

Progresos recientes en sistemas de gestión de la información científica

Recent progress in research information management and its systems

Pablo De-Castro

De-Castro, Pablo (2019). "Progresos recientes en sistemas de gestión de la información científica". *Anuario ThinkEPI*, v. 13, e13e04.

<https://doi.org/10.3145/thinkepi.2019.e13e04>

Publicado en *IweTel* el 1 de julio de 2019

Pablo De-Castro

<https://orcid.org/0000-0001-6300-1033>

University of Strathclyde & euroCRIS

pcastrmartin@gmail.com



Abstract: Se resumen algunos de los avances más recientes en el área de la gestión de la información científica y sistemas CRIS. Se hace énfasis en el creciente grado de interoperabilidad de sistemas facilitado por la implantación de las directrices *CERIF-XML* para gestores de sistemas CRIS conjuntamente publicadas por *OpenAIRE* y *euroCRIS*. Se describe una serie de mejoras que la incorporación de los sistemas CRIS a la lista de proveedores de datos para *OpenAIRE* puede suponer para un más profundo conocimiento de la investigación que se está llevando a cabo en Europa.

Palabras clave: Sistemas de gestión de la información científica; Sistemas CRIS; Informes de investigación; Colaboración internacional.

Abstract: This piece summarises some of the most recent advances in the area of research information management and Current Research Information Systems (CRIS). With a focus on the increasing level of system interoperability that the implementation of the *CERIF-XML* Guidelines for CRIS Managers jointly issued by *OpenAIRE* and *euroCRIS* is making possible, the update lists a number of improvements that adding CRIS systems to the list of *OpenAIRE* data providers may deliver to the attempt to capture a more comprehensive snapshot of the research being conducted in Europe.

Keywords: Research information management systems; CRIS; Research reporting; International collaboration.

1. Introducción

A pesar de gestarse casi en secreto para los no hispanohablantes, vienen llegando regularmente buenas noticias desde el Sur de Europa y en particular desde España en relación con los sistemas de gestión de la información científica o CRIS (*current research information systems*). La escasez de traducciones al inglés y la baja participación de instituciones públicas en los eventos dedicados a debatir los avances en la disciplina hacen que sea difícil percibir los progresos en el área desde un entorno internacional, pero las iniciativas están ahí y hay ocasionales menciones en listas y eventos de ámbito nacional que permiten un seguimiento más o menos detallado.

La nota *ThinkEPI* “Portales de producción científica” de Andoni Calderón-Rehecho, distribuida por *IweTel* en marzo de 2017 (Calderón-Rehecho, 2017) fue una muestra de que había interés por conocer el panorama de implantación de estos sistemas a nivel institucional en el país. En verano del año siguiente el subgrupo de acceso abierto de la línea 2 de *Rebiun* llevó a cabo una encuesta para obtener más detalles sobre la implantación, funcionalidad, interoperabilidad y roles profesionales asociados a los sistemas CRIS en las universidades españolas integrantes de esta red de bibliotecas (*Rebiun*, 2018a). Esta encuesta, que de manera interesante coincidió en el tiempo –y en buena medida, también en el ámbito– con la encuesta de nivel mundial sobre prácticas en el área de gestión de la información científica realizada conjuntamente por *OCLC Research* y *euroCRIS*, produjo un detallado informe del estado del arte de los sistemas CRIS en universidades españolas que se presentó en el *XVII Workshop Rebiun de Proyectos Digitales* de octubre de 2018 en Sevilla (*Rebiun*, 2018b).

En paralelo viene gestándose el proyecto *Hércules* de la *Crue* [4] con la intención de examinar la posibilidad de diseñar una arquitectura común que contribuya a armonizar la infraestructura CRIS a nivel nacional y que facilite el intercambio y la agregación de la información científica para su explotación en el ámbito de la toma de decisiones basadas en la evidencia.

