Entre Pares 2018

DATOS Y LOS PRINCIPIOS FAIR

Delfino García Alonso
Shirley Alquicira Hernández

Centro de Ciencias Genómicas-UNAM

CCG
Centro de Ciencias Genómicas
Licenciatura en
CIENCIAS
GENÓMICAS
SoIBio

Contenido

Copies and Copyright-Notice

User is not entitled to change or erase data sets of the RegulonDB database or to eliminate copyright notices from RegulonDB. Furthermore, User is not entitled to expand RegulonDB or to integrate RegulonDB partly or as a whole into other databank systems, without prior written consent from CCG-UNAM.

Citation

User is committed to cite properly the use of RegulonDB. The current most update publication concerning RegulonDB is:
Socorro Gama-Castro, Verónica Jiménez-Jacinto, Martin Peralta-Gil, Alberto Santos-Zavaleta, Monica I. Peñaloza-Spindola, Bruno Contreras-Moreira,
Juan Segura-Salazar, Luis Muñiz-Rascado, Irma Martínez-Flores, Heladia Salgado,
César Bonavides-Martínez, Cei Abreu-Goodger, Carlos Rodriguez-Penagos, Juan Miranda-Rios,
Enrique Morett, Enrique Merino, Araceli M. Huerta, Luis Treviño-Quintanilla and Julio Collado-Vides.
"RegulonDB (Version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation"
Nucleic Acids Research, 2008, vol 36. D120-D124

Release: 6.4 Date: 10-AUG-09

RegulonDB

E. coli (K12) Transcriptional Network

Transcription Factor Matrix and Alignments

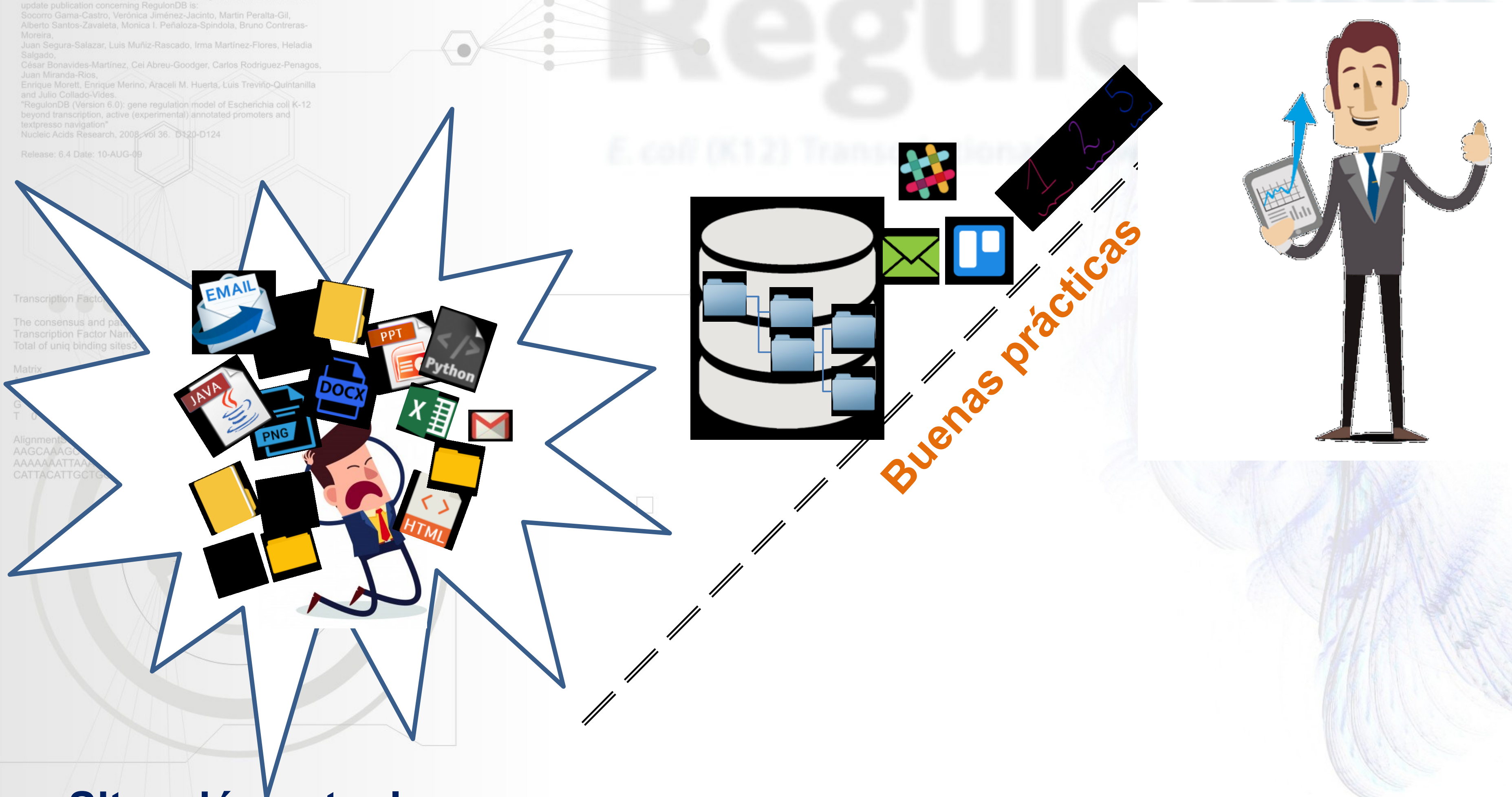
The consensus and patser programs were used to create the matrix and alignment
Transcription Factor NameAda
Total of uniq binding sites3

Matrix
A 2 3 1 1 3 2 3 0 0 1 1 2 0 0 1 0 2 2 0 3 2 0 1 1 0 0 0 1 0
C 1 0 0 1 0 1 0 0 1 0 2 0 0 2 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0
G 0 0 1 0 0 0 0 1 0 2 0 0 3 1 2 0 0 0 3 0 0 0 2 0 0 2 1 0 1
T 0 0 1 1 0 0 0 2 2 0 0 1 0 0 0 2 0 1 0 0 1 3 0 2 2 1 2 2 2

AlignmentScore
AAGCAAAGCGCAGCGTCTGAATAACGTTT20.66
AAAAAAATTAAAGCGCAAGATTGTTGGTT21.42
CATTACATTGCTGGATAAGAATGTTTTAG19.78

Antes de aplicar los principios FAIR, el primer paso es....

