

# Métricas basadas en datos

Autor: Isaac Hernández Arizo

Tutora: Antonia Ferrer Sapena

**MUGI** Curso: 2015-2016  
3ª edición

# Índice

1. Introducción
2. Estado de la cuestión
3. Métricas
4. Resultados
5. Conclusiones

# Introducción

- El término “open access” o “acceso abierto” hace referencia al acceso a fuentes de información científica en formato electrónico.
- El acceso se efectúa a través de Internet, y de forma gratuita.
- Cambio de paradigma:
- Usuarios
- Propiedad intelectual

# Introducción

El principal objetivo de este trabajo final de máster es conocer los distintos tipos de métricas que se están utilizando actualmente para medir la reutilización de los datos. Tipos de datos:

- Datos abiertos de investigación
- Datos de gobierno abierto

Análisis de repositorios de datos de investigación

# Estado de la cuestión

- Datos abiertos
- Datos abiertos de investigación
- Datos de gobierno abierto
- Tipos de archivo

# Datos abiertos

Según el Open Data Handbook los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, como mucho, a la obligación de atribución y de compartirse de la misma manera en que aparecen.

- Disponibilidad y acceso
- Reutilización y redistribución
- Participación universal

# Datos abiertos de investigación. Miembros

- Donantes: máximo provecho de los fondos de investigación.
- Infraestructuras de investigación: conservación y uso.
- Científicos: prestigio académico.
- Centros de datos: almacenamiento.
- Editores: tratar los datos que de las publicaciones.
- Bibliotecas: los datos estén identificados y accesibles.
- Bases de datos de publicaciones: enlazar publicaciones y citas de datos.

# Datos de gobierno abierto

- El gobierno abierto (open government, en inglés) es un concepto cada vez más extendido que se basa en tres pilares: transparencia, participación y colaboración.
- Modelo de democracia más participativo: Gobierno electrónico + TIC
- Para llegar a disponer de un gobierno abierto, hay dos requisitos imprescindibles:
  - Open data
  - Open action



# Tipos de archivo

## Clasificación de Tim Berners-Lee

★ Datos o documentos disponibles en la web con formato no estructurado y datos o documentos que se pueden visualizar en la web pero no procesar automáticamente.

Una imagen en formato JPG o PNG o un documento escaneado en formato PDF.

★★ Datos o documentos estructurados, procesables automáticamente y con formato propietario.

Un ejemplo sería una hoja de cálculo en formato Excel (con extensión .xls o .xlsx).

★★★ En este tercer nivel estarían los datos o documentos con un formato estructurado y abierto.

Una hoja de cálculo en formato CSV.

# Tipos de archivo

★★★★ Documentos que pueden ser referenciados mediante direcciones web. También encontraríamos documentos que siguen formatos estándares y abiertos del W3C.

Un ejemplo sería representar en el modelo RDF los edificios de un organismo público, con sus datos de contacto y ubicación.

★★★★★ En el nivel más alto tendríamos los datos enlazados con otros datasets.

Un ejemplo lo podríamos encontrar uniendo las descripciones de los edificios públicos del apartado anterior con enlaces a Geonames, una gran base de datos especializada en localizaciones geográficas.

# Métricas

- Indicadores bibliométricos
- Métricas tradicionales
- Métricas actuales
- Portales

# Indicadores bibliométricos

- Los indicadores bibliométricos son la herramienta utilizada para medir la actividad científica.
- Se utilizan para analizar el tamaño, crecimiento y distribución de la bibliografía científica (ya sean libros, revistas, patentes, etc.) (Sancho, 1990).
- Sirven para valorar la actividad científica, y la influencia (o impacto) tanto del trabajo como de las fuentes (Camps, 2008).

# Indicadores tradicionales. FI

- El factor de impacto (FI), fue ideado por Eugene Garfield en 1955 para identificar las revistas que debían estar en el Science Citation Index (SCI). El FI de una revista es el número de veces que son citados en un año los artículos de una determinada revista publicados en los dos años anteriores.
- El factor de impacto es la primera medida objetiva y cuantificable de la valoración de una revista en el ámbito científico.

# Ventajas del FI

Utilizar el factor de impacto tiene las siguientes ventajas: es universal, fácil de usar y de entender. En el año 2010 hay más de 10.500 revistas indexadas de más de 80 países y clasificadas en más de 200 categorías. También puede ser importante para los bibliotecarios o los especialistas en documentación a la hora de seleccionar las revistas para la colección (Velasco y otros, 2012).

