



## Repositorio Institucional e-IEO a finales de 2014

Informe para la Reunión de Directores del IEO, Madrid, 24-25 de noviembre de 2014

Concha Mosquera-de-Arancibia

### Resumen

Se presentan los distintos trabajos realizados en el repositorio del Instituto Español de Oceanografía (e-IEO) durante 2014: 1. Interoperabilidad semántica y técnica con el Sistema Integrado de Proyectos de Investigación del IEO (SIPI) —base de datos interna de la institución para seguimiento de proyectos y evaluación de investigadores— para evitar a los investigadores el doble archivo y para difundir sus trabajos. 2. Estandarización de los nombres de autor, mediante la implantación del control de autoridades (authority control) de DSpace. 3. Adopción de los identificadores internacionales únicos de autor ORCID (Open Researcher and Contributor ID) para la identificación y detección de ambigüedades y duplicidades en los nombres de los investigadores y la correcta adjudicación de sus obras. 4. Enlace con las principales bases de datos internacionales de autores y publicaciones (ORCID, Google Scholar, ResearcherID, ScopusID, Dialnet) y sus métricas de impacto: las clásicas y las de evaluación del impacto en las redes sociales (Altmetrics). 5. Mejoras en el módulo gráfico de estadísticas de DSpace. 6. Formación de los administradores de colección que trabajan en los distintos centros del IEO. 7. Implementación de un blog como foro de comunicación entre administradores y usuarios internos. 8. Inclusión de vocabularios controlados específicos de ciencias marinas (recursos marinos, pesquerías, etc.) como mecanismo de mejora de las capacidades de búsqueda y descubrimiento de los ítems. 9. Integración en recolectores temáticos de ciencias del mar. 10. Presentación de la situación actual del repositorio: distribución de ítems por comunidades del repositorio y Centros del IEO, y su evolución en 2014.

Diapositiva 1.- e-IEO a finales de 2014: <http://www.repositorio.ieo.es/e-ieo/>

Diapositiva 2.- Ya en la recta final del año y del proyecto e-IEO 2014, se han efectuando la mayoría (✓) de las diez tareas programadas, presentadas en la Reunión de Directores de febrero. Solo quedan por implementar las tres últimas (blog, vocabularios controlados



e integración en recolectores temáticos), que ya tenemos muy avanzadas, e integrar la institución en ORCID, para lo que el repositorio ya está preparado.

En este año, en el que se han conseguido los objetivos previstos, la gran parte de los esfuerzos se han dedicado a la *interoperabilidad* entre en el repositorio institucional y SIPI, para establecer una plataforma única de envío de publicaciones en abierto y difundir, a escala internacional, todos los trabajos del IEO, excepto los informes confidenciales, que, a petición expresa de sus coordinadores, no pasan al repositorio y, si en algún momento fuera de interés que estuviesen en abierto, habría que subirlos.

Cumplimos así con el artículo 37 de la Ley de la Ciencia y estamos preparados para la unión con el entorno de datos en abierto que exige H2020.

Diapositiva 3.- La interoperabilidad semántica y técnica entre los sistemas REPOSITORIO e-IEO y SIPI evita el doble archivo e incrementa el valor de la producción investigadora pero obliga a sincronizar datos. Esta ha sido la tarea fundamental de este proyecto, para la que se homogeneizaron los elementos y los valores de todos los metadatos en ambos sistemas y adaptaron los formularios de entrada de DSpace a las necesidades de SIPI.

Ya los sistemas *interoperables*, se programó una ingesta masiva única, desde SIPI al repositorio, de todos los registros que se iban a poner en abierto. Esta carga se realizó en tres fases, para disponer, en el momento de la evaluación anual de los investigadores, de los registros de 2014 ya normalizados. En la primera fase, se incorporaron los registros correspondientes a 2014, y los que figuraban sin año (presumiblemente, publicaciones ‘en prensa’), disponibles el día programado para la ingesta: el 11 de agosto —aprovechando el periodo vacacional, de menor actividad en el repositorio—; en la segunda, los registros de 2014 incluidos en SIPI entre las fechas de la primera y la segunda cargas, el 18 de septiembre —justo antes del inicio del proceso de evaluación de los investigadores—; y en la tercera, toda la serie histórica de SIPI anterior a 2014, el 21 de noviembre.



