

Segunda Circular, 19 agosto 2010

"Estimados colegas,

Tras ocho meses de virtual silencio sobre el tema de la patrón de especies --solo roto por contactos bilaterales breves pero interesantes-- y a menos de 3 meses para nuestra reunión en Pamplona, rompo este silencio con un mensaje me temo que largo.

En concreto, me pongo en contacto con vosotros para presentaros algunas (1) noticias, (2) ideas de desarrollo, (3) peticiones específicas y alguna (4) propuesta.

(1) De la parte de las noticias. Aquí constatar que durante el año hemos hablado algunos de vosotros y reincidente que la necesidad no solo existe, si no que corre prisa el avanzar en una solución.

He participado en reuniones de dos iniciativas que vienen sobre todo de la mano del MARM e implementadas por TRACSA Estas son:

- Definición de metodología para el cumplimiento de objetivos de la GSPC (Global Plant Conservation Strategy)
- Establecimiento de la metodología del sistema de seguimiento de la biodiversidad española.

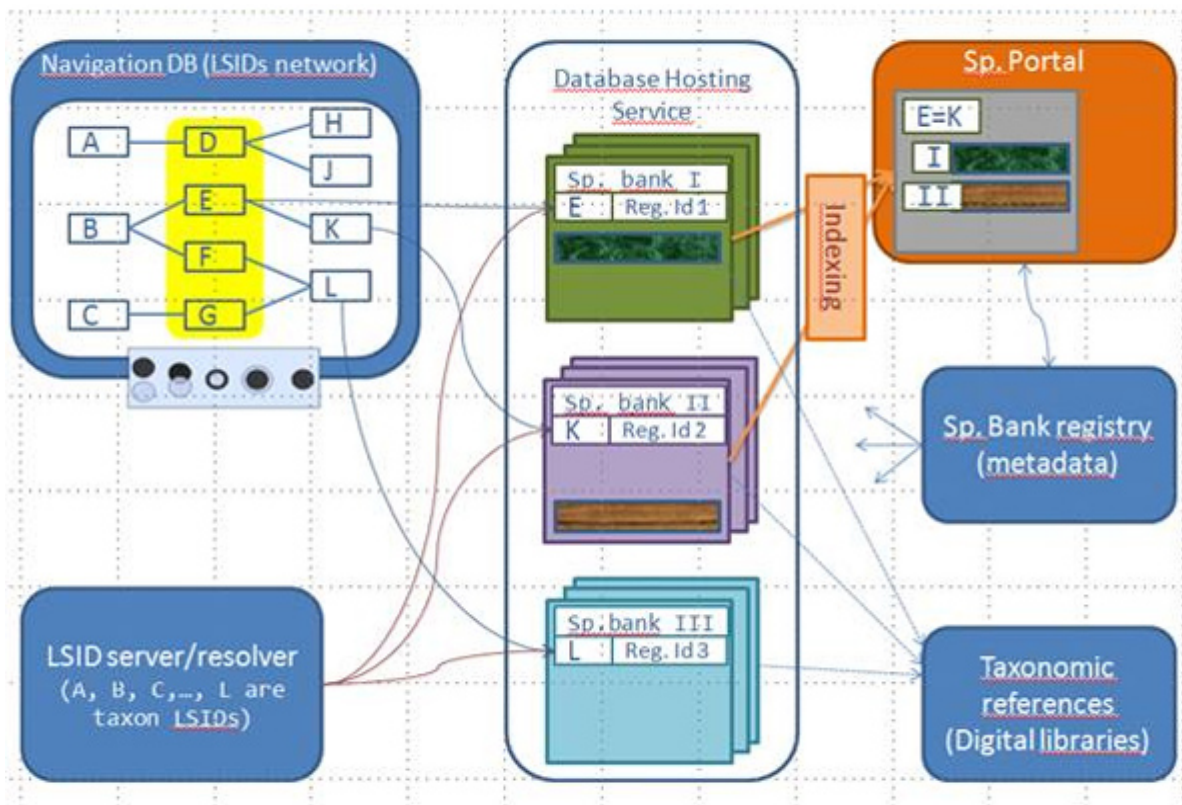
En las dos bosquejé la idea de desarrollo que os presento ahora más abajo.