Gran parte de la tradicional reticencia que desde el mundo de las bibliotecas ha existido en relación con estos sistemas CRIS se debe al elevado grado de ‘colonización’ de la disciplina por parte de plataformas comerciales en manos de bien conocidos proveedores no particularmente populares entre la comunidad de gestión de la información científica por sus prácticas en otros ámbitos. Esta divergencia se pone especialmente de manifiesto cuando se compara con la simultánea explosión de los repositorios institucionales de acceso abierto, tradicionalmente basados en plataformas de código abierto y en una intensa actividad de colaboración entre instituciones.

2. Más y mejor interoperabilidad de sistemas

El panorama está cambiando sin embargo, y el reciente acuerdo de colaboración entre *OpenAIRE* y *euroCRIS* (*EuroCRIS*, 2019a) es buena prueba de ello. Renunciar a la exploración de las indudables oportunidades que para las instituciones ofrece el área de la gestión de la información científica, únicamente porque la presencia de actores comerciales en el sector está muy extendida, es una tentación que potencialmente limita el crecimiento de las ambiciones del movimiento por la ciencia abierta.

El trabajo conjunto de *euroCRIS* y *OpenAIRE* en los últimos tiempos se basa en incrementar la interoperabilidad de los sistemas CRIS y en hacer técnicamente posible que la muy valiosa información que contienen pueda exportarse y agregarse a la extensa colección de información sobre publicaciones que ya existe (*EuroCRIS*, 2019b).

Si no es posible contar con sistemas CRIS abiertos por defecto, el objetivo es tratar al menos de que sean tan interoperables como sea factible.

En el recientemente celebrado *euroCRIS Membership Meeting* de primavera de 2019 (CSC, Espoo/Helsinki, 27 a 29 de mayo de 2019) (De-Castro, 2019) se ha presentado por primera vez la estrategia de *OpenAIRE* en relación con los sistemas CRIS. Esta estrategia, aún en una fase muy temprana, se basa en la implantación de las directrices *OpenAIRE* para sistemas CRIS basadas en *CERIF-XML* y en la aplicación de los mismos esquemas de validación y cosechamiento y agregación de metadatos que



Figura 1. Los primeros sistemas CRIS cosechados por *OpenAIRE*

“Si no es posible contar con sistemas CRIS abiertos por defecto, el objetivo es tratar al menos de que sean tan interoperables como sea factible”

Driver primero y *OpenAIRE* después aplicaron en el ámbito de los repositorios de acceso abierto. La presentación *OpenAIRE* realizada en Espoo por Andreas Czerniak, Jochen Schirrwagen y Aenne Löhden de la *Bielefeld University* (Loehden; Schirrwagen; Czerniak, 2019) mostró los seis primeros sistemas CRIS que han implantado las directrices CERIF-XML de *OpenAIRE* y han sido cosechados a nivel preliminar por la infraestructura del proyecto *OpenAIRE Advance*.

La idea subyacente es que los sistemas CRIS constituyen “cofres del tesoro de información contextual sobre la actividad científica institucional” como gusta de afirmar el presidente de *euroCRIS* Ed Simons de la *Radboud University* en Nijmegen, Países Bajos (Simons, 2017).

En tanto que los repositorios institucionales han estado tradicionalmente centrados en la recogida de información sobre resultados de investigación –publicaciones y más recientemente datos de investigación– con un gran énfasis en la puesta a disposición pública de textos completos de publicaciones, los sistemas CRIS se orientan a la colección de toda la información relacionada con la actividad investigadora. Esto incluye naturalmente metadatos sobre publicaciones y datos de investigación, pero también –entre otros– sobre patentes, proyectos de investigación, investigadores y sus afiliaciones, instalaciones y equipos de investigación en que se han llevado a cabo los experimentos o doctorandos asociados a los proyectos y sus fuentes de financiación.

De manera crítica, todas estas “entidades” –por usar la terminología del modelo de datos subyacente CERIF (*Common European Research Information Format*) (*EuroCRIS*, s.f.a)– están conectadas entre sí, de modo que si la información es suficientemente completa y detallada, resulta sencillo extraer informes basados en cualquiera de las entidades del modelo.

