

Hacia una Arquitectura para Redes de Archivos Abiertos en Ciencias y Tecnología Agrícolas

Imma Subirats, Irene Onyancha, Gauri Salokhe, Johannes Keizer

Organización de las Naciones Unidas para la Agricultura y la Alimentación,
Roma, Italia

{Imma.Subirats, Irene.Onyancha, Gauri.Salokhe,
Johannes.Keizer}@fao.org

Resumen. La Red AGRIS es una iniciativa internacional cimentada en una red de colaboración de instituciones cuyo objetivo es promover el libre acceso a información sobre ciencias y tecnología agrícolas y temas afines. Esta ponencia explica cómo emplear los modelos de Acceso Abierto (Open Access, OA) y de la Iniciativa de Archivo Abierto (Open Archive Initiative, OAI) en el marco de la Red AGRIS para zanjar las dificultades de difusión e intercambio de resultados de investigaciones agrícolas. La falta de instancias para el intercambio adecuado de información entre investigadores en ciencias y tecnología agrícolas ha coartado la posibilidad de responder debidamente a cuestiones relacionadas al desarrollo agrícola. El modelo de OA promueve la difusión de resultados de investigación a nivel internacional, nacional y regional, eliminando de esta manera las restricciones impuestas por el modelo tradicional de publicaciones científicas. Esta ponencia aborda asuntos relativos al acceso, disponibilidad e interoperabilidad en lo que dice relación con el intercambio de trabajos de investigación agraria.

Palabras clave: AGRIS, Investigación y Tecnología Agrícola, Acceso Abierto, Iniciativa de Archivo Abierto, Proveedores de Datos, Proveedores de Servicios, Perfil de Aplicación del AGRIS, Sistemas de Organización del Conocimiento, Interoperabilidad

1 Introducción

La Red AGRIS, en operación desde 1975, partió como una iniciativa internacional cimentada a partir de una red colaborativa de instituciones cuyo objetivo era desarrollar un sistema común de información de libre acceso sobre ciencias y tecnología agrícolas y otros temas afines. Hasta fines de los años noventa, los productos relacionados al AGRIS se componían fundamentalmente de una base de datos bibliográficos centralizada y productos anexos. Sin embargo, a partir del año 2000, la red AGRIS comenzó a dirigir sus acciones hacia el fortalecimiento de la

capacidad descentralizada de sus centros de recursos participantes, afianzándolos para optimizar la gestión de información agraria dentro de sus propias instituciones. Una de las tareas más importantes del AGRIS es facilitar el acceso a bibliografía científica, tecnológica y académica, así como -y especialmente- a la literatura “gris”. La literatura gris contribuye de manera significativa a la investigación agraria, especialmente en los países en desarrollo, a pesar de que la comunidad agrícola aún no cuenta con un acceso adecuado a ella.

El avance vertiginoso de la tecnología ha ido creando nuevas formas de compartir datos, información y conocimientos. El mundo de las comunicaciones ha sido testigo en estos últimos años de varios sucesos, entre ellos el surgimiento del Acceso Abierto (Open Access, OA) y de la Iniciativa de Archivos Abiertos (Open Archives Initiative, OAI). Estas iniciativas se complementan entre sí y tienen como objetivo principal mejorar la visibilidad y el acceso a resultados de investigación, maximizando así su impacto. El Acceso Abierto se define como “el libre acceso a través de Internet que permite a los usuarios leer, descargar, copiar, distribuir, imprimir, buscar o referenciar por medio de enlaces, el texto completo de artículos académicos o científicos, hacer búsquedas para indización, transferir el contenido como datos para un software, o usar la obra para cualquier otro propósito legítimo, sin otras barreras financieras, legales o técnicas más que las que suponga Internet en sí misma” (BOAI, 2002). El modelo de Acceso Abierto hace posible que los resultados de investigación sean difundidos ampliamente y con máxima visibilidad, eliminando de esta manera las restricciones impuestas por los métodos tradicionales de divulgación de bibliografía científica. El modelo permite a los investigadores acceder a documentos de texto completo de dos formas: a través de la publicación del trabajo en revistas de acceso abierto que permiten acceder libremente a artículos apenas son publicados, y por medio del Archivado de Acceso Abierto que permite a los autores “depositar un documento en formato digital en un sitio web de acceso público, de preferencia uno que se ajusta a las normas establecidas por la OAI para archivos abiertos” (Preguntas frecuentes sobre Auto Archivo, 2006). La Iniciativa de Archivos Abiertos (OAI) desarrolla y promueve soluciones de interoperabilidad con vistas a permitir una difusión eficiente de contenidos. La iniciativa, basada en el principio de la interoperabilidad, federará la distribución de archivos

abiertos, fomentado así el desarrollo de servicios con valor agregado como portales web, portales de acceso por materias (*gateways*) y motores de búsqueda especializados, y cuya ventaja radica en que aumenta la visibilidad de los Archivos Abiertos. El objetivo final es eliminar los obstáculos a la interoperabilidad que existen al utilizar archivos abiertos para todo el material digital (Van de Sompel & Lagoze, 2000).