La sincronización (mediante ficheros de intercambio) entre sistemas hace que, cada vez que el contenido del repositorio se incrementa o se actualiza, SIPI incorpore la nueva información ordenada cronológicamente.


Diapositiva 4.- La estandarización de nombres de autor, realizada gracias a la implantación del control de autoridades de DSpace, permite la validación en el repositorio (y su incorporación a SIPI) de los nombres normalizados de autores, colaboradores y editores del IEO.

La validación se realiza de forma automática (ingestas masivas) o manual (en el momento del depósito o en la edición de cada ítem y en las comprobaciones de las validaciones automáticas), coteja los nuevos valores con la base de datos de autoridades de autor incorporada al repositorio y actualizable, y asigna valores de confianza DSpace (aceptado, incierto, ambiguo, no encontrado y sin validar). De esta manera, se corrigen los falsos positivos (publicaciones adjudicadas a un autor que no es el suyo) y los falsos negativos (publicaciones que no han sido adjudicadas a su autor).






Mediante el mismo mecanismo, se ha incorporado también la validación de las revistas en las que publican los autores del IEO.



Diapositiva 5.- La adopción de los identificadores internacionales únicos de autor ORCID (Open Researcher and Contributor ID) permite la identificación y detección de ambigüedades y duplicidades en los nombres de los investigadores y la correcta adjudicación de sus obras.

ORCID comenzó hace dos años como un registro de autores individualizados y en este momento funciona más como registro de instituciones (con todos sus posibles autores incorporados).

El repositorio ya tiene preparada la conexión con ORCID y los autores que tengan ya su ID ORCID podrán conectar con sus datos de esta base de datos a través del logotipo , al lado de su nombre validado en el repositorio. En breve se efectuará la integración institucional.



Diapositiva 6.- Las conexiones con las principales bases de datos internacionales de autores y sus publicaciones (ORCID, Google Scholar, ResearcherID, ScopusID, Dialnet), a través de sus logotipos      (al lado del nombre validado de cada autor que esté dado de alta en ellas), permiten el acceso al perfil de autor y a las métricas de impacto clásicas de estas bases de datos (publicaciones, coautores, total artículos, citas, media de citas por artículo, índice h, índice i10, etc.)

Los ítems que tienen doi enlazan con las estadísticas Scopus del ítem a través del logotipo  , dentro del ítem. Todos los ítems, gracias a su URI, conectan con las estadísticas altmetrics para evaluación del impacto en las redes sociales. Se pueden consultar enlazando, dentro del ítem, a través del logotipo de Altmetric  39 .

Diapositiva 7.- Las mejoras en el módulo gráfico de las estadísticas de DSpace permite visualizar (en gráficos de barras y de línea) la evolución de las visitas y descargas en todo el repositorio, o en cada ítem, en distintos periodos de tiempo (de 1 mes a 5 años).

Diapositiva 8.- Se realizó un curso de formación para los administradores de colección de todos los Centros, con una primera fase presencial, de dos días de duración, en Madrid, y una segunda online, a continuación, de tres semanas.

Al curso en Madrid asistieron todos los administradores y sirvió también para confirmar, como tales, a las personas que realizan estas labores, a partir del curso, en los Centros de Cádiz, Málaga y Murcia, Centros que carecían de ayuda para estas tareas. Canarias y Madrid (Sede Central) continúan sin administrador de colección. Virginia Martín, investigadora de Canarias, asistió en calidad de representante de su Centro para informar a sus compañeros. Mari Cruz Iglesias, administradora de colección de Baleares, estuvo presente por videoconferencia.

Administradores de colección en cada Centro: A Coruña, Manuel Vázquez; Baleares, Mari Cruz Iglesias; Cádiz, Benjamín Ballesteros; Gijón, Rosa Vázquez; Málaga, José Miguel Álvarez; Murcia, Ignacio Franco; Santander, Regina Herrera; Vigo, Uxía Tenreiro y Alexia Luiña.