# Defectos del FI

- Los problemas principales del FI son los siguientes:
- Un tiempo de cálculo de citas de dos años se debería utilizar sólo en áreas con un rápido envejecimiento de las publicaciones.
- El factor de impacto de las revistas no se debe utilizar para evaluar a los autores.
- El factor de impacto no separa las citas dirigidas a artículos de investigación de otros documentos como cartas o editoriales.
- En el SCI se prefiere incluir revistas que se publican en inglés.
- El factor de impacto de una revista se ve afectado por el tamaño de su comunidad científica.

# Indicadores tradicionales. Índice h

- Este modelo consiste básicamente en ordenar las publicaciones de un autor de forma descendente según las citas recibidas por cada trabajo.
- Encontramos el índice h de un autor cuando la posición en la lista es mayor o igual al valor de la cita (Velasco y otros, 2012).



# Ventajas del índice h

Este indicador, se dirige en principio a investigadores y ofrece las siguientes ventajas (Aleixandre Benavent, Valderrama Zurián y González Alcaide, 2007):

- Es una cifra que sólo puede aumentar o quedarse fija.
- Se dirige a los científicos de forma individual y permite comparar la producción de científicos de edades diferentes.

# Defectos del índice h

- Los autores pueden publicar documentos científicos de disciplinas que no sean su especialidad y recibir muchas citas.
- Depende en gran manera de los años de trabajo del investigador.
- Depende de la cantidad de científicos que trabajen en un determinado área.

# Métricas actuales

- Son indicadores basados en la web 2.0 o web social.
- Se denominan alternativos porque miden la repercusión de la actividad científica de una forma no habitual. Se obtienen al contar los “me gusta”, de Facebook, menciones en blogs, número de tuits o veces que un artículo se guarda en un gestor de referencias.
- James Wilsdon y sus compañeros ofrecen la siguiente definición: Altmetrics son métricas no tradicionales que cubren además del número de citas, las descargas, las publicaciones en redes sociales y otras medidas de impacto de los resultados de la investigación. (Wilsdon y otros, 2015).

# Métricas actuales. Defectos

Los principales defectos de las altmetrics son (Peralta González, Frías Guzmán y Gregorio Chaviano, 2015):

- Algunos artículos científicos tienen muy poco impacto en las redes sociales, pero en otras fuentes (bases de datos e índices de citas) poseen resultados más visibles.
- Las fuentes de las que se extraen estos datos pueden ser herramientas inestables. En otras palabras, quizás dejen de existir en un periodo de tiempo corto.

