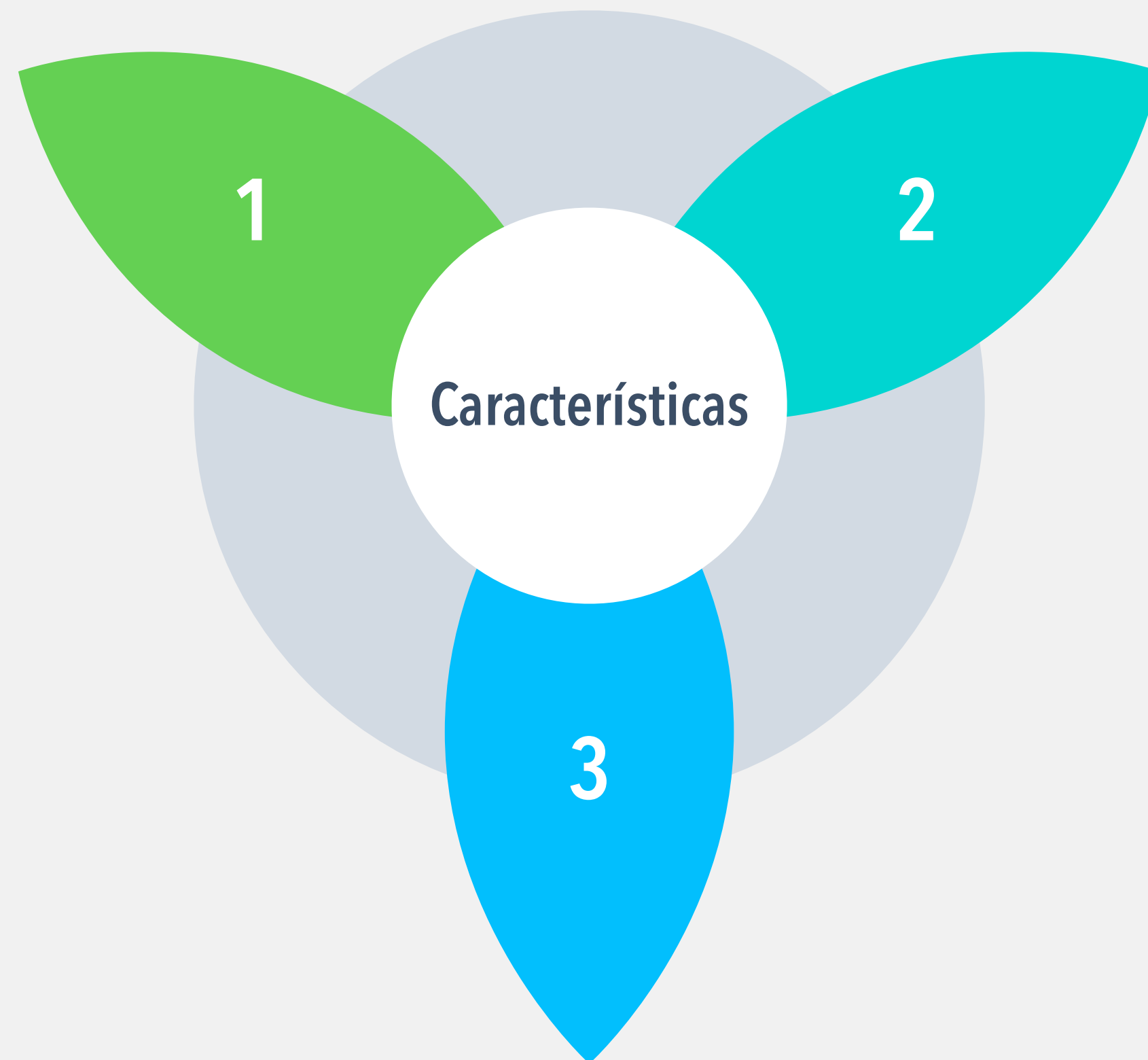
Necesidad creciente

¿Por qué es clave gestionar datos?