<https://www.eurocris.org/cerif/main-features-cerif>

Tirando del hilo, todas las demás entidades conectadas saldrán a relucir, y será posible por ejemplo determinar los resultados del uso de un determinado equipo científico: publicaciones, datos, grupos de investigación involucrados dentro y fuera de la institución, colaboraciones con la industria. Esto resulta particularmente valioso para las agencias de financiación que invierten ingentes cantidades de dinero público en la adquisición o el desarrollo de infraestructuras de investigación tipo canales hidrodinámicos, telescopios o buques oceanográficos (EPSRC, 2017).

La filosofía es exactamente la misma que la que aplica a la ciencia abierta: transparencia, reproducibilidad y el viejo mantra (todavía utópico) de introducir los datos una vez en un sistema para su múltiple reutilización en múltiples sistemas. Los repositorios de acceso abierto, con todas sus ventajas y aspiraciones, no recogen como norma esta información contextual que permite un salto cualitativo en la calidad del *reporting*, pero la recolección y agregación de la información científica que albergan los sistemas CRIS y su integración con la que ya se está obteniendo desde los repositorios mediante estrategias de deduplicación basadas en identificadores persistentes abrirían las puertas a este tipo de prácticas avanzadas en la gestión pública de la información de la investigación (financiada a su vez y en su inmensa mayoría con fondos públicos).

euroCRIS CRIS: Treasure Chests of metadata

A CRIS is an information system that holds (interlinked) metadata about virtually all aspects of research:

- The **researchers** (ID, name, title, affiliation, ...) and their role in the research.
- The **projects** (ID, title, description, key words, start- and enddate, ...)
- The **organisations** (institutes, universities,...) involved and their role (owner, funder, ...)
- The **input** invested in the research both in time (f.t.e.) and money.
- The **output** from the research (publications, datasets, software, patents, images, etc...)
- The **equipment, services and tools** (e.g. which version of which software) used.
- The **field or subject** of the research.
- The **impact of the research**, both in- and outside of academia (metrics, impact indicators,...).
- **Semantic classifications of the research** on various dimensions (typologies, ...),
- **Rights metadata**: who is authorised or which conditions may apply to access (some of the) information in the CRIS
- **Cooperations** with other projects, researchers, groups, organisations.
- **Links to other systems** (HRM, financial, external: ORCID, WoS, Scopus, ...)

Figura 2. Sistemas CRIS como “cofres del tesoro” de metadatos (Simons, 2017)

“Los sistemas CRIS se orientan a la colección de toda la información relacionada con la actividad investigadora”

“La filosofía es la misma que en la ciencia abierta: transparencia, reproducibilidad y el viejo mantra (todavía utópico) de introducir los datos una vez en un sistema para su múltiple reutilización en múltiples sistemas”

3. Cultura de colaboración y sistemas CRIS internacionales

Más que en la generalizada presencia de actores comerciales en el sector, el auténtico desafío lo plantea la evolución de una cultura basada en la competición entre instituciones (por los proyectos, por la financiación, por el impacto y por la imagen pública) hacia dosis más elevadas de colaboración cuando menos a nivel técnico para hacer posible la recogida de los frutos del trabajo conjunto. Un ámbito éste en el que las oficinas institucionales de gestión de la investigación podrían aprender mucho de la tradicional actitud colaborativa entre bibliotecas de investigación en múltiples áreas. Es por esto que el rol de las bibliotecas en la gestión de la información científica –destacado tanto por el informe *Rebiun/Crue* en España como por el elaborado conjuntamente por *OCLC* y *euroCRIS* a nivel internacional (*OCLC*, 2019)– es particularmente crítico para aportar esta actitud colaborativa, idealmente en el ámbito internacional además de en el estrictamente nacional o regional.