Los modelos de publicación de Acceso Abierto y de la Iniciativa de Archivos Abiertos están estrechamente vinculados a los principios de AGRIS. La comunidad AGRIS, al aplicar los conceptos del OA, fortalecerá la capacidad de sus centros de recursos para difundir y compartir sus resultados de investigación y, a su vez, facilitará el acceso y aumentará la visibilidad de las publicaciones académicas sobre temas agrícolas. De más está decir que la red AGRIS también tendrá que ajustarse a las normas y condiciones específicas establecidas para Archivos Abiertos.

Este documento define y describe una arquitectura de alto nivel basada en la estructura de Archivos Abiertos para la comunidad AGRIS. Dicha arquitectura consta de tres componentes: (i) la creación de contenidos empleando las normas acordadas para la descripción de contenido, (ii) la recolección de estos contenidos usando normas comunes de intercambio, y (iii) proporcionar servicios con valor agregado a usuarios usando contenidos estándar intercambiados. Esta ponencia, por lo tanto, se divide en tres secciones. La Sección 3 presenta una síntesis de lo que es la arquitectura *workflow* e incluye una descripción de los diversos elementos que la componen y su interrelación. La Sección 4 hace referencia al objetivo general de esta nueva arquitectura, basada en el modelo de Acceso Abierto, y a los dos servicios vinculados a éste, es decir, los proveedores de datos y los proveedores de servicios, el papel específico que cada uno de ellos cumple y sus formatos comunes, y el protocolo de integración por medio de recolección de metadatos. La Sección 5 explica la base lógica de las normas de intercambio y su utilización, incluyendo los metadatos, los Sistemas de Organización del Conocimiento tales como los tesauros y formatos comunes de integración, considerados la piedra angular de la web semántica. Finalmente, en la Sección 6 se presentan las conclusiones y se hace

referencia a las etapas siguientes necesarias para implementar la arquitectura.

2 Objetivos

La red AGRIS promoverá el paradigma de la publicación de Acceso Abierto y aplicará las normas y metodologías de la OAI dado que estas últimas contribuirán de manera significativa al cumplimiento de una de las metas de larga data de AGRIS, esto es, mejorar tanto el acceso a la información y conocimientos sobre ciencias y tecnología agrícolas como el intercambio de éstos entre países en desarrollo. Entre los objetivos específicos se incluye: aumentar la visibilidad y mejorar el acceso a contenidos, incluso a aquellos a los que nos se tenía acceso anteriormente, y promover normas temáticas comunes, afianzando los principios fundamentales de la red AGRIS. La implementación de una nueva arquitectura utilizando herramientas relacionadas al Acceso Abierto traerá grandes beneficios a la comunidad de investigadores agrarios ya que aumentará tanto la visibilidad de sus trabajos como el número de lectores. Es más, los Archivos Abiertos permiten a los potenciales lectores acceder fácilmente a documentos, quienes de otra manera tendrían acceso limitado dado los altos costos de las revistas académicas tradicionales. La iniciativa permite enlazar la investigación local con la internacional, así como entregar una mejor visión de la cantidad de artículos de investigación agraria publicados en un determinado país. El nuevo modelo provee una arquitectura que permite establecer una comunicación entre dos niveles de socios: los proveedores de datos, o sea, los poseen los metadatos que se expondrán, y los proveedores de servicios que recolectan estos metadatos. La comunicación entre ambos se hace posible gracias al modelo de interoperabilidad conocida como “recolección de metadatos”, que permite a los proveedores de datos desplegar sus metadatos a través de una interfaz abierta a los proveedores de servicios que utilizan estos metadatos como base para sus servicios con valor agregado. La calidad de la descripción de la información agraria dependerá del uso que se haga de las normas temáticas comunes y de la tecnología a emplear. Entre estas normas se incluyen las del Perfil de Aplicación del AGRIS (AGRIS AP) y de los Sistemas de Organización del Conocimiento (SOC), entre ellos el Tesauro AGROVOC, el cual se describe en mayor detalle en la Sección 5. Esto permitirá a los

proveedores de servicios temáticos en ciencias agrarias desarrollar mejores servicios a partir de los metadatos recolectados. La red AGRIS promoverá el cumplimiento de las normas y el uso de estas herramientas a objeto de consolidar la nueva arquitectura.