Para ser datos abiertos

Características básicas de los datos de investigación abiertos



- **Volver a usar**

Para reproducir y comprobar los resultados. Se requiere de datos completos, con información de contexto, sin restricciones técnicas.

- **Utilizar para otros fines**

Para lo que se requiere acceso completo, sin restricciones legales, con licencias apropiadas.

- **Distribuir**

Sin restricciones financieras, para cualquier uso, en beneficio de la sociedad.

Publicar los datos

Las tres formas más comunes

De acuerdo con el tipo y el destino de publicación. Siempre usando PID

**Material
suplementario**

**Artículo de
datos**

En revistas evaluadas por pares. Siempre usando PID y publicando de forma abierta

Utilizando un PID y asegurando que el repositorio cumpla con las condiciones

**Repositorio
de datos**

Qué son datos de investigación

Características y tipología

3 tipos de datos:

Observaciones. Registros históricos, irrepetibles.

Experimentales. Datos que acompañan a los experimentos.

Computacionales. Datos que acompañan las simulaciones.

- Experimentos científicos
- Modelos y simulaciones. Modelos con metadatos asociados.
- Observaciones. Formularios, censos, registros de votos, etc.
- Datos derivados. Resultado del proceso de combinación de datos sin procesar con otros tipos de datos
- Datos referenciales. Fórmulas, estructuras químicas, etc.
- Material complementario. Instrucciones de codificación, gráficos sobre los datos, diccionarios de datos, etc.

¿Cómo abrir los datos?

Recomendaciones del **Open Data Handbook**

- **Mantenlo simple**

Iniciar con algo pequeño, sencillo y rápido.
Se vale innovar e ir aprendiendo de la experiencia.

- **Relacionarse temprano y con frecuencia**

Con los usuarios actuales y potenciales de los datos, tan pronto y tan a menudo como sea necesario
Con otros actores que asegurarán la interpretación, la visualización, el nivel de apertura

- **Soluciona temores y malentendidos comunes**

Temáticas relacionadas con la seguridad, la libertad, los aspectos legales, la participación comercial, etc.



Los datos FAIR

Estructura y la ruta de los datos

<https://fairaware.dans.knaw.nl/>



Ruta de los datos FAIR



Tratamiento

Organización, estructura,
nombres, versiones

Cuáles se abrirán
públicamente, apego a
reglamentación,
autorización

SELECCIÓN

- A. Eliminar identificadores directos (nombres, direcciones).
- B. Agregar o reducir pseudónimos (reemplazar fecha de nacimiento por edad).
- C. Generalizar el signo de un ejemplo "cuerpo de negocios").
- D. Anonimizar datos y cruzarlos con otras fuentes (teléfono, email, etc).
- E. Anonimizar datos geográficos (coordenadas o IP) para permitir acceso.
- F. Utilizar pseudónimos.

Ruta de los datos FAIR



Anonymization Wizard

Restart

Source

Anonymized

Hierarchy

Algorithms

Solution Graph

Results

Amnesia Site

Results

version: 1.2.7 beta

Save To Local

Save To Zenodo

Save To Dataverse

DataSet

Show 10 entries

Saludo	Nombre	Apellidos	Depto	Cargo	Organizacion
Ing.	Aaron	Lescarez	Redes	Administrador	FLACSO - Facultad Latinoamericana de Ciencias Sociales
Dr.	Abelardo	Marina Flores	Economia	Jefe de Departamento de Economia	UAM - Universidad Autonoma Metropolitana Unidad Acatzaco
Lic.	Abigail	Pizarro		Asistente de Jefatura de adquisiciones directas	UACJ - Universidad Autonoma de Ciudad Juarez
	Abraham	Mosqueda Salas	Adquisiciones y Contratos	Jefe Adquisiciones y Contratos	COLPOS - Colegio de Postgraduados
Lic.	Abraham	Garcia		Biblioteca jefe administrativo	UA - Universidad Iberoamericana Puebla

