Gestión de la calidad y la preservación en repositorios institucionales

CONFERENCIA MAGISTRAL

4 DE OCTUBRE DE 2022

MARISA RAQUEL DE GIUSTI

Universidad Nacional de La Plata (UNLP) (Argentina)

Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) (Argentina)

RESUMEN EXTENDIDO

La conferencia hace un recorrido por los conceptos de repositorios institucionales y de la preservación digital, considerando que los repositorios institucionales son depositarios de la producción de una institución, cualquiera sea su tipología de acuerdo a lo que la propia institución determine. Y que la preservación digital es el conjunto de estrategias, procesos y técnicas que dan respuesta a los problemas que plantea la conservación de los materiales digitales y de los medios (hardware y software) que se emplean para su almacenamiento y consulta, y que en general está destinada a mantener los objetos digitales y sus características de acceso a largo plazo.

Cuando se piensa en preservación digital, se debe hacer referencia como estándar al Modelo OAIS y la norma ISO 14721: ajustar las funciones del repositorio. Lo que plantea es un modelo abstracto que tiene seis entidades: ingesta, gestión de datos, almacenamiento de archivos, preservación, administración y acceso. Esas seis entidades que van desde ingresar el paquete de información, es decir, el contenido que se desea ingresar al repositorio a través de un autoarchivo o bien por operaciones de la administración, u operaciones automáticas de naturaleza informática que toma la forma de un zip.

Durante el proceso se genera un paquete de preservación que se llama AIP, se realizan funciones como separar todos los elementos que son la información descriptiva y enviarla a esta otra entidad que se llama gestión de datos a fin de coordinar el acceso para cuando alguien solicite una información.

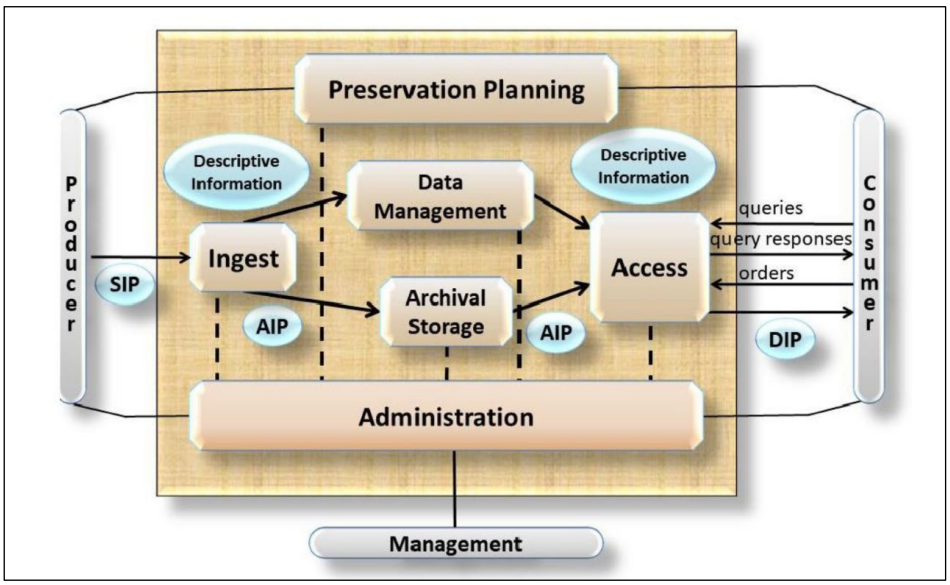

FIGURA 1: OAIS Functional Entities. Management Council of the Consultative Committee for Space Data Systems (CCSDS). (2019). OAIS final v3 draft with changes wrt OAISv2 20190924-rl.docx. P. 41

El centro de todo: el IP del modelo OAIS

La información descriptiva de la preservación va a tener que contener ciertos elementos, metadatos e información que tiene que ver con la referencia a ese objeto digital y está directamente relacionado con la manera unívoca en que se identifica, por ejemplo, un identificador persistente tipo handle, DOI, etc.