Es interesante observar que el tema principal del evento *euroCRIS* de junio en Helsinki no fue la consolidación de una interoperabilidad avanzada de sistemas, sino la gradual emergencia de sistemas internacionales de gestión de la información científica. La posibilidad de mostrar la información de los sistemas CRIS al exterior plantea múltiples retos, tales como determinar qué parte de la información es susceptible de ser expuesta y en qué medida esta tendencia puede verse afectada por la legislación europea de protección de datos, así como los imprescindibles progresos necesarios en la armonización de áreas como las clasificaciones científicas. Pero esta exposición de la información al exterior abre asimismo la puerta a la posibilidad de trabajar en el ámbito internacional asociado al *Área Europea de Investigación (ERA)*, máxime cuando tanto los proyectos de investigación como la utilización de las infraestructuras de investigación son con frecuencia internacionales. En un momento en el que la actualidad política europea está con frecuencia marcada por la división, la posibilidad de trabajar conjuntamente a nivel internacional para alcanzar estos objetivos resulta particularmente gratificante.

4. El directorio *euroCRIS* de *Sistemas de Información Científica (DRIS)*

En el encuentro de Helsinki, *euroCRIS* presentó también su iniciativa para elaborar un directorio comprensivo de sistemas de gestión de información científica, el directorio DRIS actualmente en desarrollo y que cuenta ya con alrededor de 300 entradas.

<https://www.eurocris.org/eurocris-directory-research-information-systems-drjs>

La descripción del panorama de implantación de sistemas CRIS a nivel nacional o internacional ha constituido tradicionalmente un objetivo muy difícil de alcanzar, en parte debido a esta cultura de competencia entre instituciones, sumada al hecho de que los sistemas CRIS se diseñaron originalmente para su uso interno por parte de universidades y centros de investigación. Uno no recuerda haber escuchado nunca cuestiones acerca de los motivos para el desarrollo de un directorio de repositorios de acceso abierto como *OpenDOAR*. Cierto es que en la génesis de proyectos como *OpenDOAR* o *ROAR* estaba muy presente la necesidad de recolectar los metadatos de los repositorios a través del protocolo OAI-PMH por parte de agregaciones de contenido que explotaran la naturaleza distribuida de la infraestructura. Y cierto es también que apenas había proveedores comerciales de repositorios, sino que se trataba en su mayoría de desarrollos propios basados en plataformas de código abierto.

Con la gradual emergencia de una infraestructura de gestión de la información científica igualmente distribuida, de cuya explotación conjunta podrían obtenerse grandes beneficios, puede cobrar más sentido tratar de mantener un mapa de la implantación de sistemas CRIS cuando menos a nivel europeo. Aun así, abundan aún las actitudes reticentes entre proveedores comerciales a la hora de facilitar la (mínima) información necesaria para mantener una entrada para un sistema CRIS que no quede desactualizada en breve tiempo. Será por tanto la actitud colaborativa de las instituciones a la hora de facilitar información sobre sus sistemas la que determinará si la iniciativa de describir el ‘paisaje CRIS’ tiene o no éxito a medio plazo.

Este ámbito de los proveedores comerciales dista no obstante de ser blanco o negro. La imagen que por el momento proyecta el directorio DRIS de *euroCRIS* es una mezcla de soluciones de desarrollo propio y de plataformas comerciales, con importantes diferencias en los perfiles por países. En el área específica de la interoperabilidad de sistemas CRIS, la implantación de las directrices *CERIF-XML* de *OpenAIRE* puede realizarse de manera individual para cada una de las soluciones institucionales de desarrollo propio, lo que permite un elevado nivel de control institucional sobre los ritmos y las funcionalidades a abordar, pero puede también realizarse de un golpe para una plataforma comercial con gran implantación como *Pure* (o como *Omega-PSIR* en Polonia). El papel de los proveedores comerciales en este sentido es muy relevante, porque representan un solo proceso de implantación de estrategias para incrementar la interoperabilidad que todos sus clientes en diversos países pueden potencialmente explotar para exponer públicamente la información de gestión de su investigación (*Semicrol*, 2019).

Por otro lado, soluciones como los repositorios institucionales basados en plataformas de código abierto gozan de amplia popularidad en universidades, sobre todo entre las de cierto tamaño, pero son más bien escasas en centros de investigación y prácticamente inexistentes en otras entidades como hospitales, pequeñas y medianas empresas o fundaciones, organizaciones todas ellas de creciente importancia en el ámbito de la gestión de proyectos europeos de investigación. Desde una perspectiva pragmática es por tanto conveniente no centrarse excesivamente en el carácter de un sistema CRIS específico tanto como en el valor de la información que contiene y en los beneficios de poder exponerla para su recolección por agregadores externos de información científica.