3 Los elementos de la arquitectura: un nuevo *workflow* o secuencia de procesos

La arquitectura propuesta para la Red AGRIS, como se detalla en la Figura 1, proporciona los elementos principales y la dinámica prevista para el flujo de información a través de la red. En términos conceptuales, el *workflow* está compuesto de tres actividades principales: la gestión de contenidos, la exposición o despliegue de metadatos y la prestación de servicios con valor agregado.

La actividad que tiene relación con la gestión de contenidos describe el actual proceso para la captura, descripción y almacenamiento de información utilizando herramientas y metodología adecuadas. Los sitios web de AGRIS y de AIMS¹ (Agricultural Information Management Standards) ofrecen una lista de las normas, métodos y software apropiados. Un nuevo elemento en el proceso de captura es el del auto archivo de documentos de texto completo por los mismos autores investigadores. Los centros AGRIS promoverán el archivo de documentos de texto completo, ya sea por el autor o por un representante institucional.

¹ Sitio web de AIMS. <http://www.fao.org/aims/>. Ultimo acceso: enero 2007

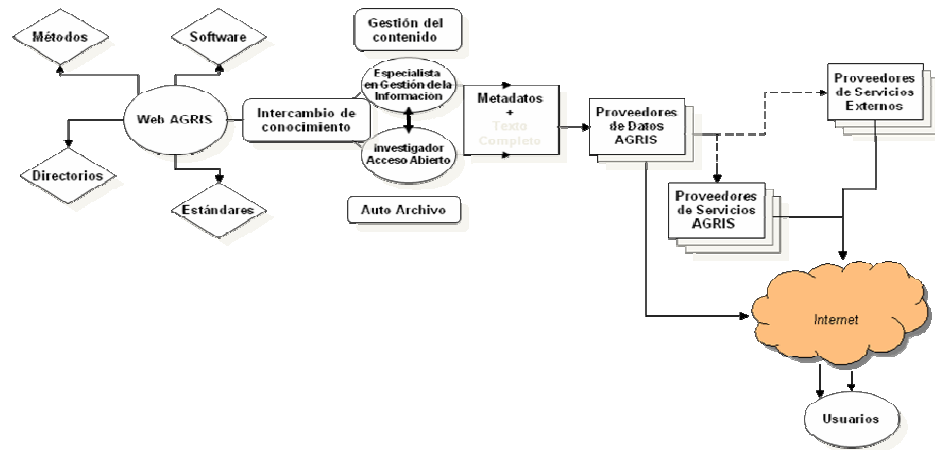


Figura 1. El nuevo *workflow* de la red AGRIS

La segunda parte del *workflow* es la exposición o despliegue de los metadatos. Esto requiere que los socios de AGRIS se transformen en proveedores de datos y que éstos interoperen con los proveedores de servicios mediante la implementación de normas y procedimientos de la OAI. Los proveedores de datos deberán registrar y describir sus colecciones en un repertorio ubicado en el sitio web de AGRIS a fin de garantizar su visibilidad. La Red AGRIS busca ir más allá del intercambio de metadatos Dublin Core simples (un requisito para participar en la comunidad de la OAI) y recomendar el uso del AGRIS AP – un formato de intercambio más completo para recolectar e intercambiar documentos agrícolas. El AGRIS AP y los SOC son elementos fundamentales para la descripción de contenidos ya que permiten a los proveedores de servicios seleccionar y recolectar datos de calidad que, a la postre, mejorará la calidad de sus servicios. La última acción del *workflow* corresponde a la de los proveedores de servicios temáticos, los cuales aportan valor agregado a los metadatos recolectados al proporcionar plataformas sobre las cuales los investigadores podrán interactuar y compartir información. La red AGRIS ha creado comunidades temáticas que realizan actividades afines y comparten servicios; estos grupos ya han establecido una base sobre la cual se pueden definir las tareas y responsabilidades de los proveedores de servicios.

Los proveedores de servicios también podrán crear portales de acceso para temas específicos en diversas áreas dentro del ámbito de las ciencias agrarias. Asimismo, el material que mantienen los proveedores de datos también será recolectado por otros proveedores de servicios fuera de la red AGRIS. En términos generales, la nueva red AGRIS se visualiza como una instancia en la cual cada uno de los elementos descritos anteriormente cumple un papel activo en el intercambio de metadatos y conocimientos. La implementación de redes de archivos abiertos es la unidad básica de la nueva arquitectura ya que le permitirá a la comunidad de investigadores agrarios acceder a y compartir sus publicaciones.