La procedencia (o *provenance*) tiene que ver no solamente con el origen del archivo, sino con los sucesivos tratamientos que se le dieron, es decir, la información con la que está relacionado ese objeto de información. A su vez, *fixity* se refiere a la información de fijeza que asegure que ese archivo no es

alterado; por su parte los derechos se refieren a la licencia para dar difusión a ese contenido, así como la licencia de uso que se le otorgue.

En la imagen siguiente se podrá notar que la información de contenido, si se mira a la izquierda de la figura, se trata de formatos, y a la derecha, se visualizan los metadatos. A su vez, esto se relaciona con tres formas que tomará este paquete de información, mientras se vaya procesando desde la ingesta a las actividades que se realizarán dentro del repositorio, hasta el paquete que se conoce como DIT y que es el paquete que se le entregará a un sistema informático o a una persona que esté solicitando el archivo.

El IP del modelo OAIS

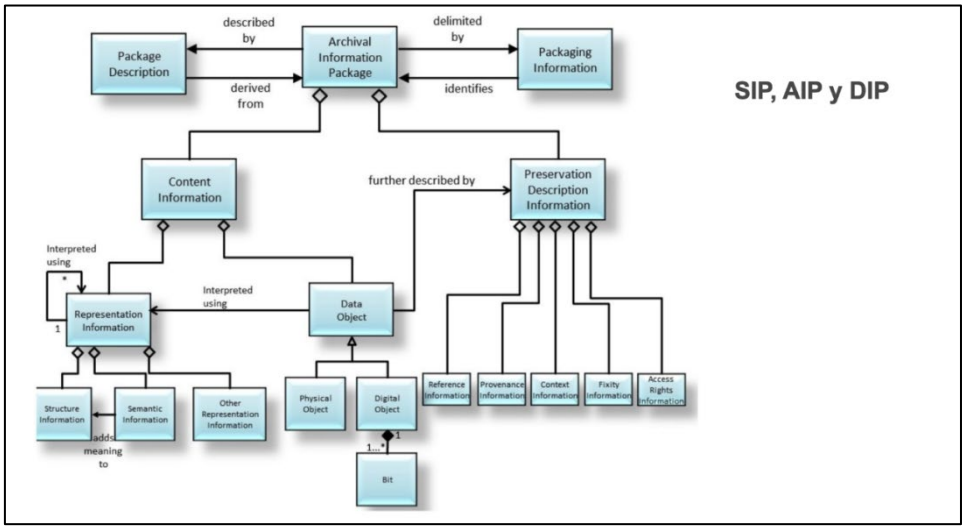

FIGURA 2: Archival Information Package (Detailed View) and its associated Package Description and Packaging Information Management Council of the Consultative Committee for Space Data Systems (CCSDS). (2019). OAIS final v3 draft with changes wrt OAISv2 20190924-rl.docx. P.4-43

El siguiente punto tiene que ver con el trabajo de vigilancia de los contenidos de un repositorio, el cual deberá ejecutarse a lo largo de todo el ciclo, desde que se crea el contenido, se organiza y se clasifica, incluyendo el momento de revisión, en distintas operaciones de verificación e incluso si se elimina el repositorio. Todo ello implica conocer el origen del archivo y si se ejecutaron

otros eventos como migración del archivo original, verificación de su autenticidad por algún tipo de algoritmo, derechos, etc.

Preservación de alto nivel

Lo primero que se debe considerar al hablar de preservación a alto nivel es generar una política de preservación digital; los siguientes puntos se refieren al abordaje sugerido para su realización.

La política de preservación digital de un repositorio debe tener como principal referencia la política institucional de acceso abierto, la cual puede estar en consonancia con la legislación a nivel nacional sobre publicaciones y datos.

Debe establecer la preservación a largo plazo de los contenidos del repositorio.

En lo posible debe hacer referencia a una guía de evaluación de los procesos de preservación en el repositorio (por ejemplo, NDSA), incluso aludir a otros marcos de referencia, por ejemplo, recomendaciones de COAR.

Debe contener un apartado de “Alcances y propósitos”.

Para la creación de un plan de preservación se consideran los siguientes puntos: el plan debe exponer los motivos, principios y sobre qué contenidos va a centrarse para garantizar la conservación, el acceso y la comprensión a largo plazo de esos fondos. Además, se deben identificar necesidades y prioridades y aportar un cronograma pormenorizado que muestre la distribución de las tareas en el período de vigencia del plan.