Anonymized DataSet

Saludo	Nombre	Apellidos	Depto	Cargo	Organizacion
Ing.	A****	Lescarez	Redes	Administrador	FLACSO - Facultad Latinoamericana de Ciencias Sociales
Dr.	Ab****	Marina Flores	Economia	Jefe de Departamento de Economia	UAM - Universidad Autonoma Metropolitana Unidad Acatzaco
Lic.	Ab****	Pizarro	[null]	Asistente de Jefatura de adquisiciones directas	UACJ - Universidad Autonoma de Ciudad Juarez
[null]	Abraham	Mosqueda Salas	Adquisiciones y Contratos	Jefe Adquisiciones y Contratos	COLPOS - Colegio de Postgraduados
Lic.	Abraham	Garcia	[null]	Biblioteca jefe administrativo	UA - Universidad Iberoamericana Puebla

Ruta de los datos FAIR



Dirección web única que nos permite acceder a un recurso digital de forma inequívoca, independientemente de la ubicación física o lógica en la que se encuentre alojado

URI ó HANDLE

Ruta de los datos FAIR

ORCID
Personas

ROR
Organizaciones de investigación

DOI + ORCID + ROR = Colección pública de metadatos de investigación

I4OC (Initiative for Open Citations)

PID Graph

FORMATO

Apertura W3C

Anonimización

Identificación, precisión, generalización, pseudónimos

Tratamiento

Organización, estructura, nombres, versiones

SELECCIÓN

Cuáles se abrirán públicamente, apego a reglamentación, autorización



Ruta de los datos FAIR

Difusión



Licencias

Creative Commons u Open Data Commons, NO restrictivas



IDENTIFICADOR

PID

FORMATO

Apertura W3C

Anonimización

Identificación, precisión, generalización, pseudónimos



Tratamiento

Organización, estructura, nombres, versiones

SELECCIÓN

Cuáles se abrirán públicamente, apego a reglamentación, autorización



Cita tus datos de investigación



Por qué es importante citar los datos:

- Los conjuntos de datos también son resultados de investigación como los artículos, monografías, etc.
- Facilita la identificación y el acceso a los datos y de esta forma su localización, validación y reutilización.
- Permite reconocer la autoría de sus creadores.
- Facilita la métrica e impacto de los datos.
- Favorece la transparencia de la investigación científica.

Buenas prácticas para citar datos:

- Se debe facilitar la identificación, localización y el acceso a los datos mediante un identificador único y persistente (DOI, Handle, etc.).
- Cada conjunto y subconjunto de datos (dataset) debe citarse de forma independiente.
- Las citas de los datos utilizados han de aparecer en la sección de referencias bibliográficas de la publicación resultante.
- Se recomienda incluir un identificador único de autor (ORCID, etc.).



Enlaza los datos con los documentos resultado de investigación y viceversa, y crea las referencias bibliográficas de los mismos.

Elaboración de la cita

- Existen elementos mínimos obligatorios (O) y otros recomendados (R) que se combinan para elaborar la cita en cualquier estilo estándar (APA, MLA, Chicago, etc.) o los propuestos por los principales repositorios de datos (Dataverse, Dryad, etc.).

Autoría (O)

Identificador autor (R)

Fecha (O)

Título (O)

Identificador único persistente (O)

Tipo de recurso (O)

Versión y/o Edición (O)

Repositorio de datos (R)

Publicación (R)

Productor (R)

Ámbito geográfico (R)

Ámbito temporal (R)

Ejemplo de cita estilo APA

Autoría

Título

Remesar Betloch, X., Antelo, A., Ulvina, C., Albà, E., Berdié, L., Agnelli, S., ... Alemany, M., (2015). *Influence of a hyperlipidic diet on the composition of the non-membrane lipid 6 pool of red blood cells of male and female rats*. [Dataset]. Versión de 22 de junio de 2015. Dipòsit digital de documents de la UAB. <http://hdl.handle.net/2445/66010>.