5. Bibliotecas de investigación y gestión de la información científica para la ciencia abierta

El acceso de las bibliotecas de investigación a la gestión –cuando menos parcial– de los sistemas de información de la investigación permite asimismo su aplicación a la implantación institucional de la ciencia abierta. En áreas como el acceso abierto o la gestión de datos de investigación, la posibilidad de acceder a la información detallada sobre proyectos de investigación asociados y agencias que los financian proporciona un contexto de suma utilidad en la materia (**De-Castro, 2018**). En el Reino Unido, esto se ve favorecido además por una política de mandato de acceso abierto a los manuscritos aceptados en revistas y conferencias, un mandato asociado al ejercicio de evaluación de la actividad científica a nivel nacional. Los resultados del reciente ranking del instituto CWTS en Leiden sobre porcentajes de la producción científica institucional disponible en acceso abierto son buena muestra de la efectividad de estas políticas, con 25 de las 30 primeras instituciones en el ranking localizadas en Reino Unido.

<http://www.leidenranking.com/ranking/2019/>

Más aún, la obligatoriedad de depositar una copia del texto completo tan pronto como sea posible tras la aceptación del manuscrito genera una situación en la que los registros bibliográficos de las publicaciones se crean en primer lugar en las instituciones, como norma mucho antes de que existan registros en los editores, en *CrossRef* y en las bases de datos internacionales. Esto lleva inevitablemente a plantearse si una extensión de estas políticas a otros ámbitos geográficos no haría posible que las instituciones retomaran el control de sus datos sin verse obligadas a adquirirlos a proveedores exter-