4 Los proveedores de datos: incorporación del modelo de Acceso Abierto

La implementación a nivel institucional del modelo de Acceso Abierto (u OA por su sigla en inglés para Open Access) obligará a los proveedores de datos a adoptar mecanismos que apoyen la implementación de repositorios digitales. Estos son ficheros de acceso abierto que pueden incluir contenidos como artículos *pre-prints* (que no han pasado por un proceso de arbitraje) o *post-prints* (que han sido revisados por pares y aceptados, publicados o se encuentran en proceso de publicación), tesis, manuales, material docente u otros documentos que los autores o instituciones desean poner a disposición del público sin costo y sin restricciones. La infraestructura técnica necesaria para montar repositorios institucionales se puede adquirir fácilmente dada la gran cantidad de software de código abierto que existe actualmente que cumple las normas establecidas y cuyo costo de instalación es mínimo. La puesta en marcha de estrategias y políticas institucionales de Tecnologías y Gestión de Información y Comunicaciones que contemplen un *workflow* idóneo para publicaciones y documentos también fomentará la institucionalización del modelo de Acceso Abierto. Otros asuntos a considerar son los derechos de autor asociados al autoarchivo de artículos publicados. La base de datos ROMEO (Rights Metadata for Open Archiving) disponible a través del sitio web SHERPA (Securing a Hybrid Environment for Research Preservation and Access) contiene información relativa a las políticas de copyright editoriales y para autoarchivo. Para facilitar este proceso, las ventajas del OA deben ser divulgadas y exhibidas tanto a los investigadores como a los responsables de tomar la decisión respecto a su utilización

pues son ellos quienes puedan avalar y/u exigir el uso de Archivos Abiertos. Las bibliotecas tienen el deber de alentar el uso de esta iniciativa en sus instituciones y de participar activamente en el proceso de adopción de OA a través de campañas de sensibilización para educar tanto a los investigadores como a los responsables de la toma de decisiones, sobre el papel que cumple el modelo y sus ventajas al aumentar la visibilidad e impacto de los resultados de las investigaciones desarrolladas en sus instituciones.

4.1 La exposición de metadatos

Los centros de recursos de AGRIS no son los únicos proveedores de datos en la nueva arquitectura; cualquier institución de investigación y tecnología agraria que quiera publicar sus obras en la red AGRIS puede inscribirse como socio y valerse de las normas generales y las tecnologías de la red. Al proveedor de datos se le debe aclarar que un repositorio o un archivo abierto no es meramente una base de datos con documentos de texto completo y sus metadatos. La Figura 2 muestra los diversos componentes para el suministro de datos: el proceso de captura de contenidos, el lugar donde los investigadores pueden autoarchivar sus obras; la catalogación por los gestores de información, y el despliegue de los metadatos en distintos formatos. La red AGRIS ha definido tres tipos de mecanismos para el suministro de datos: un Repositorio Dinámico o capa de software que permite que los repositorios cumplan las normas de la OAI y que actúan como intermediarios entre los archivos abiertos y los proveedores de servicios basados en la OAI. La capa acepta solicitudes de la OAI y genera respuestas dinámicas; un Repositorio Estático para colecciones pequeñas de metadatos que puede ser generado periódicamente a través de un *script* o secuencia de comandos que extrae información de una base de datos existente (Hochstenbach, Jerez & Van de Sompel, 2003), y un Repositorio Hospedado para repositorios que no pueden desplegar sus metadatos en Internet y que requieren de otras instituciones para hospedarlos y desplegarlos. Un proveedor de datos puede desplegar distintos formatos de metadatos. Estos pueden ser recolectados por diversos proveedores de servicios, ofreciéndoles canales adicionales de salida y mayor visibilidad. El Dublin Core no calificado es el formato básico de metadatos para poder participar en la comunidad de la OAI, si bien aún no responde a cabalidad con las necesidades de la comunidad científica agraria. El Perfil de Aplicación del AGRIS (AGRIS AP) fue

creado a objeto de mejorar la calidad e interoperabilidad de los metadatos en este dominio. Los proveedores de datos pueden ofrecer tanto Dublin Core no calificado como AGRIS AP a fin de lograr mayor visibilidad en un número mayor de motores de búsqueda y proveedores de servicios temáticos tanto en la red AGRIS como en otros medios. La Iniciativa AGRIS fomentará el intercambio de metadatos bajo las normas del AGRIS AP a fin de garantizar una descripción de alta calidad para los recursos de información agrícola.