Buena organización y uso de los Principios FAIR



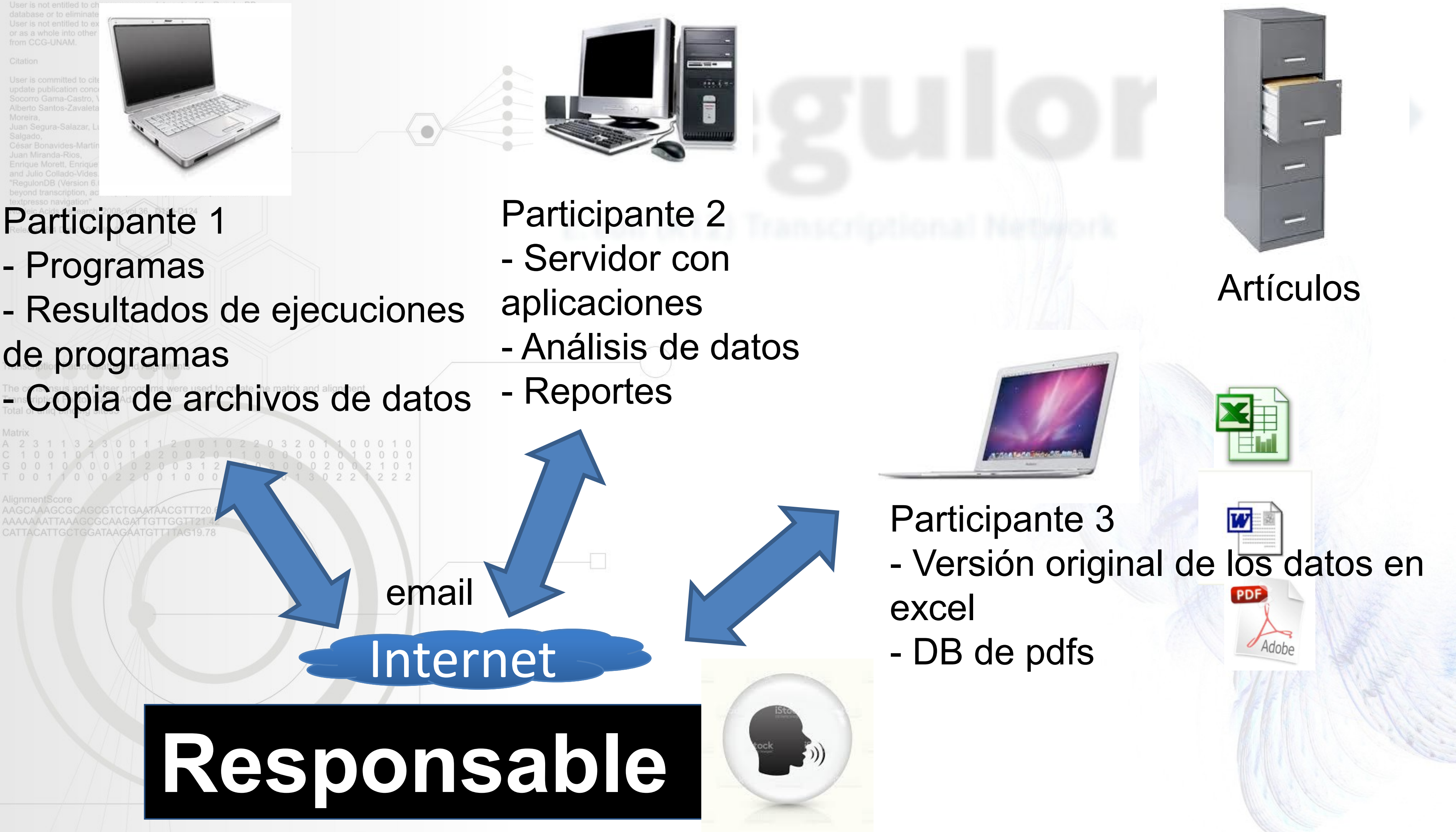
Buenas prácticas

Situación actual

Trabajando de manera personal

En el trabajo diario manejamos demasiada información, y en la mayoría de las ocasiones se vuelve complicado tener un orden de la misma.

Cuando trabajamos en un proyecto colaborativo…



Participante 1
- Programas
- Resultados de ejecuciones de programas
- Copia de archivos de datos

Participante 2
- Servidor con aplicaciones
- Análisis de datos
- Reportes

Artículos

Participante 3
- Versión original de los datos en excel
- DB de pdfs

Responsable

- Reportes

Integrando a nuevos participantes al equipo

Jefe

Los que quedan
les toca integrar al nuevo
miembro del equipo:

Biólogo
Programador
Analista
Estudiante

Nuevo integrante

Discutiendo el proyecto

¡Muy interesante por cierto!

¿Dónde están los programas
¿Dónde están los datos?
¿A quién le reporto?
¿Y qué reporto?

Retrasos en entregas
Trabajos a la carrera
Pérdida de información
Estrés, frustración, enojo, etc…

MEJORES PRÁCTICAS

Base de Conocimiento (BC)

Modelo de Procesos para la Industria del Software en México, desarrollado a solicitud de la Secretaria de Economía en convenio con la Facultad de Ciencias de la UNAM

BC: BASE DE CONOCIMIENTO

Repositorio de **todos los productos de trabajo** resultado de los proyectos: software, planes, reportes, registros, lecciones aprendidas, tesis, artículos, plantillas, seminarios, datos, imágenes, etc.*

*NYCE, Norma Mexicana NMX-I-059/01-NYCE-2005, 2005. Definición de Conceptos y Productos

BC: BASE DE CONOCIMIENTO

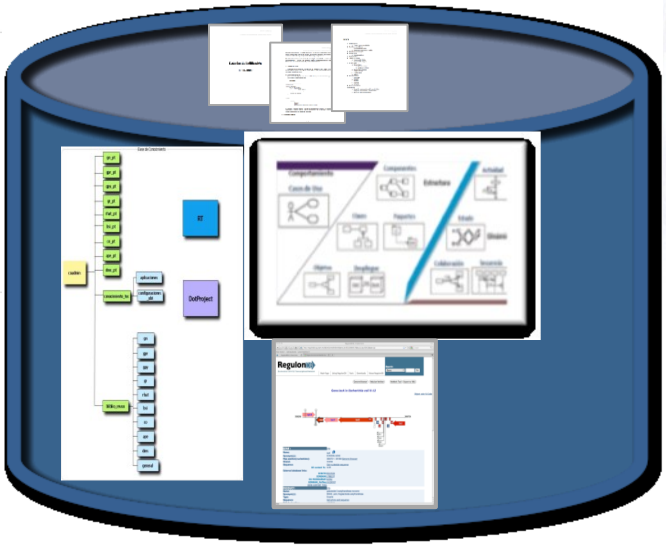
PROYECTO

- Artículos
- Software
- Procedimientos
- Investigación
- Presentaciones
- Docencia

Colaborador

Alimentación/ poblado

¿Dónde están los programas?
¿Dónde están los datos?
¿A quién le reporto?
¿Y qué reporto?