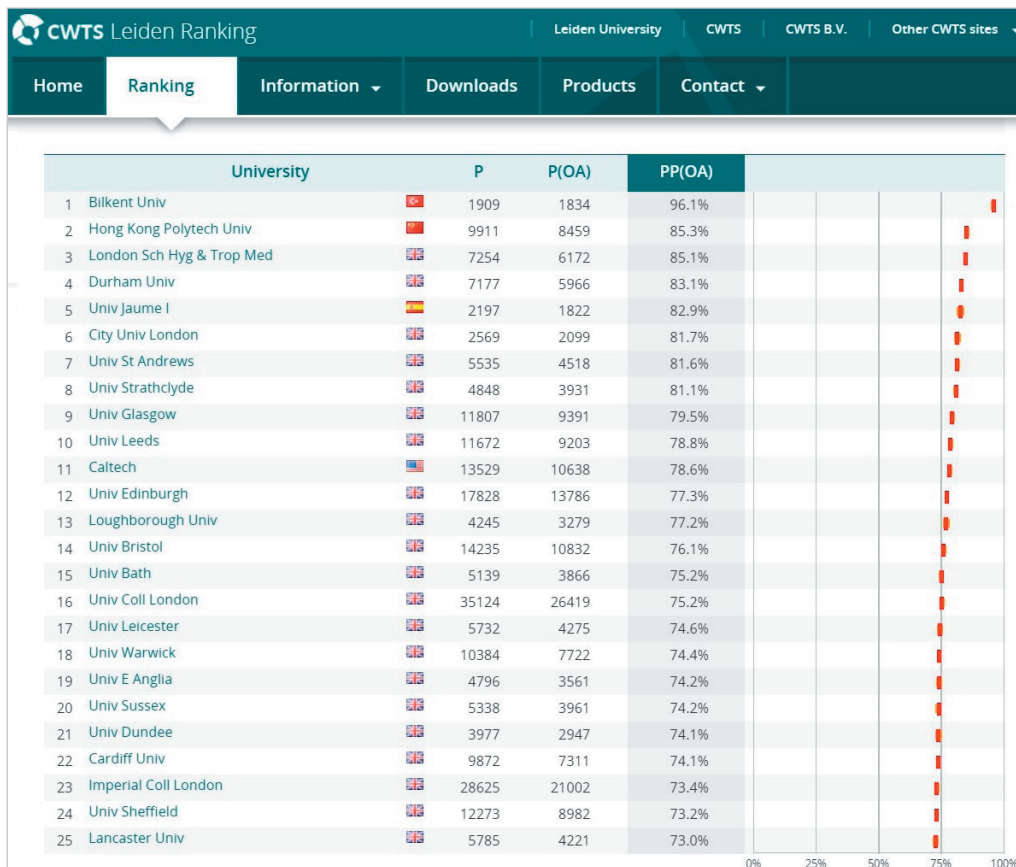


Figura 3. Ranking CWTS Leiden de instituciones por porcentaje de producción científica en abierto

nos. Ciertamente es que las bases de datos internacionales contienen mucha más información que los meros registros bibliográficos, pero toda esta información adicional puede generarse por parte de cualquier agente que cuente con los datos originales.

Estas consideraciones, junto al bien conocido factor de la baja cobertura de las publicaciones en las áreas de ciencias sociales y humanidades por parte de bases de datos como *Scopus* o *WoS*, están en la base del proyecto *Enressh* (*European Network for Research Evaluation in the Social Sciences and Humanities*)

<https://enressh.eu>

Enressh, que cuenta con participación española desde el CCHS-CSIC, es una iniciativa internacional para explorar el nivel de cobertura de la producción científica en ciencias sociales y humanidades ofrecido por bibliografías de ámbito nacional en países europeos. Tal como dice el título del proyecto, el objetivo es poder basar las actividades de evaluación de la actividad científica en el área de ciencias humanas y sociales en una colección de datos suficientemente comprensiva. Este es el mismo objetivo que tienen los sistemas CRIS, y el encuentro *euroCRIS* de Helsinki contó con una sesión dedicada a difundir los progresos del proyecto *Enressh* y a explorar posibles sinergias con el entorno de los sistemas de gestión de la información científica.

Es desde el liderazgo de este proyecto *Enressh* desde donde se está progresando de manera más firme en el concepto de un sistema CRIS de ámbito internacional. En los países nórdicos, donde el ámbito de la gestión de la información científica está muy consolidado y donde existe una fuerte tradición de colaboración internacional, se ha presentado ya una propuesta para el sistema *NordRIS* que habría de servir a Finlandia, Suecia, Noruega, Dinamarca e Islandia (Røeggen, 2018).

Las más recientes ediciones del ranking *Webometrics* de repositorios han comenzado también a incluir algunos sistemas CRIS como una categoría adicional a tener en cuenta (*Cybermetrics Lab*, 2019). La expansión del modelo de datos de plataformas de repositorios de código abierto como *DSpace* y *EPrints ranking* ha llevado a una convergencia de sistemas, de modo que en la actualidad soluciones *open source* como *DSpace-CRIS* o *Haplo* constituyen de hecho una integración de repositorio y CRIS. Al contrario que en el caso de los repositorios institucionales, no está claro que la visibilidad en *Google Scholar* constituya una métrica adecuada para elaborar un ranking genérico de sistemas CRIS, salvo quizá en estos casos en los que los sistemas CRIS se han creado como repositorios expandidos. El número de registros exportados a *OpenAIRE* para las diversas entidades CERIF (publicaciones, datasets, patentes, proyectos de investigación, equipos científicos, etc.) sería como norma un indicador mucho más fiable. Dado que esta exportación todavía no está teniendo lugar, el directorio DRIS de *euroCRIS* está entretanto públicamente disponible para su exploración como fuente de información para posibles servicios por parte de actores externos.

6. Referencias

Calderón-Rehecho, Andoni (2017). "Portales de producción científica". *Anuario ThinkEPI*, v. 11, p. 247-255.

<https://doi.org/10.3145/thinkepi.2017.45>

Cybermetrics Lab (2019). *CCHS-CSIC Webometrics Ranking May 2019: CRIS*. Cybermetrics Lab CCHS-CSIC.

<https://repositories.webometrics.info/en/cris>

Crue (2018). *Hércules: Semántica de datos de investigación de universidades*. Comisión sectorial de Tecnologías de la Información y las Comunicaciones de la Crue.