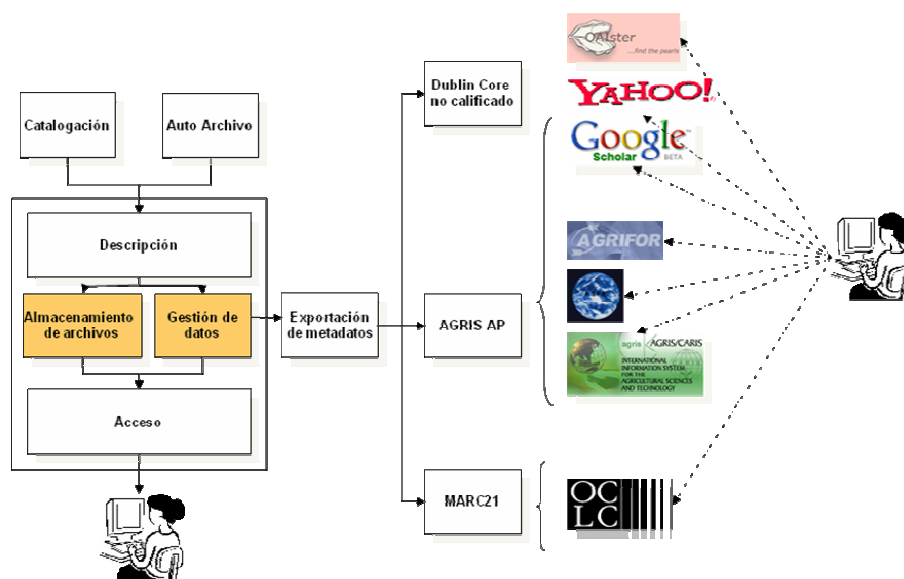


Figura 2. Los proveedores de datos

5 Recolección de metadatos: creando el canal de comunicación

La recolección de metadatos es una solución de interoperabilidad que permite a los proveedores de servicios comunicarse con los proveedores de datos. En este caso, la solución de interoperabilidad consta de la denominación uniforme (metadatos estandarizados), una sintaxis común y un protocolo de recolección de metadatos. El Protocolo para la Recolección de Metadatos de la Iniciativa de Archivos Abiertos (Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH) es un protocolo que se desarrolló para recolectar los metadatos de registros que se encuentran en los archivos abiertos. El OAI-PMH está construido sobre la base de una arquitectura cliente-servidor (los

proveedores de servicios solicitan metadatos a los proveedores de datos/archivos abiertos). En consecuencia, los proveedores de servicios pueden ampliar sus servicios utilizando metadatos de varios archivos. El protocolo OAI-PMH se ha ido empleando cada vez más para establecer un vínculo con los archivos abiertos distribuidos. La razón principal que llevó a su aceptación es la simplicidad de su implementación. Además, se han desarrollado varias herramientas para facilitar la creación, gestión e interoperabilidad de los archivos abiertos que cumplen las normas de la OAI. La denominación uniforme se logra a través de la estandarización de la semántica para metadatos. Se utiliza una sintaxis común, XML, para representar y transportar conjuntos de metadatos específicos de un archivo. La mayor parte del software disponible tiene la capacidad de generar registros XML.

6 Los Proveedores de Servicios: mejorando el acceso

La red AGRIS cuenta actualmente con diversos proveedores, entre ellos, la base de datos AGRIS de FAO, la base de datos AGRIS comercial de Ovid Technologies, el servicio AGRI2000 del Sistema de Información y Documentación Agropecuaria de las Américas (SIDALC), el centro de información virtual del Grupo Consultativo sobre Investigación Agrícola Internacional (CGIAR) y otros proveedores regionales de servicios más pequeños. La nueva arquitectura esboza un ambiente mucho más simple y eficiente para establecer servicios específicos. Los Proveedores de Servicios AGRIS recolectan registros que sólo contienen metadatos, a partir de dos o más archivos abiertos, dependiendo de la envergadura y características de sus colecciones, y no documentos de texto completo. La facilidad con que un proveedor de servicios pueda recolectar datos con éxito dependerá de los esquemas de metadatos del proveedor, de sus colecciones de metadatos y la frecuencia con que actualiza el archivo abierto, el acervo de archivos abiertos, las condiciones de uso de los registros, si se menciona la fuente, y el vínculo que haya entre el proveedor de servicios y el proveedor de datos. Foulonneau y Dawson (2003) enumeran los principales puntos a considerar para la implementación de un proveedor de servicios, la mayoría de los cuales ya han sido emprendidos por la red AGRIS, estos son, la recopilación y análisis de datos (la mayoría de los miembros de AGRIS ya está aportando a la base de datos de AGRIS), esquemas de metadatos comunes (el AGRIS AP se ha convertido en la norma establecida para