Fecha

Repositorio

Tipo de recurso

Identificador único y persistente

Versión



El personal de tu Biblioteca te puede asesorar

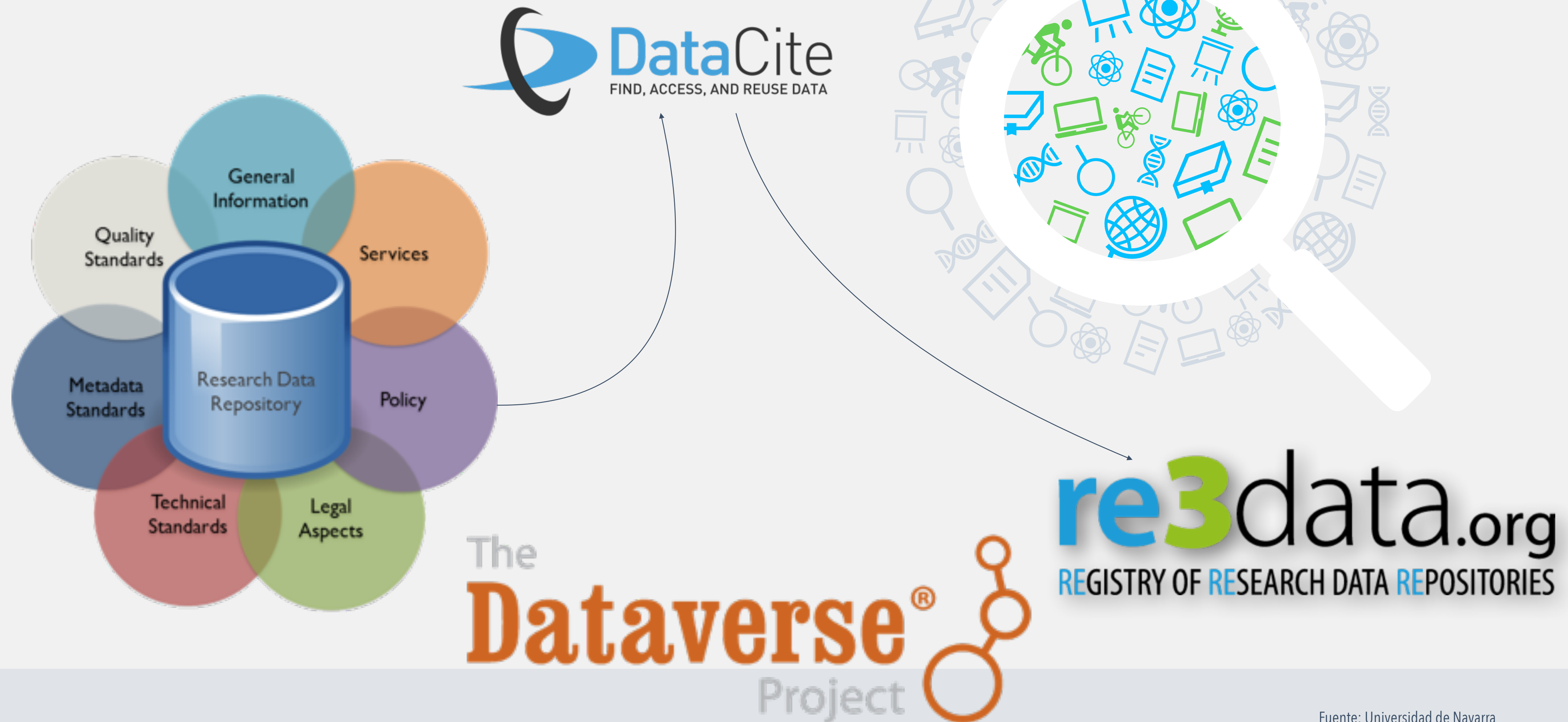
Los datos FAIR

Estructura y la ruta de los datos



Repositorios de datos

Difusión y posicionamiento





Gestión de datos de investigación en repositorios

Seminario Internacional de Repositorios Digitales y Acceso Abierto



Joel Torres
Octubre, 2021