Investigador

Consulta

Administrador

Mantenimiento
Respaldo

Sugerencia de herramientas para crear una BC

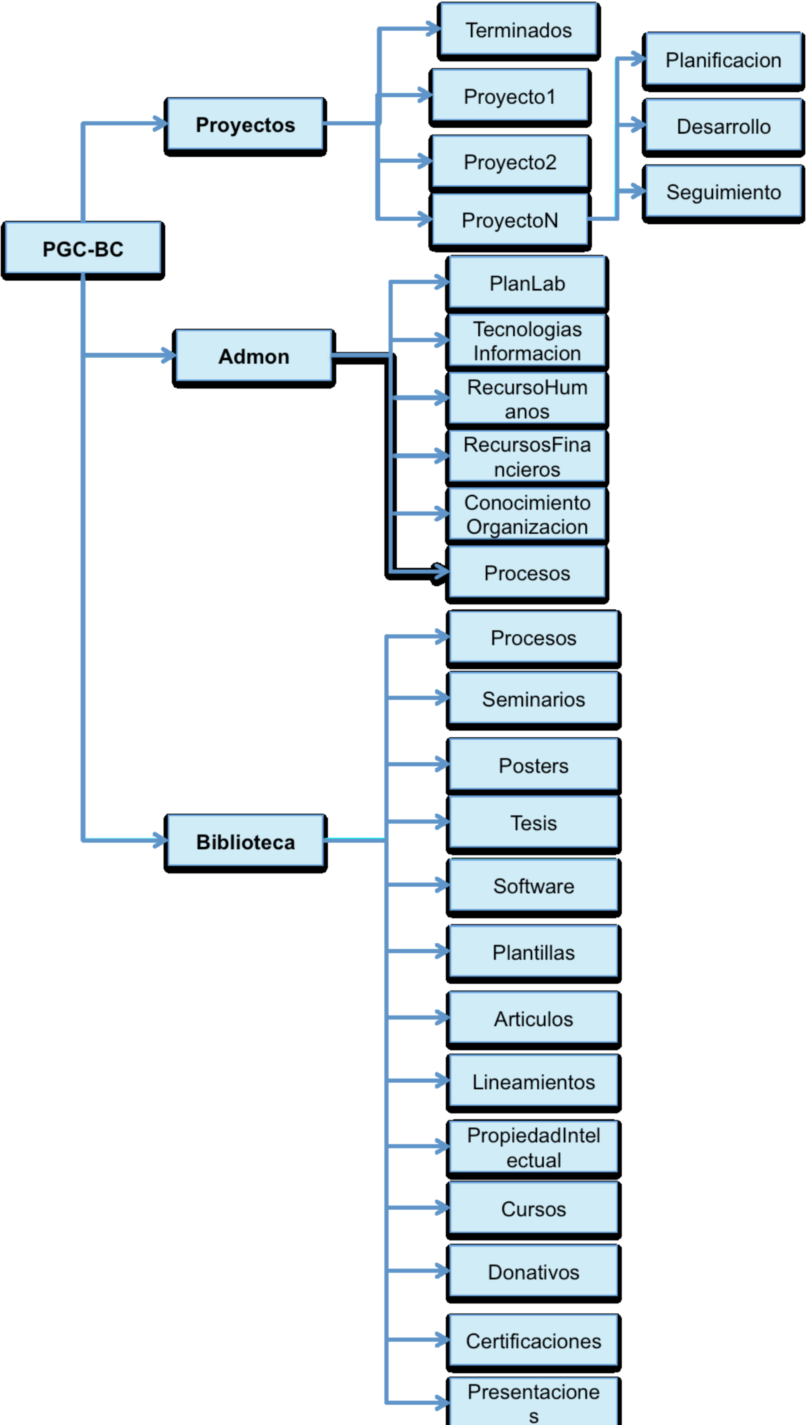
Dropbox

Google Drive

OneDrive

- Almacenamiento de archivos y carpetas
- Compartir de archivos y carpetas
- Control de cambios y versiones
- Recuperación de archivos/carpetas eliminadas
- Manejo de usuarios
- Disponibilidad en dispositivos móviles
- Multi-plataforma
- **Trabajo colaborativo** permite la edición de múltiples usuarios sobre un mismo documento

Ejemplo

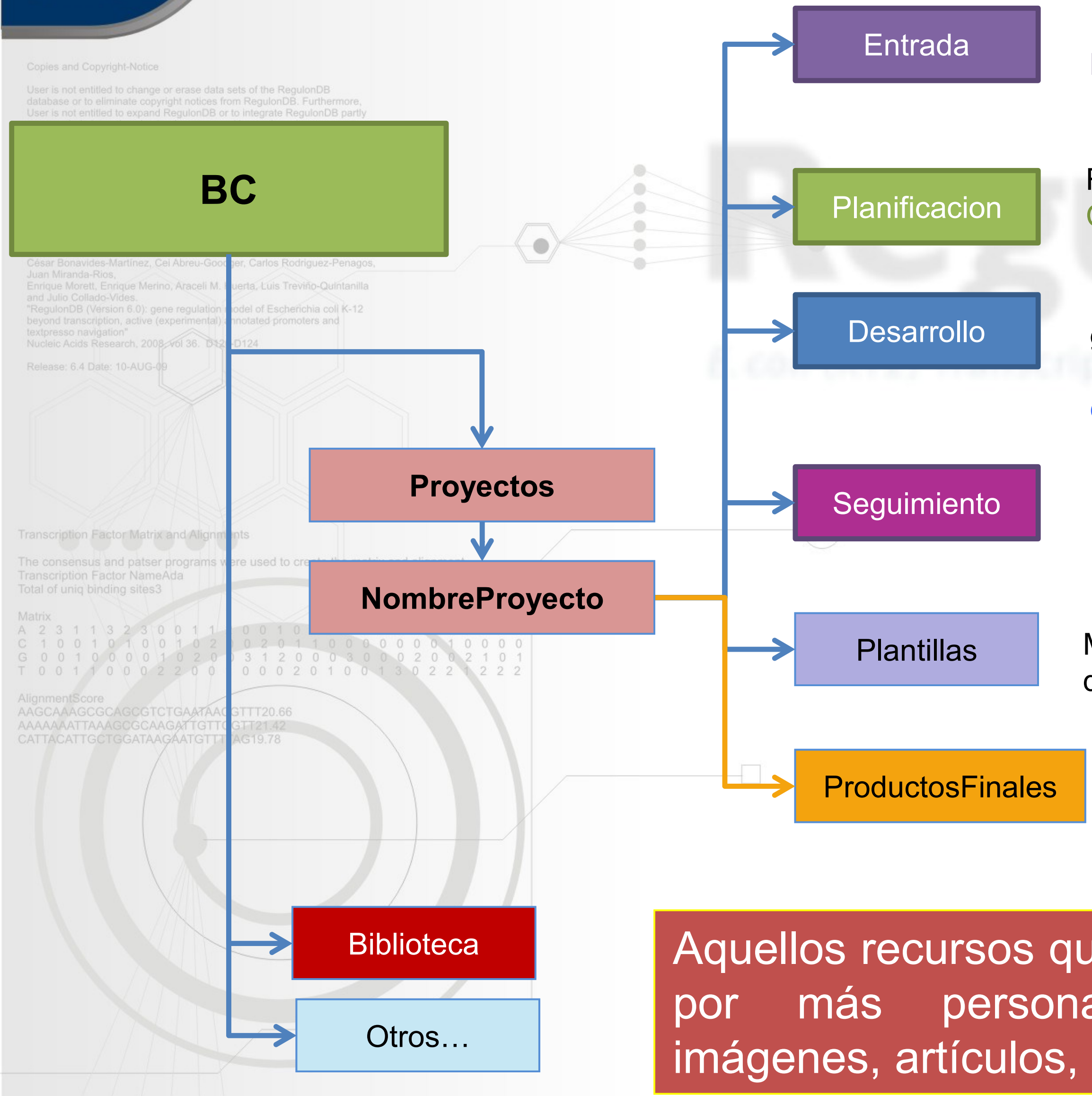


- PGC-BC
  - Proyectos
    - Terminados
    - Proyecto1
    - Proyecto2
    - ProyectoN
      - Planificacion
      - Desarrollo
      - Seguimiento
  - Admon
    - PlanLab
    - Tecnologias Informacion
    - RecursoHumanos
    - RecursosFinancieros
    - Conocimiento Organizacion
    - Procesos
  - Biblioteca
    - Procesos
    - Seminarios
    - Posters
    - Tesis
    - Software
    - Plantillas
    - Articulos
    - Lineamientos
    - PropiedadIntelectual
    - Cursos
    - Donativos
    - Certificaciones
    - Presentaciones

Directorios de trabajo en una BC

Administrador

Mi carpeta de trabajo



Entrada: Datos fuente de Bases de datos, documentación exter programas de otros, datos, etc..

Planificacion: Repositorio para el resguardo de planes de, Cronograma, Descripción del proyecto

Desarrollo: Repositorio para el resguardo de toda la información generada en la ejecución del proyecto: Documentación, Diagramas, Programas, Manuales, etc.

Seguimiento: Repositorio para el resguardo de reportes sobre el estado del proyecto., Minutas

Plantillas: Material guía para la elaboración de la documentación del proyecto.

ProductosFinales: Versiones finales de TODO, en el cierre

Biblioteca: Aquellos recursos que pueden ser reutilizados por más personas: plantillas, posters, imágenes, artículos, programas, etc.

Herramientas de comunicación

Correo electrónico: Cuenta institucional, gmail, hotmail, etc.