metadatos para intercambiar información agrícola), el nivel de detalle de los descriptores (el AGRIS AP proporciona metadatos calificados para describir de manera más delimitada los documentos agrarios y cuenta con esquemas específicos relacionados a la agricultura), cruces de metadatos (existen cruces del AGRIS AP con otros esquemas de metadatos), temas relacionados con la terminología (se han desarrollado y empleado varios SOC en ciencias agrarias) y un entorno multilingüe, una de las características principales de la red AGRIS.

Los metadatos recolectados pueden proporcionar textos completos asociados y el acceso a éstos es otorgado por los proveedores de servicios. Entre los diversos proveedores de servicios previstos se incluyen los portales temáticos, las bibliotecas digitales y portales o páginas web que sólo despliegan metadatos. Entre los servicios que se pueden implementar empleando los datos recolectados se incluyen: la indexación de datos; las búsquedas avanzadas y navegación; las estadísticas de uso; el análisis sintáctico de citas bibliográficas; la posibilidad de comentar documentos; soporte en varios idiomas; la autenticación de usuarios; lectores automatizados de noticias (aggregated news feeds); foros electrónicos; notificaciones automatizadas por email, o sea, un servicio de Difusión Selectiva de Información (Selective Dissemination of Information, SDI); sistemas automatizados de envío de noticias (RSS) sobre conferencias, eventos y cursos de diversos disciplines, y servicios personalizados (que el mismo usuario adapta conforme a sus necesidades).

7 La interconexión para la interoperabilidad: el sitio web de AGRIS y su Directorio de Archivos Abiertos

El sitio web de AGRIS es un canal de comunicación que conecta los datos y el proveedor de servicios con un determinado usuario. Es un punto de enlace entre diversos recursos de información que consta de un depósito para los productos de las Unidades Operativas, el directorio de Archivos Abiertos de AGRIS, las normas y herramientas para la gestión de datos y cualquier otra información relevante. El sitio web cumple un papel esencial en la arquitectura propuesta para la red AGRIS y responde al objetivo de fomentar su uso. El sitio web es una herramienta fundamental para promover una visión amplia de coherencia para la información agrícola en el contexto de la nueva iniciativa AGRIS.

El Directorio de Archivos Abiertos de AGRIS es la plataforma de registro para todos los archivos abiertos de AGRIS. Se podrá utilizar para hacer búsquedas en el sitio web de AGRIS y para identificar los proveedores de datos que forman parte de la comunidad. El directorio clasificará en categorías y enumerará una amplia gama de archivos abiertos empleando conjuntos de metadatos que proporcionarán la información pertinente. Se podrán realizar búsquedas por país, ubicación, tema, formato de metadatos, software, tipo, material incluido y otros criterios de búsqueda.

8 Calidad garantizada en la creación de metadatos

La descripción del contenido mediante la creación de metadatos de calidad y la utilización de terminologías estándar es la piedra angular de una gestión de contenidos eficaz y el desarrollo de servicios con valor agregado. Las normas para metadatos como el Dublin Core (DC) y el Perfil de Aplicación del AGRIS (AGRIS AP) aportan mecanismos para compartir información en forma estandarizada al recomendar el uso de una semántica común y sintaxis interoperables. Las normas de la OAI exigen la utilización del esquema de metadatos Dublin Core no calificados para desplegar metadatos a través del Protocolo para Recolectar Metadatos de la OAI (OAI-PMH), a pesar de que no cumplen a cabalidad con los requisitos de la Comunidad Científica Agrícola. Al utilizar Dublin Core no calificado, se dificulta el desarrollo de servicios con valor agregado eficaces a raíz de la mala calidad y pérdida de datos, entre otras razones. Como fuera mencionado anteriormente, el perfil AGRIS AP, una norma para definiciones precisas, fue creado para mejorar la calidad de los metadatos y para permitir un mayor intercambio de ellos entre la comunidad científica agrícola. La calificación del elemento de los metadatos permite agruparlos de mejor manera y recuperar los contenidos posteriormente. Los metadatos estandarizados, en términos de la sintaxis y la semántica empleada, facilitan la interoperabilidad, mientras que los metadatos de mala calidad pueden convertir un recurso dentro del repositorio en algo prácticamente invisible que terminará no siendo utilizado (Barton, Currier & Hey, 2003). Los proveedores de servicios temáticos más especializados utilizarán el perfil AGRIS AP como el formato básico de metadatos para recolectar metadatos.

