

Porqué y para qué? Desafíos y beneficios de la gestión de datos abiertos de investigación

Maria Esteva

Gestión de Datos de Investigación en la UNR

Lunes 26 de Octubre, 2020

Datos de investigación

- Información que se recoge, utiliza, genera, produce, calcula, infiere, desde la planificación y a través de la conducción de una investigación.
- Elementos concretos sobre los que se basan las interpretaciones, resultados, y conclusiones de la investigación.
- Datos generados por otros investigadores o en otros contextos.
- Códigos creados para analizar los datos, paginas web y materiales didácticos, informes con explicaciones sobre datos.

Se generan datos en las investigaciones teóricas? Y [en Leyes?](#)

Porqué hay que publicarlos?

- Nuevos paradigmas de investigación.
- Beneficios sociales y económicos.
- Registro de la actividad académica.
- Evidencia del trabajo de investigación.
- Permiten validar, verificar, replicar la investigación.
- Permite su reutilización – retorno de la inversión.
- Beneficio para el investigador – impacto y exposición.

Todos los datos son abiertos?

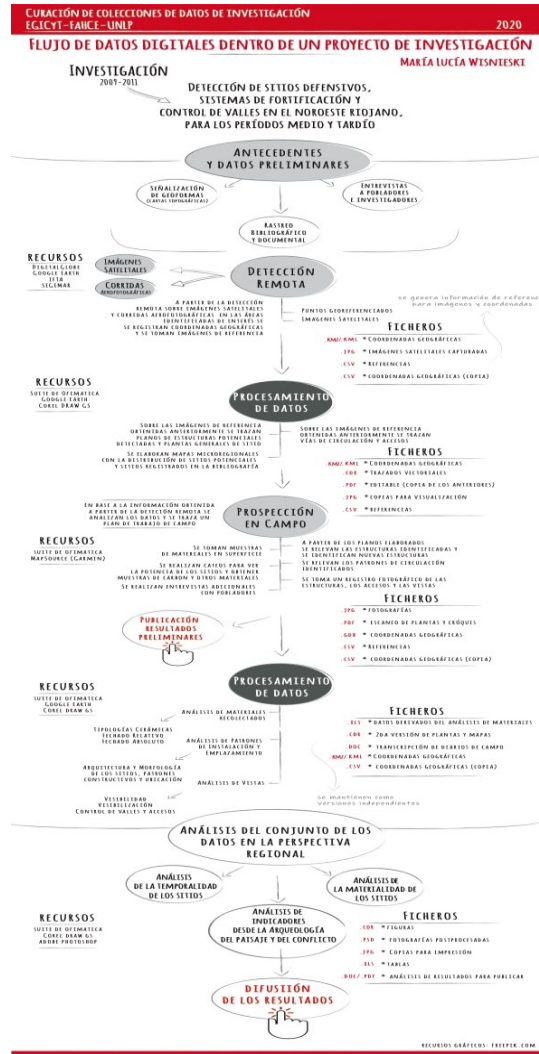
Curación de datos

- Para publicar datos primero hay que curarlos.
- Procesos y practicas mediante las cuales se aseguran la disponibilidad de los datos a través del tiempo.
- Organizar, describir, normalizar, depurar, corregir, validar.
- Comienza con el plan de investigación, atraviesa los pasos de la investigación hacia la publicación y continua durante la reutilización de los datos.

Como se van a entender, encontrar, y acceder los datos dentro de 10-20-100 meses/años?



Flujo de datos de la investigación



- Describan los pasos de su investigación hasta la publicación.
- Identificar los datos que se generan en cada paso.
- Determinar que elementos son necesarios describir para que otros pueden entenderlos y reutilizarlos datos.

Realizado por Maria Lucia Wisnieski durante el curso de Curación de Datos de Investigación, EGICyT, UNLP 2020

Principios de datos FAIR

• Encontrables

- Tienen DOI
- Tienen metadatos estandarizados?
- Se encuentran en Google?
- Se encuentran en un repositorio?

• Accesibles

- Son gratis?
- Tiene limitaciones de uso?
- Se pueden descargar?
- Se pueden abrir/ver?

• Interoperables

- Son estandarizados?
- Están en un formato abierto? (ASCII, CSV, TIFF, WAV, ...)
- Pueden ser fácilmente utilizados por humanos y computadoras?
- Requieren diccionarios para ser entendibles?

• Reutilizables

- Calidad de los datos
- Completos
- Entendibles

File Preview: 02072018@092140@114054@64.5rpm.bin

File Metadata

Copy Download

Unsupported File Type
Preview for this item is not supported.

Metadatos

- Descripción de los datos, su proveniencia, su lógica, sus creadores, sus sponsors.
- Existen estándares/vocabularios para describir datos en general y en particular para áreas científicas/de humanidades.
- Aplicarlos es una tarea interdisciplinaria.
- Permiten el descubrimiento, la búsqueda, la promoción, y el análisis de los datos por personas y por computadoras.
- Dos tipos generales:
 - Para describir la publicación y que se encuentre en línea
 - Para describir el contenido de los datos y que estos puedan ser reutilizados.
 - Quien, como, donde, porque?