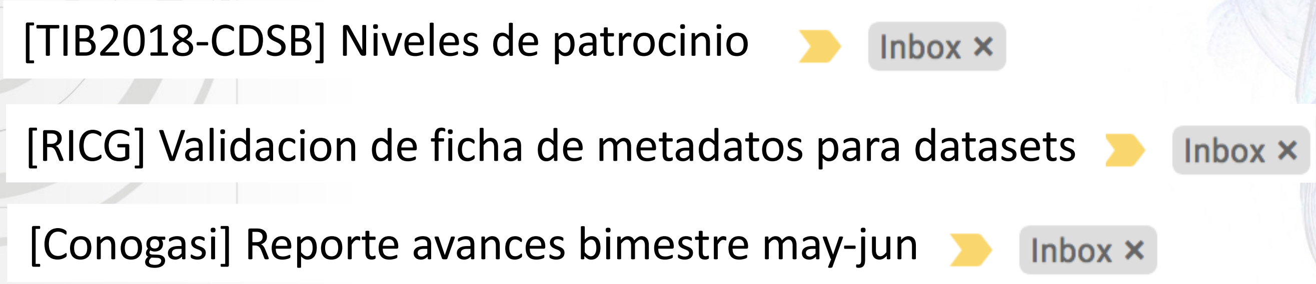
“Asunto” en correos electrónicos

Usar **palabras clave** para localizar su búsqueda y localización.

Asignar un “**identificador**” o “**nombre corto**” del proyecto, el cual es utilizado para localizar todos los correos relacionados al proyecto.

Ejemplo:

[NombreCortoProyecto] Palabras clave



[TIB2018-CDSB] Niveles de patrocinio

[RICG] Validacion de ficha de metadatos para datasets

[Conogasi] Reporte avances bimestre may-jun

Herramientas de comunicación

Trello: Herramienta para la administración de proyectos. Representa un **tablero** virtual en el que se pueden **definir** ideas, **tareas**, imágenes o enlaces.

Slack: Herramienta de comunicación en equipo, ofrece canales de chat organizados por temas, así como grupos privados y mensajes directos.

Versionamiento

Versionamiento

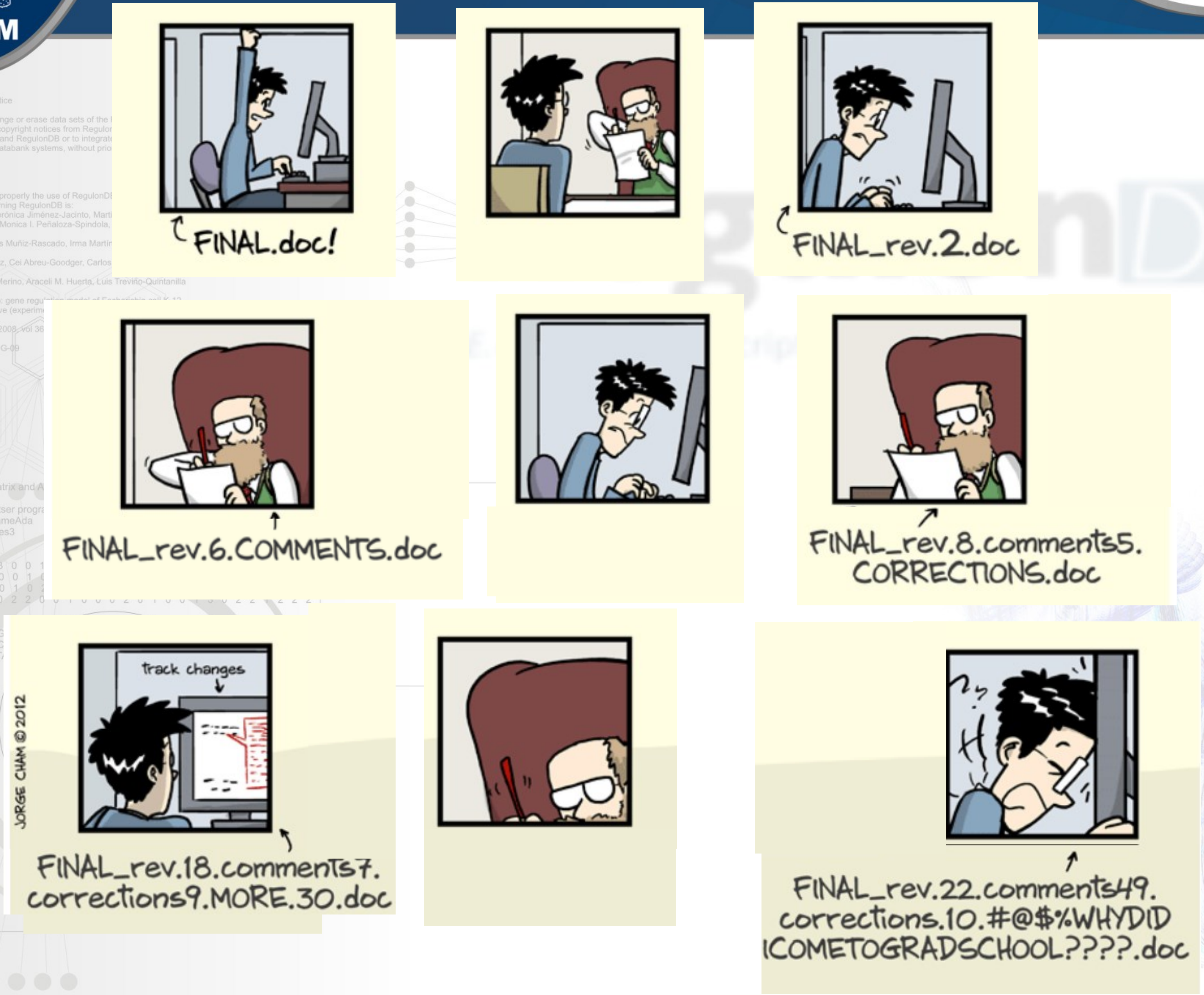
Contar con un *control de versiones* ayuda a gestionar y manejar los cambios realizados en un producto de trabajo.

¿Por qué es útil?

La razón por la cual el control de versiones es universal es porque **ayuda** en todos los aspectos al **dirigir un proyecto**:

- Comunicación entre los participantes del proyecto, todos los participantes saben a que documento se refieren
- Control entre aquellos documentos liberados y los nuevos cambios en la documentación
- Identificación de los cambios realizados por los participantes del proyecto

"FINAL".doc


FINAL.doc!
FINAL_rev.2.doc
FINAL_rev.6.COMMENTS.doc
FINAL_rev.8.comments5.
CORRECTIONS.doc
track changes
JORGE CHAM © 2012
FINAL_rev.18.comments7.
corrections9.MORE.30.doc
FINAL_rev.22.comments49.
corrections.10.#@$%WHYDID
ICOMETOGRADSCHOOL????.doc

Ejemplos

Proceso_Extraccion_datos_**v0.5.ppt**

Expo_Admon_Proyectos_**v2.0.ppt**

Analisis_Experimento1_**v1.0.xls**

Uso de metadatos

- Los **metadatos** ayudan a **describir** el contenido de un recurso (**archivo de datos/literatura**). El uso de metadatos:
  - Facilita la localización y posicionamiento de los recursos en internet
  - Nos ayuda a que nuestros datos puedan ser referenciados por otros
  - Que otras personas puedan acceder a nuestros datos y puedan trabajar con ellos

Existen normas y estándares internacionales* dedicados a promover el uso e intercambio de información por medio del uso de metadatos.

**Dublin Core, OpenAIRE, Lineamientos FAIR*

Ejemplo

- Antes y después del uso de los metadatos

En resumen...