A fin de garantizar un nivel mínimo de calidad será necesario consensuar algunos campos de metadatos y el uso de vocabularios controlados pertinentes. El perfil AGRIS AP logra esto al exigir el uso de cinco elementos esenciales y fomentando el uso de vocabularios controlados de términos agrícolas tales como AGROVOC, los códigos CABI, etc.

8.1 El Perfil de Aplicación del AGRIS: ajustándose a los requerimientos agrícolas

El Dublin Core es un formato de metadatos cuya estructura fue establecida como resultado de un consenso internacional. El Conjunto de Elementos de Metadatos Dublin Core (DCMES) consta de quince elementos -usados para describir recursos de manera simple y poder recuperarlos- cuyo uso es recomendado pero no obligatorio. El DC se ha ampliado para incluir otros elementos opcionales, elementos delimitadores y vocabularios. El Dublin Core no calificado es el formato para metadatos que permite asegurar la interoperabilidad básica necesaria para el Protocolo OAI-PMH. No obstante, probablemente será necesario emplear otros descriptores para metadatos para ciertas áreas temáticas y comunidades, por ejemplo, para describir recursos con estructuras complejas de manera especializada, como es el caso de la red AGRIS (Liang, Salokhe, Sini & Keizer, 2006). El perfil AGRIS AP fue diseñado para el dominio agrícola conforme a sus características y requerimientos específicos. Es una norma creada específicamente para mejorar la descripción, intercambio y eventual recuperación de objetos informativos digitales agrícolas (Document-Like Information Objects, DLIOs). El perfil AGRIPS AP es un formato que permite compartir información a través de sistemas bibliográficos dispersos. Este perfil de aplicación acota aún más el conjunto de elementos Dublin Core simple; extensiones locales consideradas necesarias para una descripción exhaustiva de los recursos de información agrícola. La norma, desarrollada a tenor de la nueva visión de AGRIS, mejora la calidad de la descripción de los recursos de información agrícola, lo que posibilita a los proveedores de servicio procesar una mayor cantidad de información.

El hecho de que los proveedores de datos utilicen el perfil AGRIS AP dentro de la comunidad AGRIS conlleva ciertas ventajas para los proveedores de servicios ya que define una capa más completa de

interoperabilidad que permite revelar los recursos secundarios y terciarios mediante el uso de un conjunto calificado de elementos; proporciona mayor flexibilidad al entregar resultados de búsquedas más relevantes a través de búsquedas dirigidas; proporciona metadatos variados y calificados; entrega información adecuada respecto al contenido del recurso a través de la utilización de calificadores específicos al ámbito agrícola, entre ellos, esquemas de clasificación, tesauros, etc.; y permite examinar recursos por tema, país, año, etc. A pesar de que, para ajustarse a las normas “oficiales” de la OAI, los proveedores de datos deberán cumplir las normas del Dublin Core simple, para efectos de la red AGRIS a través de la cual se proporcionan "servicios temáticos", se recomienda que los proveedores de datos también cumplan las normas del perfil AGRIS AP.

8.2 Sistemas de Organización del Conocimiento: apoyo para la navegación semántica

Los Sistemas de Organización del Conocimiento (SOC) son estructuras organizadas de conocimientos. Algunos ejemplos de estos son: ficheros de autoridad, sistemas/esquemas de clasificación, mapas conceptuales, listados controlados, diccionarios, fichas de datos, glosarios, ontologías, encabezados temáticos y diccionarios geográficos. Estos recursos mantienen un flujo continuo, según el grado de precisión de su semántica y posibilidad de ser interpretado de manera automatizada. Los SOC son palabras o frases extraídas de un conjunto estandarizado que permite zanjar dos cuestiones básicas: cuando existen dos o más palabras para describir el mismo concepto, como por ejemplo, Embarcación pesquera/Galera pesquera o Riesgo para la salud/Peligro para la salud; o cuando dos o más palabras se escriben de la misma manera pero se usan para describir conceptos distintos, por ejemplo, galera (nave)/ galera (sombbrero) o Ling (brezo)/ Ling (un tipo de bacalao). Los SOC permiten crear servicios más específicos como búsquedas por palabras clave o por país de cobertura y búsquedas por tipo de documento (patente, libros, etc.), y circunscribir los resultados de las búsquedas a uno o más años específicos, ciertos idiomas o navegación semántica dentro del conjunto de resultados basado en las palabras clave definidas en vocabularios controlados. Estos servicios se habilitan mediante el uso de vocabularios controlados y su referencia explícita en los datos compartidos, situación que se da en el perfil AGRIS AP. La utilización de formatos automatizados como el XML y

los vocabularios estándar para el intercambio de metadatos son la piedra angular de la futura web semántica en el ámbito agrícola. Son los componentes básicos sobre los cuales se pueden crear las aplicaciones que aprovechan las ventajas de la expresividad semántica en los metadatos bibliográficos para mejorar el acceso a información agrícola.