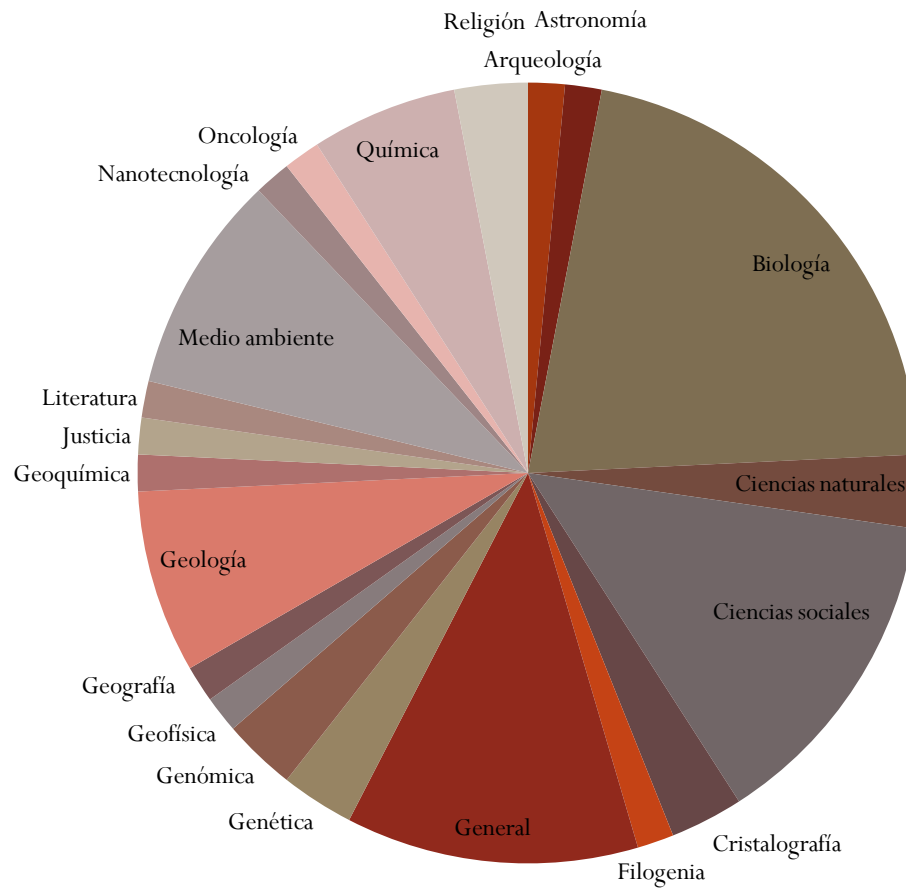
# Portales

- ImpactStory
- PLoS
- PlumX
- Altmetric
- ResearchGate
- Google Scholar
- Microsoft Academic Search

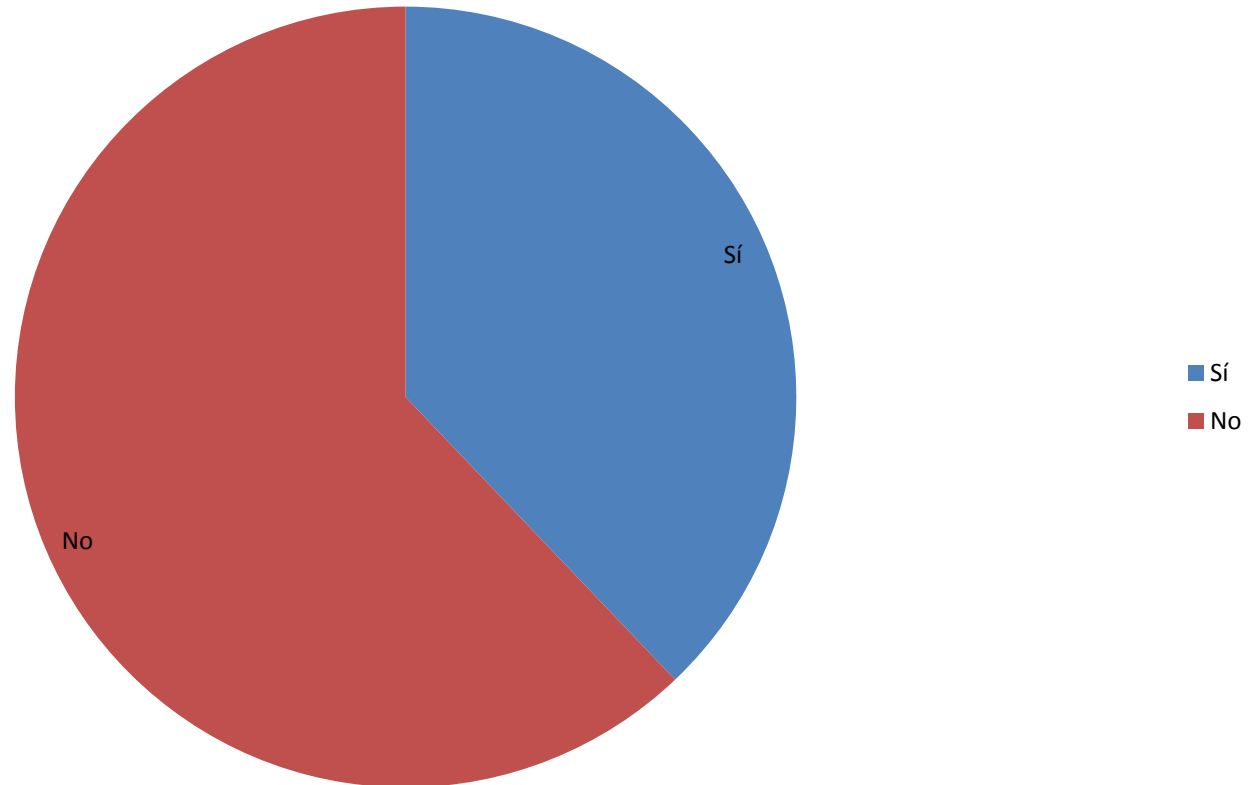
# Resultados

- Repositorios por área científica
- Referencia bibliográfica del dataset
- Identificador del dataset
- Visualizaciones de página
- Número de descargas
- Difusión en redes sociales
- Número de veces que se difunde el dataset
- Difusión en gestores bibliográficos