- Buenas práticas aplicadas a mi proyecto y ambiente de trabajo....
  - Tengo una Base de Conocimiento
  - Uso adecuadamente herramientas de comunicación
  - Versiono mis datos y archivos

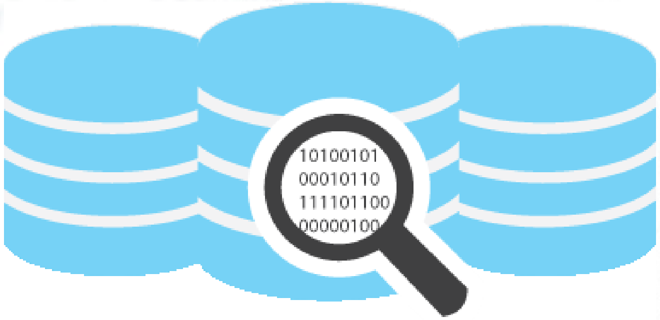
Ahora que he terminado mi proyecto (investigación) y que tengo datos y resultados finales...

¿Qué tengo que hacer para publicar mis datos/resultados?

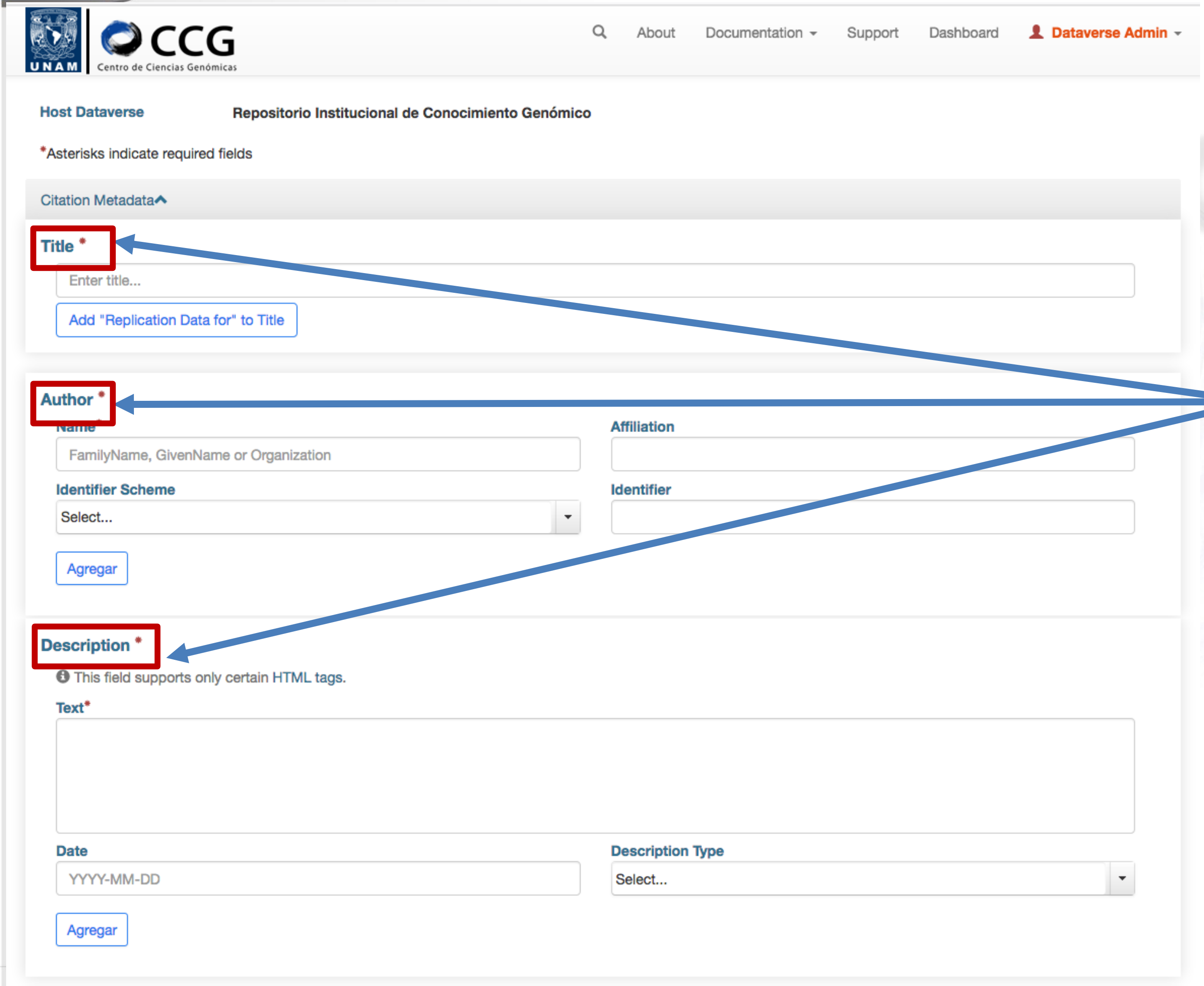
¿Qué características deben tener mis datos/resultados para que puedan ser utilizados por los demás?

Principio: FINDABLE (F)

- Selecciona algún repositorio para colocar mis datos
- Asegurarme de que puedan asignar un **IDENTIFICADOR** (doi, handle, purl, etc) a mis datos
- Describir mis datos por medio de **METADATOS**



Ejemplo: Principio-FINDABLE (F)



Host Dataverse Repositorio Institucional de Conocimiento Genómico

*Asterisks indicate required fields

Citation Metadata

Title *

Enter title...

Author *

Name	Affiliation
FamilyName, GivenName or Organization	
Identifier Scheme	Identifier
Select...	