En cuanto a temas de calidad, para ello es necesario pensar en cómo los repositorios sirven a las personas usuarias, investigadoras, docentes, ¿qué se percibe? ¿qué servicios se brindan? Y en los parámetros e indicadores tanto cualitativos como cuantitativos para objetivar esa percepción de calidad. Considerando las necesidades o expectativas de grupos distintos: cómo son

las diversas comunidades de usuarios y otros sistemas o tecnologías con las que hay que asegurar una interoperabilidad.

Forma de evaluación/autoevaluación/revisión

Sin importar el tamaño o propósito del repositorio, se debe alentar la utilización con una lista de verificación como una herramienta para la evaluación objetiva, interna o externa. La auditoría es la base para comparar las capacidades con un conjunto de criterios centrales para un repositorio digital confiable. La certificación es un paso adicional que algunos repositorios tomarán y/o deberán tomar para el reconocimiento formal y objetivo. El resultado de cualquier auditoría debe verse en el contexto en el que se realizó. Algunos ejemplos de tipos de certificación son DINI (2006) de origen alemán y la ISO 16363.

Estructura organizativa

En términos institucionales, se hace referencia a seis puntos claves a considerar para la creación y el correcto funcionamiento de los repositorios. El primero de estos se refiere a la viabilidad de la organización y su gobierno que considera la misión, el plan estratégico de conservación y la política de colección de los repositorios.

Seguidamente, se plantea la importancia de la estructura organizativa y provisión de personal, cuya base es la identificación de las responsabilidades, la importancia de personal competente, un plan que refleje las funciones de cada parte, y un programa de desarrollo profesional para el personal.

En cuanto al marco de procedimiento de responsabilidad y política de conservación debe tomar en cuenta: la comunidad científica, las políticas de conservación, la historia documentada de los cambios, la transparencia y rendición de responsabilidad, así como las mediciones de integridad de la

información, y finalmente, estar comprometido con un programa regular de autoevaluación y/o certificación externa.

Para el cuarto punto considera la sostenibilidad financiera que involucre la planificación económica a corto y largo plazo, las prácticas financieras y procedimientos transparentes y el análisis e informe de riesgos.

Como quinto y último aspecto de este apartado, se presentan los contratos, licencia y pasivos, que integran los contratos de depósito, la adquisición, mantenimiento y acceso y retirada, las políticas de conservación de contenido, y las políticas de propiedad/derechos.

Autoevaluación NSDA

Para los procesos de autoevaluación una buena práctica es considerar el almacenamiento (copias y localización), la integridad de los datos, la seguridad de la información y los responsables, los metadatos y los formatos.

A su vez existen cuatro niveles para el almacenamiento (copias y localización) y estos tienen una complejidad creciente, estos son: proteger, conocer, controlar y reparar los datos.

Seguidamente, es importante realizar una revisión tanto de prácticas y procedimientos como de contenidos y formatos. Y en cuanto al resguardo de datos, escribir un plan de riesgos que considere la ciberseguridad, fallas de hardware, errores humanos, desastres naturales, etc. El requerimiento de copias de seguridad se define según su cantidad, rotación y ubicación.

Lo anterior se complementa con el abordaje de acciones de preservación en cuanto a integridad (análisis de checksum), el control de cambios (permisos y grupos), los metadatos (uso de identificadores persistentes, formatos estándar, vocabularios controlados y metadatos técnicos), el control de contenido (actualizaciones, migraciones, formatos, virus).

Como conclusiones, la importancia de NDSA es que permite la autoevaluación y porque es relativamente simple. Hay acciones que exceden a

los objetos digitales y la infraestructura, por ejemplo, acciones que hacen a lo organizacional, que no cubre NDSA, básicamente ahí debe escalarse a una norma como la ISO 16363 que además permite certificar el repositorio, pero es bastante compleja.

Como puntos recomendados, se debe definir documentación prioritaria, controlar la ejecución de backups, mejorar la trazabilidad y los metadatos, y valorar la importancia de la política de formatos y de preservación.

Ver conferencia en YouTube