# Repositorios por área científica

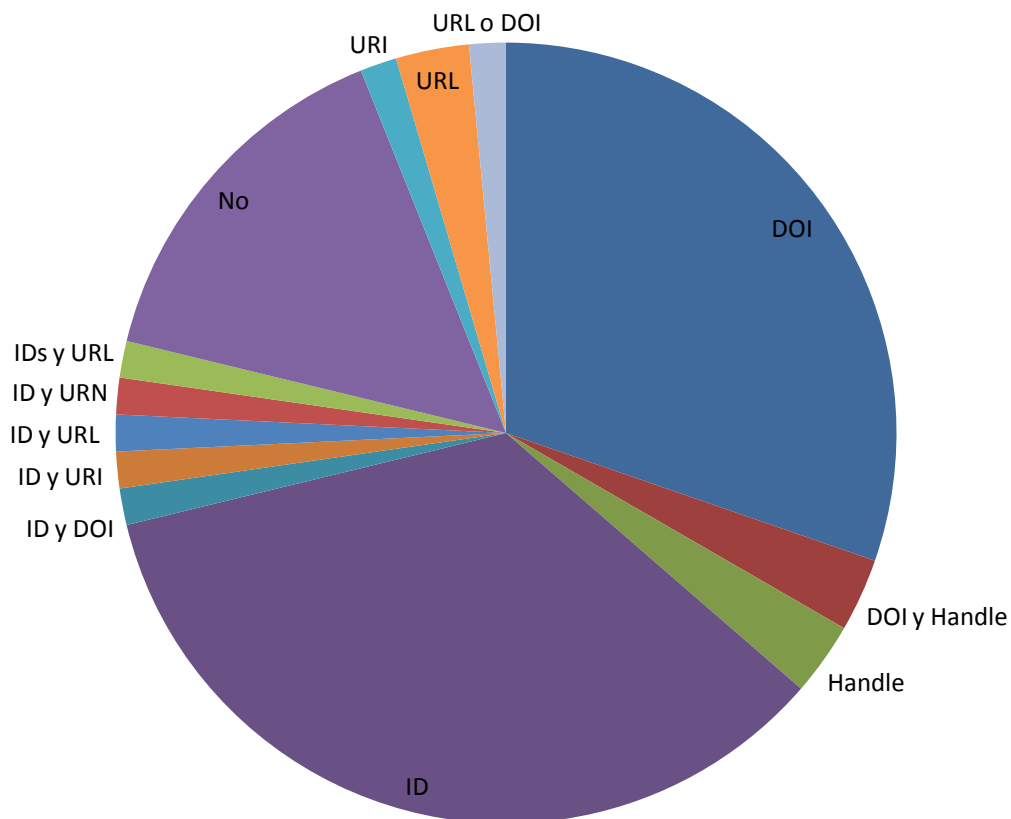


# Referencia bibliográfica del dataset

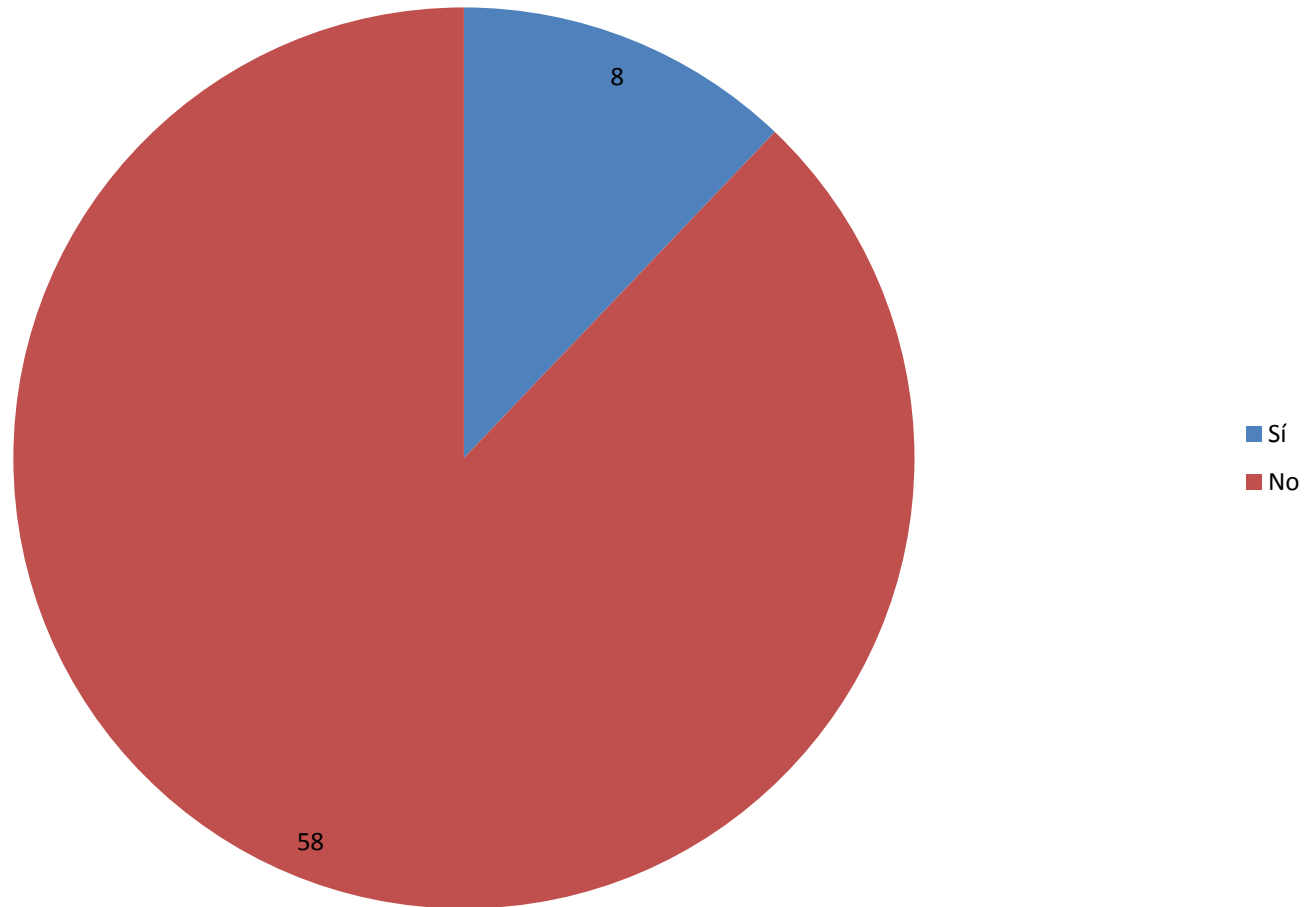