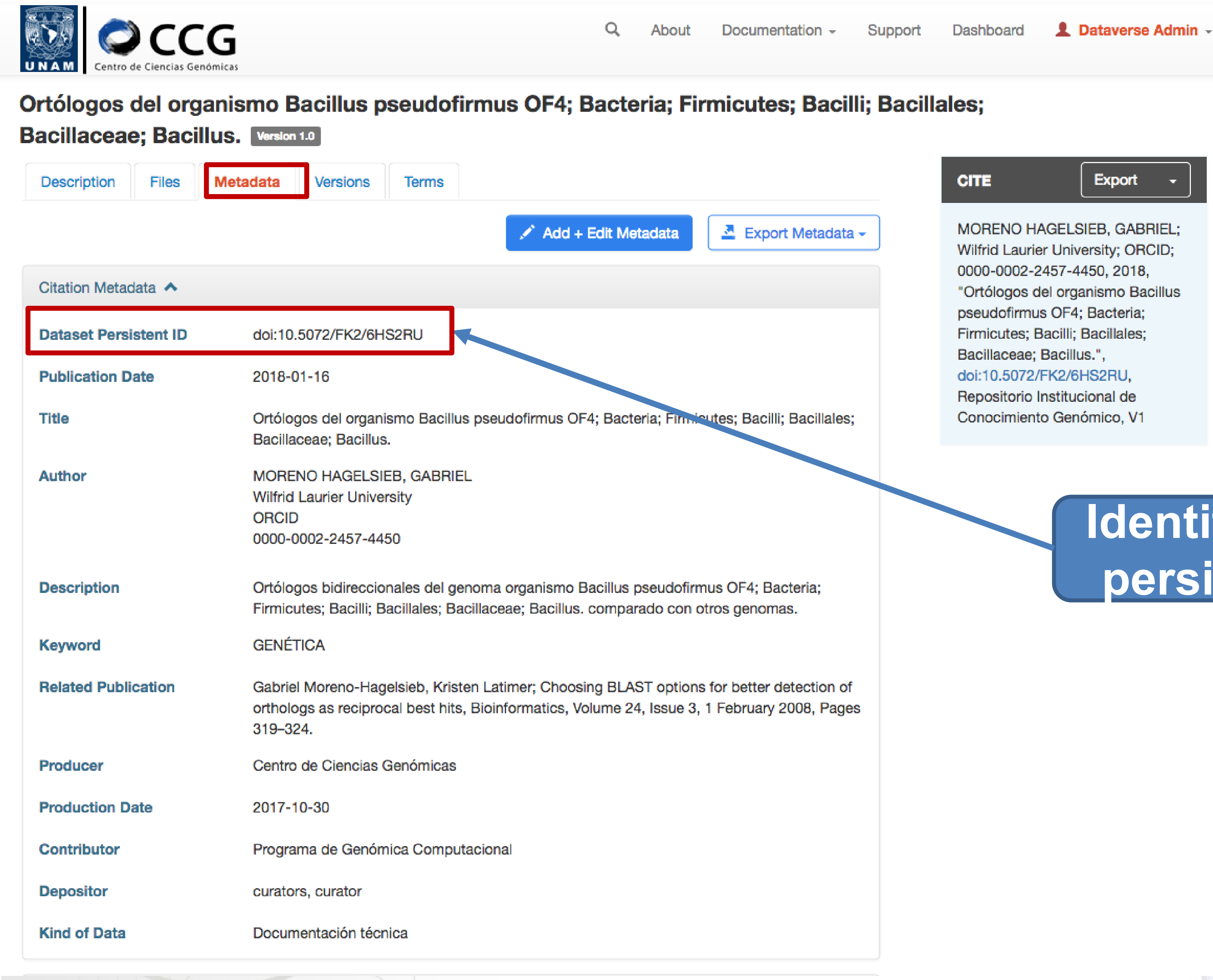
Description *

This field supports only certain HTML tags.

Text*

Date	Description Type
YYYY-MM-DD	Select...

Metadatos Dublin Core



Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus. Version 1.0

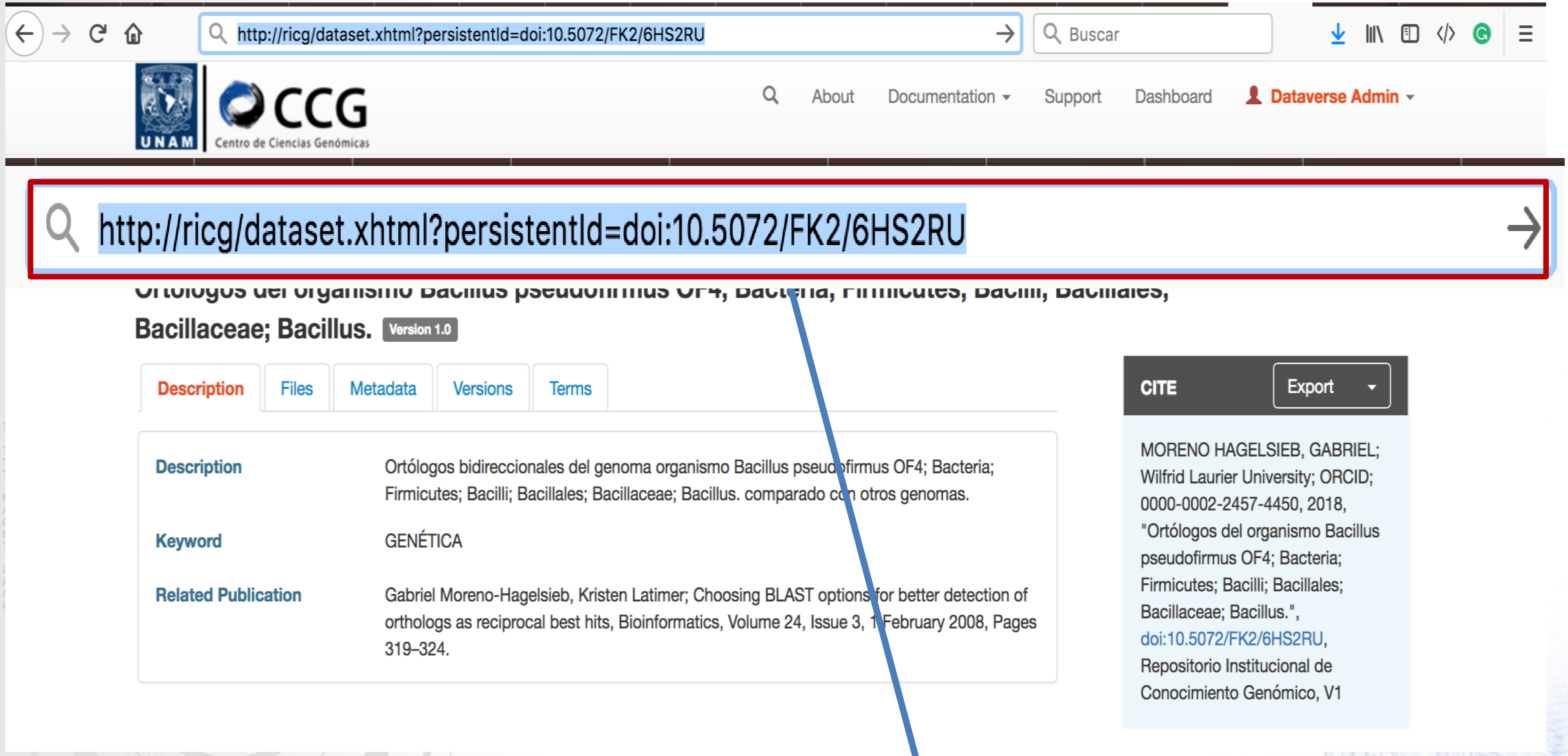
Citation Metadata

Dataset Persistent ID	doi:10.5072/FK2/6HS2RU
Publication Date	2018-01-16
Title	Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.
Author	MORENO HAGELSIEB, GABRIEL Wilfrid Laurier University ORCID 0000-0002-2457-4450
Description	Ortólogos bidireccionales del genoma organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus. comparado con otros genomas.
Keyword	GENÉTICA
Related Publication	Gabriel Moreno-Hagelsieb, Kristen Latimer; Choosing BLAST options for better detection of orthologs as reciprocal best hits, Bioinformatics, Volume 24, Issue 3, 1 February 2008, Pages 319–324.
Producer	Centro de Ciencias Genómicas
Production Date	2017-10-30
Contributor	Programa de Genómica Computacional
Depositor	curators, curator
Kind of Data	Documentación técnica

MORENO HAGELSIEB, GABRIEL; Wilfrid Laurier University; ORCID; 0000-0002-2457-4450, 2018, "Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.", doi:10.5072/FK2/6HS2RU, Repositorio Institucional de Conocimiento Genómico, V1

Identificador persistente

Ejemplo: Principio-Accesible (A)



http://ricg/dataset.xhtml?persistentId=doi:10.5072/FK2/6HS2RU

Ortólogos del organismo Bacillus pseudofirmus OF4, Bacteria, Firmicutes, Bacilli, Bacillales, Bacillaceae; Bacillus. Version 1.0

Description	Ortólogos bidireccionales del genoma organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus. comparado con otros genomas.
Keyword	GENÉTICA
Related Publication	Gabriel Moreno-Hagelsieb, Kristen Latimer; Choosing BLAST options for better detection of orthologs as reciprocal best hits, Bioinformatics, Volume 24, Issue 3, 1 February 2008, Pages 319–324.