9 Conclusiones y etapas siguientes

Esta ponencia formula una propuesta de arquitectura para archivos abiertos dentro de la Red AGRIS. Dicha arquitectura conjuga la experiencia e historia de la Red AGRIS con el nuevo paradigma de publicaciones académicas de Archivo Abierto y la Iniciativa internacional de Archivo Abierto. La implementación de esta arquitectura creará un lugar en la web donde miles de proveedores de datos podrán subir su material y sobre el cual se basarán muchos servicios. Esto podría llegar a transformarse en un gran logro, comparable sólo con otros grandes proyectos como PubMed en el ámbito de la medicina. Asimismo, se requerirá una inversión mucho menor para desarrollar normas y herramientas universales, herramientas de código abierto (*open source*) para la gestión de datos y metadatos para proveedores de datos, esquemas y vocabularios para datos, y conjuntos de herramientas para garantizar que los actuales sistemas cumplan las normas de la OAI. Se ha avanzado mucho en todos estos ámbitos, gracias en parte a la Red AGRIS y fundamentalmente a su esquema de metadatos DC y las herramientas universales del OAI, el perfil de aplicación AGRIS basado en el DC, los vocabularios controlados o sistemas de organización de conocimientos como AGROVOC, el conjunto de herramientas de gestión de metadatos de WebAGRIS, la adaptación de DSpace a los requerimientos específicos de la Red AGRIS y el desarrollo de un motor de búsqueda de código fuente abierto AGRIS (basado en Lucene) para proveedores de servicios.

El siguiente paso es lograr un consenso entre los principales asociados a la Red AGRIS respecto a su adaptación y la posterior planificación de proyectos individuales que aborden componentes específicos de la arquitectura y, finalmente, la presentación de proyectos individuales en diversos países y regiones miembros de la red.

Referencias

- [1] Segunda Consulta sobre Gestión de Información Agraria (2002) *AGRIS – Estrategia para una red internacional de información sobre ciencias y tecnología agrícola en el marco del WAICENT*. <http://www.fao.org/DOCREP/MEETING/005/AC502E.HTM> Ultimo acceso: enero 2007.
- [2] Budapest Open Access Initiative. <http://www.soros.org/openaccess/read.shtml> Ultimo acceso: enero 2007.
- [3] Eprints: *Supporting Open Access. Self-archiving FAQ*. <http://www.eprints.org/openaccess/self-faq/#self-archiving> Ultimo acceso: enero 2007.
- [4] Van de Sompel, Herbert; Lagoze, Carl (2002) *The Santa Fe Convention of the Open Archives Initiative*. Revista D-Lib, 6(2). <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html> Ultimo acceso: enero 2007.
- [5] Sitio web de Agricultural Information Management-AIMS (Normas para la gestión de información agraria). <http://www.fao.org/aims/> Ultimo acceso: enero 2007.
- [6] Hochstenbach, Patrick; Jerez, Henry; Van de Sompel, Herbert (2003). *The OAI-PMH Static Repository and Static Repository Gateway*. JCDL 2003. <http://public.lanl.gov/herbertv/papers/jcdl2003-submitted-draft.pdf> Ultimo acceso: enero 2007.
- [7] Foulonneau, Muriel; Dawson, David. (2003) *Expert Report 3 – Open Archives Initiative –Protocol for Metadata Harvestin - Practices of cultural heritage actors*. http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf Ultimo acceso: enero 2007.
- [8] Barton, Jane; Currier, Sarah; Hey, Jessie M. N. (2003) *Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice*. Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications Washington (USA). <http://eprints.rclis.org/archive/00001972/> Ultimo acceso: enero 2007.
- [9] Berners-Lee, T. and Connolly, D. (1995). *Hypertext Markup Language - 2.0* http://www.w3.org/MarkUp/html-spec/html-spec_5.html Ultimo acceso: enero 2007.
- [10] Liang, Anita; Salokhe, Gauri; Sini, Margherita; Keizer, Johannes (2006) “Towards an infrastructure for semantic applications:

Methodologies for semantic integration of heterogeneous resources”
que se publicará en la revista *Cataloging & Classification Quarterly*.