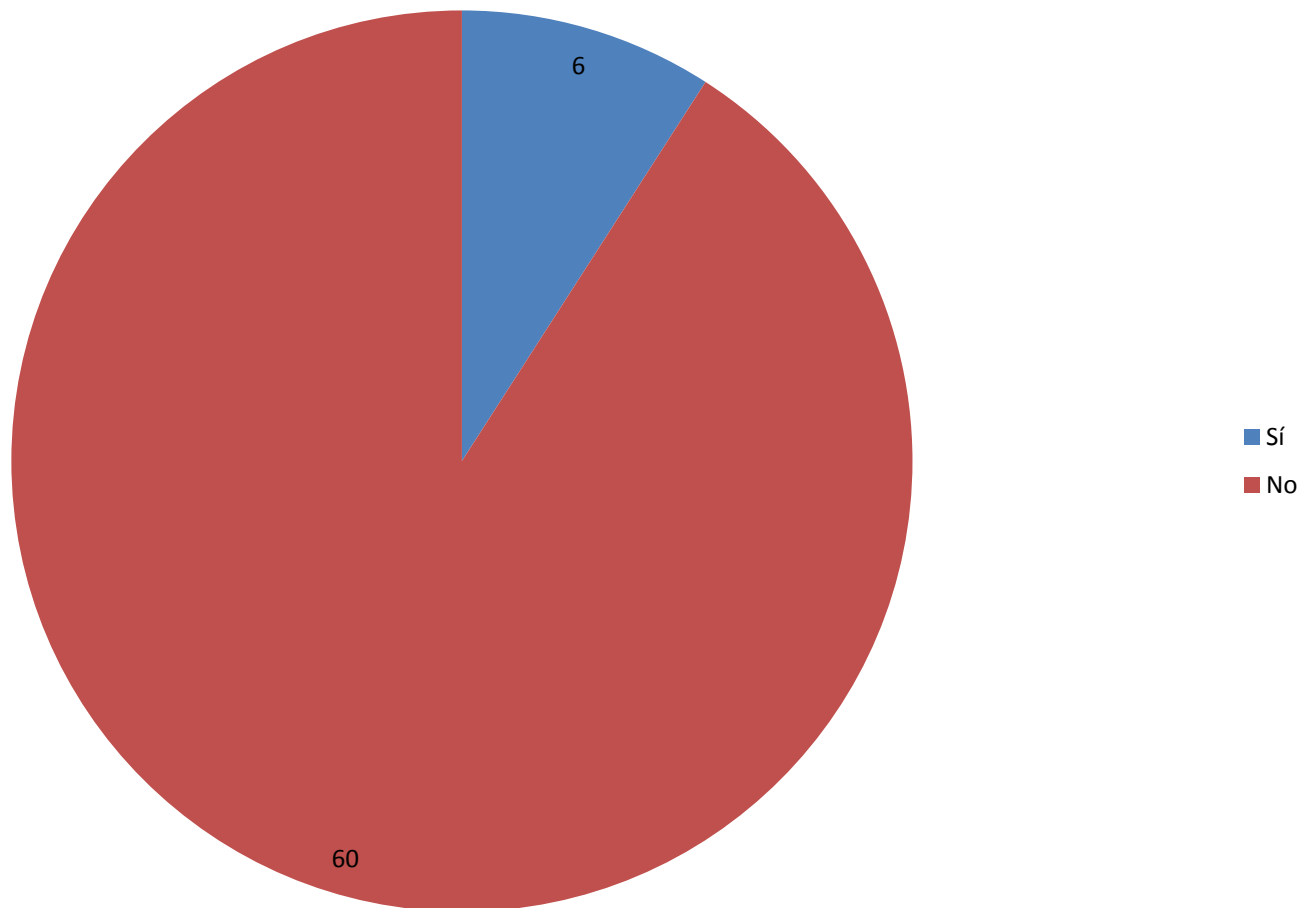
# Identificador del dataset



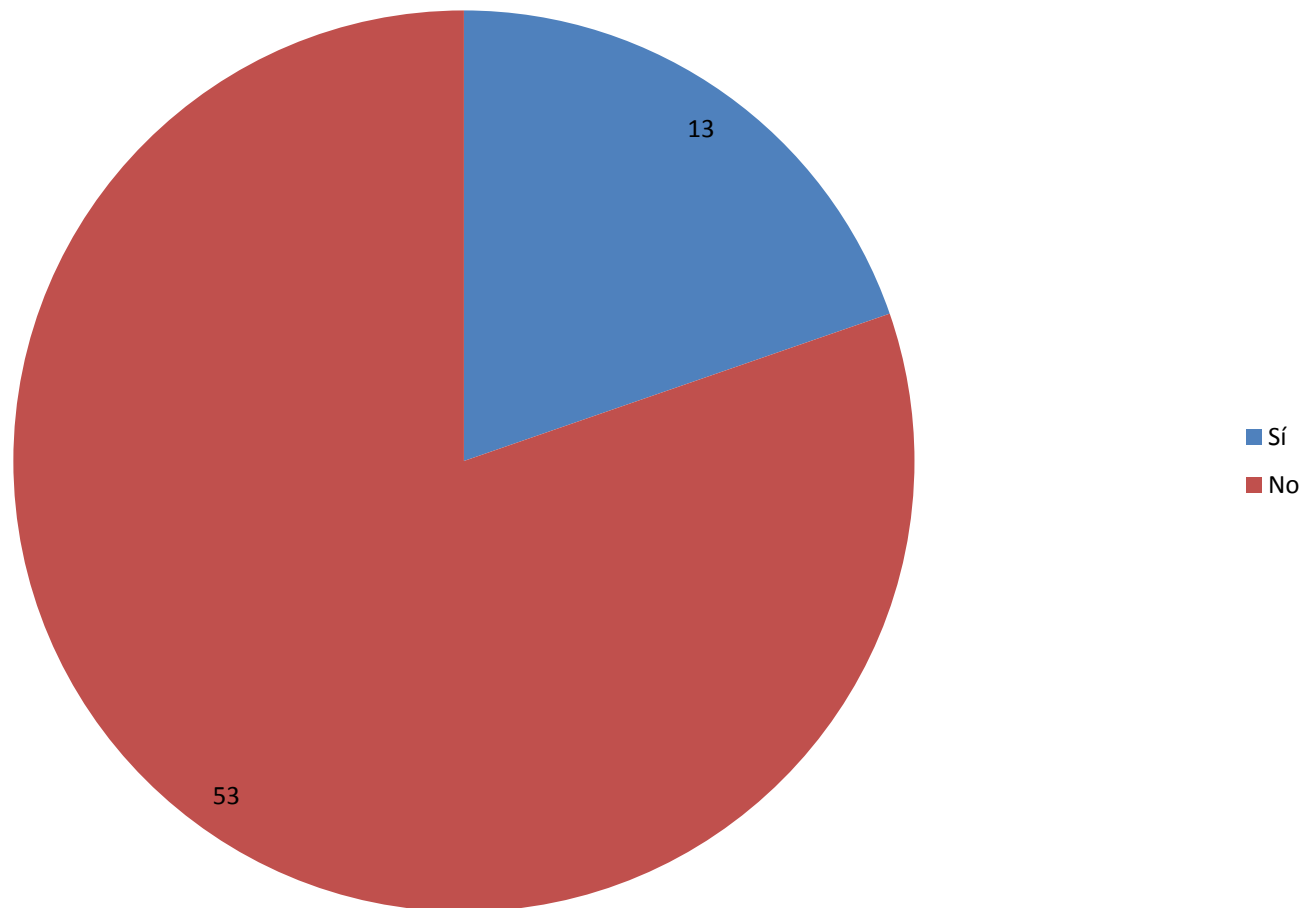