MORENO HAGELSIEB, GABRIEL; Wilfrid Laurier University; ORCID; 0000-0002-2457-4450, 2018, "Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.", doi:10.5072/FK2/6HS2RU, Repositorio Institucional de Conocimiento Genómico, V1

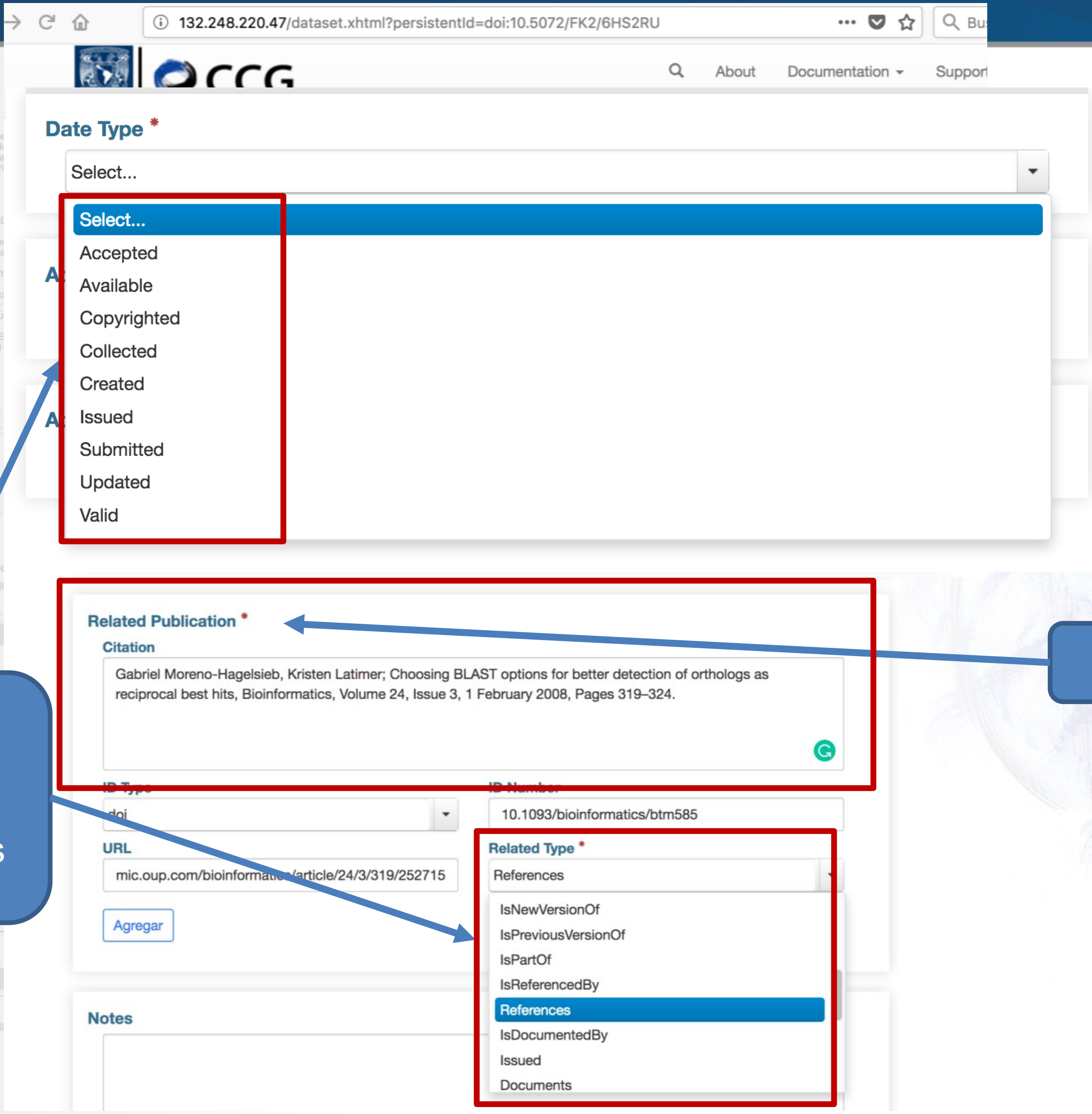
[redacted]

Internet

Protocolo de comunicación

(HTTP (S) FTP SMTP)

Ejemplo: Principio-Interoperable (I)



Date Type *

Select...

Related Publication *

Citation

Gabriel Moreno-Hagelsieb, Kristen Latimer; Choosing BLAST options for better detection of orthologs as reciprocal best hits, Bioinformatics, Volume 24, Issue 3, 1 February 2008, Pages 319–324.

ID Type: doi

ID Number: 10.1093/bioinformatics/btm585

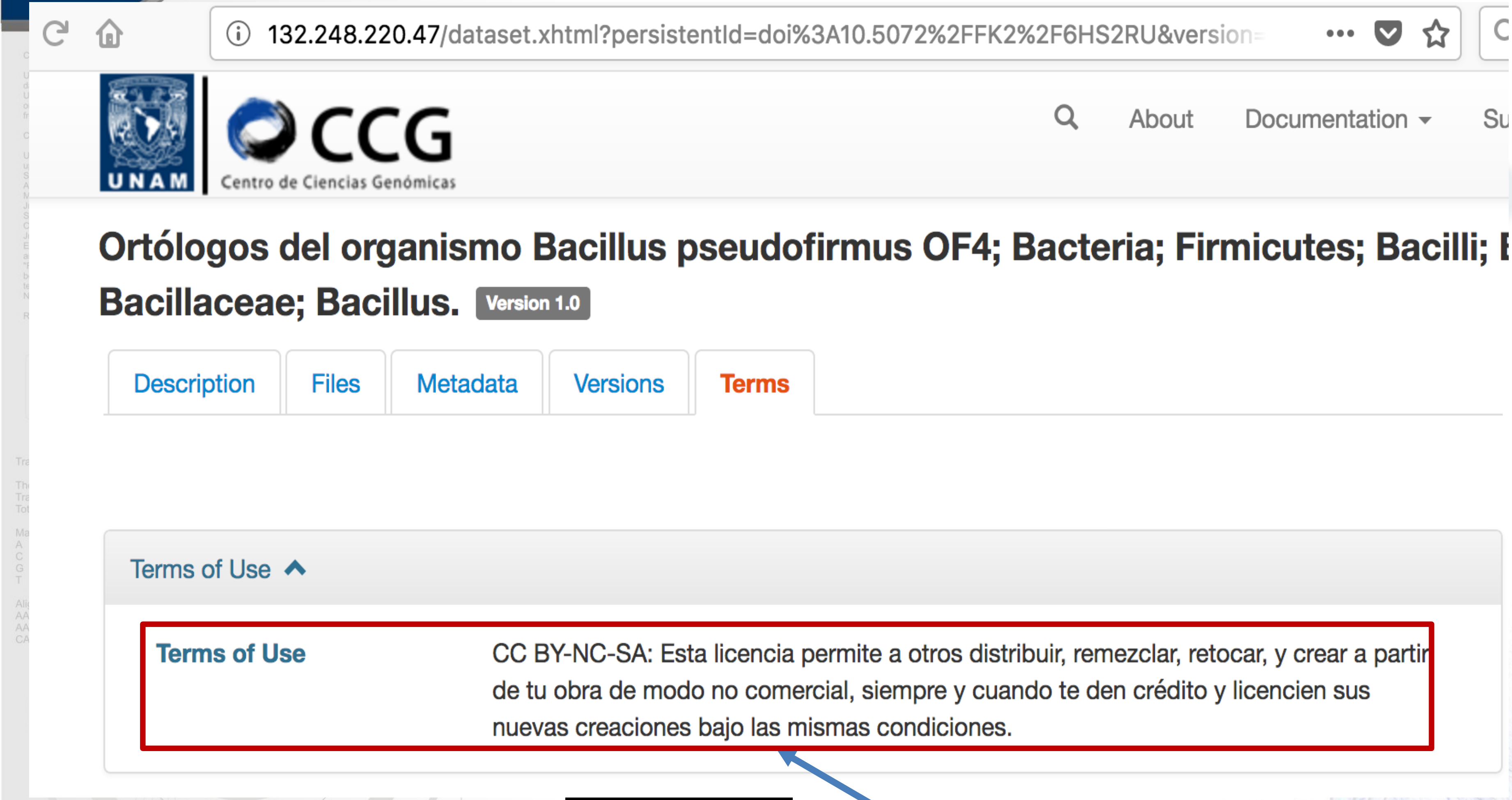
URL: mic.oup.com/bioinformatics/article/24/3/319/252715