<http://tic.crue.org/hercules>

De-Castro, Pablo (2018). "The role of current research information systems (CRIS) in supporting open science implementation: the case of Strathclyde". *ITlib. Informacné technológie a knižnice*, special issue, pp. 21–30.

<https://doi.org/10.25610/itlib-2018-0003>

De-Castro, Pablo (2019). "Notes on the Spring 2019 *euroCRIS Membership Meeting* at CSC in Espoo-Helsinki". *euroCRIS blog post*, 6 junio.

<https://www.eurocris.org/blog/notes-spring-2019-eurocris-membership-meeting-csc-espoo-helsinki>

EPSRC (2017). *Strategic equipment scheme. Maximising existing equipment sharing in physical sciences*. Engineering and Physical Sciences Research Council.

<https://epsrc.ukri.org/funding/calls/strategic equipmentscheme>

EuroCRIS (2019a). "euroCRIS and OpenAIRE sign memorandum of understanding". *EuroCRIS. Current research information systems*, 27 mayo.

<https://www.eurocris.org/news/eurocris-and-openaire-sign-memorandum-understanding>

EuroCRIS (2019b). "Metis2OpenAIRE project". *EuroCRIS. Current research information systems*, 1 marzo.

<https://www.eurocris.org/projects/metis2openaire>

Löhden, Aenne; Schirrwagen, Jochen; Czerniak, Andreas (2019). "CRIS in OpenAIRE - we take you onboard". En: *Spring 2019 euroCRIS Membership Meeting* (CSC, Espoo-Helsinki, Finland, May 27-29, 2019). <http://hdl.handle.net/11366/976>

OCLC (2019). "An increasing role for libraries in research information management". *OCLC blog post*. OCLC Research. <http://oc.lc/rim>

Rebiun (2018a). *Estado de la cuestión de los CRIS en las universidades españolas. Subgrupo de Acceso Abierto Línea II del Plan Estratégico de Rebiun*. Crue; Rebiun. <http://213.32.37.214/xmlui/handle/20.500.11967/221>

Rebiun (2018b). "Estado de la cuestión de los CRIS en las universidades españolas. Subgrupo de Acceso Abierto Línea II del Plan Estratégico de Rebiun". En: *XVII Workshop Rebiun de Proyectos Digitales (Universidad Pablo de Olavide, Sevilla, 25-26 octubre 2018)*. <https://rio.upo.es/xmlui/handle/10433/5830>

Røeggen, Vidar (2018). *Proposal to NordForsk for integrating the Nordic list with other development projects: the Nordic Research Information System (NordRIS)*. <https://enressh.eu/wp-content/uploads/2016/10/V.R%C3%B8eggen-ENRESSHCopenhagen.pdf>

Semicrol (2019). "Semicrol se adapta a los requisitos establecidos por euroCRIS para los sistemas de gestión de la investigación". *Semicrol Noticias*, 5 junio. <https://www.semicrol.es/Noticias/IDI245/Semicrol-se-adapta-a-los-requisitos-establecidos-por-euroCris-para-los-sistemas-de-gestion-de-la-investigacion>

Simons, Ed (2017). "Open science: The crucial importance of metadata". En: *euroCRIS Strategic Membership Meeting Autumn 2017* (CVTI SR, Bratislava, Slovakia, Nov 20-22, 2017). <http://hdl.handle.net/11366/627>

Pablo De-Castro

University of Strathclyde & euroCRIS
pcastrmartin@gmail.com

Si te interesan los

INDICADORES EN CIENCIA Y TECNOLOGÍA,

y todos los temas relacionados con la medición de la ciencia, tales como:

Análisis de citas, Normalización de nombres e instituciones, Impacto de la ciencia en la sociedad, Indicadores, Sociología de la ciencia, Política científica, Comunicación de la ciencia, Revistas, Bases de datos, Índices de impacto, Políticas de open access, Análisis de la nueva economía, Mujer y ciencia, etc.

Entonces **INCYT** es tu lista. Suscríbete en:

<http://www.rediris.es/list/info/incyt.html>