# Visualizaciones de página



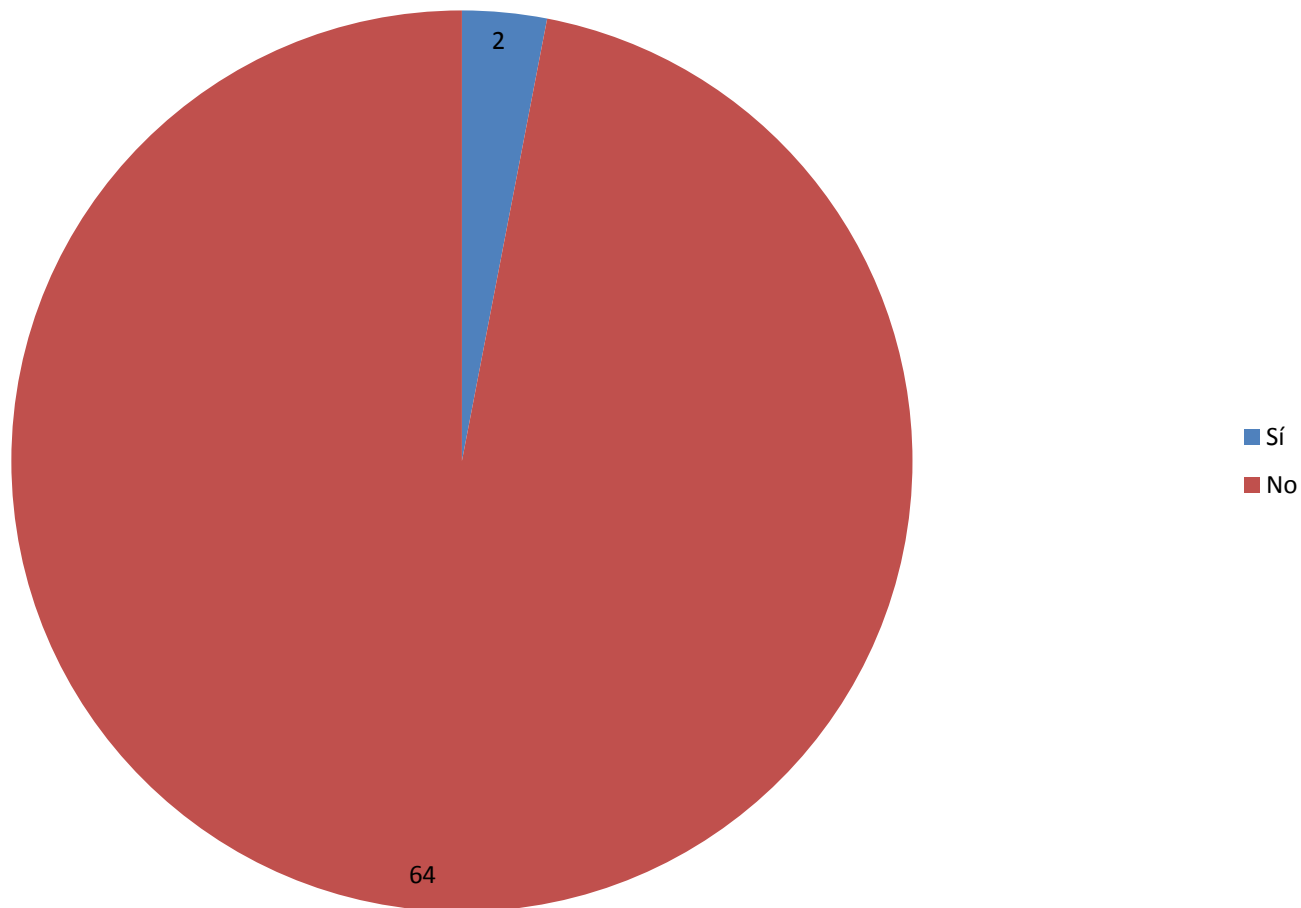
# Número de descargas



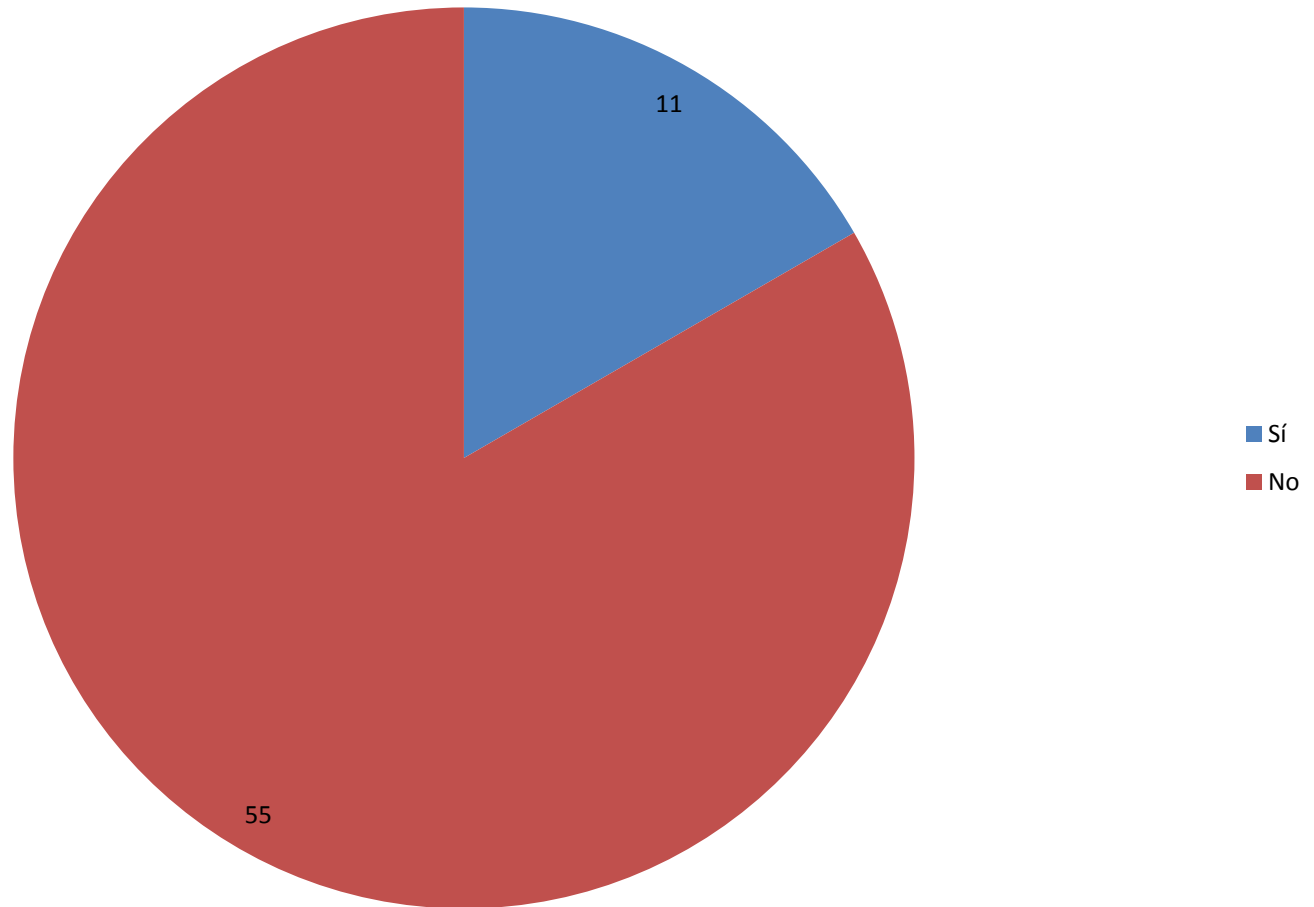
# Difusión en RRSS



# Veces que se difunde el dataset



# Difusión en GGGB



# Conclusiones

- Problemas de las métricas tradicionales
- Aparición de las métricas alternativas
- Análisis de repositorios
- Gran cantidad de trabajo para los repositorios
- Abanico de información
- Las “altmetrics”: un complemento