Licencias y derechos

- Los datos no tienen derechos de autor.
- Las licencias de uso de datos indican las condiciones mediante las cuales los autores comparten sus datos con los demás.
- Licencias abiertas y mas cerradas.
- <https://opendatacommons.org/licenses/>
 - Compartir, crear, adatar
- Palabras claves:
 - Atribuir
 - Compartir del mismo modo (usar la misma licencia para publicar).
 - No restringir acceso
 - De dominio publico

Confidencialidad

- Ética de la investigación en relación a participantes.
- Datos personales y datos sensibles.
- Comité de ética?
- Compromiso en el plan de investigación con respecto a la confidencialidad y datos de los participantes.
 - Nos autorizaron a publicar datos, cuales?
 - Nos autorizaron a publicar fotos, videos, audio?
- Como voy a proteger los datos personales o sensibles durante la investigación y durante su publicación?
 - Anonimizar.
 - Evitar la divulgación involuntaria o por negligencia

Share or Perish

- Las publicaciones de datos se mantienen relevantes de diferentes formas.
 - Marketing
 - Quien es la audiencia?
 - Como llegar a ella?
 - Conocer el impacto de los datos
 - Visitas a la pagina de publicación
 - Descargas
 - Citas
- Formas de promover los datos
 - A través de las publicaciones sobre los resultados.
 - Ubicando las
 - Mediante exposición en conferencias.
 - Descubribles en Google.
 - Google Data Search.
 - Colocar las citas en el CV.
 - Mediante el repositorio universitario.

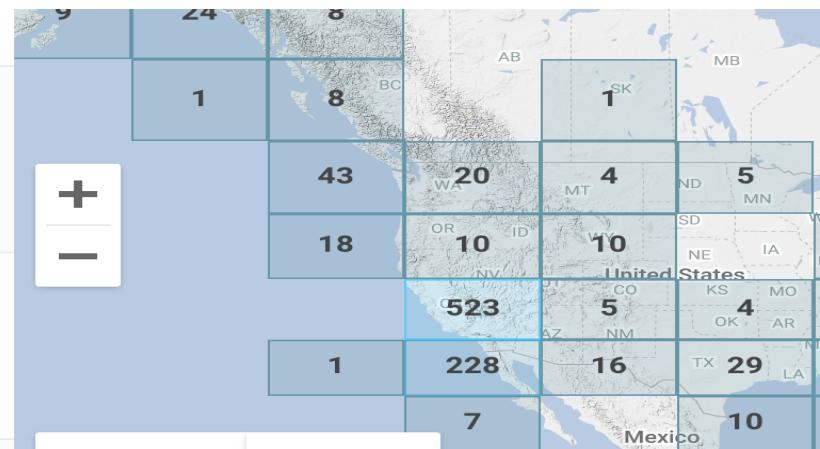
My Search
natural hazards

Filter by:

- ▶ Data attribute
- ▶ Annotation
- ▶ Data files
- ▶ Member Node
- ▶ Creator

Virginia Hatfield. 2017. **Islands of Four Mountains Artifact Analysis, Aleutian Islands, Alaska, 2013.** Arctic Data Center.
doi:10.18739/A2CC0TT7F. 136 downloads, 247 views

Michael Brady. 2017. **Arctic Slope Shoreline Change Susceptibility Spatial Data Model, 2015-16.** Arctic Data Center.
doi:10.18739/A27W6761N. 248 views



Consideraciones a tener en cuenta

- Que los datos no sean utilizados no quiere decir que la investigación sea mala.
- Los datos pueden ser reutilizados mucho tiempo después de ser creados.
- Diferentes tipos de datos tendrán diferentes grados de reutilización.
 - Observaciones.
 - Datos devenidos de hipótesis.
 - Datos sesgados.

Partes involucradas en la curación de datos

- El creador
 - La investigación es relevante para la comunidad científica de pertenencia y para la sociedad.
 - Datos validados, bien documentados, en formatos in.
- El repositorio/servicio
 - Opera los principios FAIR
- Los usuarios
 - Reutilizan los datos en correspondencia con las licencias de uso y las condiciones de confidencialidad.
 - Citan la utilización de los datos.

Discusiones grupales

- Presentarse e indicar que tipo de datos gestiona durante su investigación.
- Elegir un moderador del grupo que tomara notas.
- Elegir un presentador dentro del grupo que hará la síntesis de lo conversado a todos los participantes.

Discutir en grupo: Actividad I

- Cada participante comente:
 - Si ha publicado datos con anterioridad.
 - Si usa datos de otras fuentes en su investigación (no bibliografía)
 - Si conoce lo que es y para que sirve un DOI

Observen un caso de publicación de datos según su especialidad y comenten si entienden las descripciones y de que tratan los datos. En que coinciden y en que difieren?

Abogacía:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/27327>

Agricultura

<https://data.worldagroforestry.org/dataset.xhtml?persistentId=doi:10.34725/DVN/25YDHZ>

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/RJOYV7>

Arquitectura:

<http://io.morphocode.com/urban-layers/>

<https://data.buenosaires.gob.ar/dataset/tejido-urbano>

Salud pública:

<https://data.buenosaires.gob.ar/dataset/operativo-detectar>

<https://datos.uchile.cl/dataset.xhtml?persistentId=doi:10.34691/FK2/J8QTPH>

Ciencias Políticas

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/4BPRUI>

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OZGUH9>

Discutir en grupo: Actividad II

- Identifiquen las ventajas de compartir/ publicar los datos.
 - por ej.
 - Quien/es utilizarían sus datos y cuando?
 - Qué nuevos descubrimientos/conclusiones habilitarían?
- Identifique desventajas de compartir/publicar los datos.
- Qué apoyo logístico, técnico, de capacitación serian necesarios para estimular la publicación de datos en sus áreas de investigación?