Todos los administradores de colección realizaron la fase online del curso.



Diapositiva 9.- El blog como foro de comunicación entre los administradores y los usuarios del IEO, ya está en pruebas en la máquina de producción y se implementará próximamente.

Diapositiva 10.- La inclusión de vocabularios controlados específicos de ciencias marinas (recursos marinos, pesquerías, etc.), con el empleo de vocabularios especializado y tesauros, mejorará las capacidades de búsqueda y descubrimiento de los ítems gracias a su interoperabilidad semántica con los recolectores internacionales. Se implementará en breve el vocabulario AGROVOC-FAO de ASFA.

Diapositiva 11.- Está prevista la integración en recolectores temáticos de ciencias del mar, a través del protocolo internacional de comunicación OAI-PMH, para seguir aumentando la visibilidad de la producción del IEO.

Diapositiva 12.- El 21 de noviembre, en cuanto a la distribución de ítems por centro oceanográfico, el C. O. de Vigo sigue siendo el de mayor número de ítems (2.423, el 31 % del total), con más de tres veces el siguiente, la Sede Central, Madrid (882, el 11 %), donde están ubicados todos los artículos de *Boletín. Instituto Español de Oceanografía* desde 1996 a 2007, cuyos autores eran, en su inmensa mayoría, ajenos a la institución. El resto de los Centros aparecen representados con porcentajes entre el 9 % de Málaga (698 ítems) y el 4 % de Gijón (303 ítems). Santander (662 ítems), Murcia (612), Baleares (610) y Canarias (602) tienen una representación del 8 %. A Coruña (513) y Cádiz (494) representan el 7 % y el 6 %, respectivamente.

La distribución por comunidades del repositorio, a finales de noviembre de 2014, es:

- Área de Medio Marino y Protección Ambiental, el 43 % (3.328 ítems).
- Área de Pesquerías, el 38 % (2.994 ítems).
- Área de Acuicultura, el 17 % (1.346 ítems).
- Apoyo a la Investigación, el 1 % (64 ítems).
- Gabinete de Prensa el 1 % (68 ítems).

Hay que recordar que, aunque los nombres de las comunidades del repositorio coinciden con los de las áreas del IEO, las comunidades incluyen los ítems relativos a esas



materias de estudio, independientemente de que los primeros autores pertenezcan, o no, a esas áreas. Así, por ejemplo, hay trabajos relativos al medio marino que están firmados por autores de pesquerías.

En este momento, y de forma provisional, en la comunidad Gabinete de Prensa, están almacenados los ítems del último volcado desde SIPI en los que el valor de 'área de investigación' figuraba como 'desconocido' (62), que se pasarán a sus comunidades respectivas.

En cuanto a la evolución del total de ítems en el repositorio en estos diez meses de 2014, desde el informe presentado en la primera Reunión de Directores del año (18-febrero-2014), vemos que el número se ha multiplicado por 7,5 (1.058 ítems en febrero y 7.885 en noviembre, cuando el repositorio acaba de cumplir 3 años). Si bien hay que tener en cuenta que existirán algunas duplicidades que se irán corrigiendo en las próximas semanas. En este intervalo de tiempo, se realizó la ingesta masiva de registros, desde SIPI al repositorio, en las tres fases explicadas en la diapositiva 3. La primera, en agosto, de 226 registros; la segunda, en septiembre, de 88; y la tercera, el 21 de noviembre, de 5.887 registros. El número de registros se multiplicó por 1,9 en estos diez meses, antes de la 3ª fase de la ingesta masiva, y por 4,2, después de esta. De los registros en el repositorio antes de la 3ª fase de la ingesta masiva, 314 fueron cargados desde SIPI en las dos primeras fases (226 + 88) y, el doble: 626, enviados directamente por los usuarios del IEO.

El número de visualizaciones de ítems se ha incrementado en 1,8 veces (368.908 en febrero y 659.003 en noviembre, antes de la tercera fase de carga) y la media mensual de visualizaciones, en 1,3. Las estadísticas de visualizaciones, finalizada la ingesta masiva, se podrán analizar pasado un tiempo.