Related Type *: References

Referencias

Vocabulario controlado
Ontologías
Diccionarios

Ejemplo: Principio-Reusable (R)



Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli;
Bacillaceae; Bacillus. Version 1.0

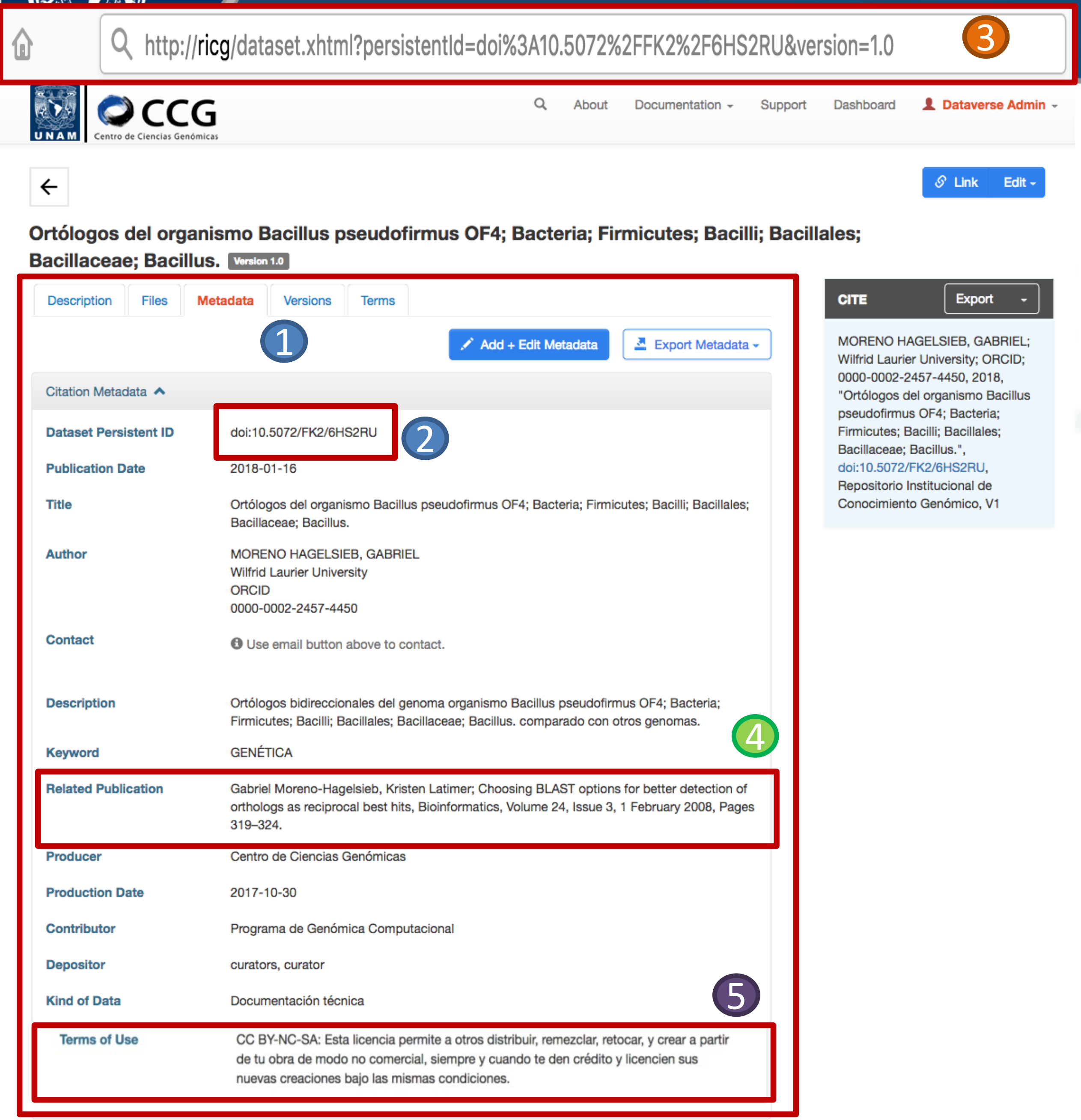
Terms of Use	CC BY-NC-SA: Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de tu obra de modo no comercial, siempre y cuando te den crédito y licencien sus nuevas creaciones bajo las mismas condiciones.

[redacted]

Monetising the access to new and existing research results is profoundly at odds with the ethos of science (Merton, 1973).

Licencia de uso
de los datos

Al final tenemos!!!



http://ricg/dataset.xhtml?persistentId=doi%3A10.5072%2FFK2%2F6HS2RU&version=1.0 ③

Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus. Version 1.0

Citation Metadata

Dataset Persistent ID	doi:10.5072/FK2/6HS2RU ②
Publication Date	2018-01-16
Title	Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.
Author	MORENO HAGELSIEB, GABRIEL Wilfrid Laurier University ORCID 0000-0002-2457-4450
Contact	Use email button above to contact.
Description	Ortólogos bidireccionales del genoma organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus. comparado con otros genomas. ④
Keyword	GENÉTICA
Related Publication	Gabriel Moreno-Hagelsieb, Kristen Latimer; Choosing BLAST options for better detection of orthologs as reciprocal best hits, Bioinformatics, Volume 24, Issue 3, 1 February 2008, Pages 319–324.
Producer	Centro de Ciencias Genómicas
Production Date	2017-10-30
Contributor	Programa de Genómica Computacional
Depositor	curators, curator
Kind of Data	Documentación técnica ⑤
Terms of Use	CC BY-NC-SA: Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de tu obra de modo no comercial, siempre y cuando te den crédito y licencien sus nuevas creaciones bajo las mismas condiciones.

CITE Export

MORENO HAGELSIEB, GABRIEL; Wilfrid Laurier University; ORCID; 0000-0002-2457-4450, 2018, "Ortólogos del organismo Bacillus pseudofirmus OF4; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus.", doi:10.5072/FK2/6HS2RU, Repositorio Institucional de Conocimiento Genómico, V1

1. Metadatos	F
2. Identificador	
3. Protocolo de comunicación	A
4. Referencias	I
5. Licencia	R

*Only results that can be discussed, challenged, and, where appropriate, tested, and reproduced by others qualify as **scientific**.*

*The chain, whereby **new scientific discoveries are built on previously established results**, can only work optimally if all research **results are made openly available** to the scientific community.*

*In the 21st century, **science publishers should provide a service to help researchers disseminate their results.** They may be paid fair value for the services they are providing, but no science should be locked behind paywalls!*

Science Without Publication Paywalls: cOAlition S for the Realisation of Full and Immediate Open Access: https://doi.org/10.3389/fnins.2018.00656

Only results that can be discussed, challenged, and, where appropriate, tested, and reproduced by others qualify as scientific.

The chain, whereby new scientific discoveries are built on previously established results, can only work optimally if all research results are made openly available to the scientific community.

Monetising the access to new and existing research results is profoundly at odds with the ethos of science (Merton, 1973).

In the 21st century, science publishers should provide a service to help researchers disseminate their results. They may be paid fair value for the services they are providing, but no science should be locked behind paywalls!