(2) Esta idea de desarrollo si bien no es copia de nada que exista (hasta donde yo sé) si que se basa fuertemente en proyectos y desarrollos punteros y activos en el campo de la informática de la biodiversidad y en los foros más cercanos a GBIF y al TDWG (www.tdwg.org)

De manera muy resumida y recapitulando, decir que apuntamos a construir un sistema que permita publicar y utilizar información de especies; que también permita establecer equivalencias entre distintas listas y catálogos de especies y que haga posible combinar información de distintos catálogos de acuerdo a relaciones de equivalencia entre las fichas de los mismos.

También se plantea como objetivo (o especificación) de este sistema el que los catálogos que conecten puedan estar distribuidos y administrados allá donde se compilen (los distintos proyectos o administraciones que los tengan), que soporte versiones sucesivas (e.g. lista de aves del 2009, del 2011), también queremos que funcione como infraestructura, esto es, que lo que se crea tenga vocación de permanecer en el tiempo y que se le dote de las necesarias "pasarelas" (en la práctica web services o APIs) para que aplicaciones y sistemas externos puedan utilizar la información a la que da acceso.

Con esta idea, el sistema que planteo se esquematiza en el siguiente gráfico:



Los componentes son:

Los "**Species banks**" (o catálogos). En su forma más simple serían listas de nombres con referencias a una fuente; en su más completa, catálogos de fichas de especies con descripciones, información de de manejo, legislación, imágenes, etc. El estándar de intercambio que propongo para estas fichas es el Plinian Core:

<http://www.gbif.es/plinian/doku.php?id=inicio>

En una fase inicial estos tendrían que estar alojados en una misma máquina, para simplificar su **indexación** por parte del portal. Más adelante se podrían implementar en una arquitectura realmente distribuida.

El **portal** es que permite visualizar, buscar y combinar la información de las "fichas", navegar por los catálogos.

La referencia para esto sería

<http://psybugs.biota.biodiv.tw/>; de apariencia es similar a otros, pero con la plataforma tecnológica ideal (moderna, modular, basada en Drupal y en LAMP) y con módulos desarrollados específicamente para manejo de nombres, fichas de especies y clasificaciones disponibles. Estos están desarrollados por el proyecto EoL www.eol.org. Además ya he invitado al desarrollador de Taiwán y le tengo por aquí en octubre.

El catálogo de **metadatos** de los species banks del sistema. Para tener la información de que es lo que hay, de quien es, que cobertura (taxonómica y espacial) tiene cada catálogo, cuando fue la última actualización, etc. Un ejemplo de metadatos (aunque para citas) puede ser:

<http://es.mirror.gbif.org/datasets/resource/252/>

Las **referencias taxonómicas** son fundamentales para hacer que el sistema no dependa de unos pocos humanos que ahora están y luego no tiene por qué, que todas las fichas de especies del sistema tengan una referencia explícita (descripción) y estable (publicación). Este módulo necesario, la manera de implementarse abierta a discusión.

Cada ficha ha de tener un identificador único y permanente. Lo mismo cada taxon. La tecnología recomendada para asignar estos identificadores es la de los LSIDs (Life Science Identifiers). El sistema funciona parecido a los códigos de barras de los productos. Tu necesitas un identificador para tu taxon o ficha, lo registras y te lo dan. Ahora desde la Unidad de GBIF estamos explorando lo que su pone montar un **servidor de LSIDs** para dar servicio a las entidades que gestionan y producen información sobre biodiversidad en nuestro ámbito. El servicio funcionaría de manera automatizada (no habrá señor en una ventanilla al cargo del registro...).

De referencia sobre el tema paso el siguen enlace y el adjunto que va a este mensaje <http://www.gbif.org/communications/news-and-events/showsingle/article/call-for-review-of-gbif-lsid-guid-task-group-report/>

Y el componente garantiza la interoperabilidad entre los catálogos es la base de datos de **la red de LSIDs**. Es aquí donde se establece que dos especies de dos catálogos son la misma, o que una incluye a otra, o que esta especie del catálogo del 2009 se ha dividido en dos en el del 2011, etc. Esta bases de datos aunque contenga y maneje los LSIDs usados en los catálogos de especies, es independiente de estas. Este módulo implica la construcción del interfaz web para establecer las relaciones con un relativamente complejo sistema de control de acceso para que los expertos puedan establecer los mapeos entre catálogos. En esa red de LSIDs es donde se define "la lista patrón" o "las listas patrón" según los distintos contextos; "tirando" de las relaciones se saca la información asociada, o por ejemplo, se traduce una lista en otra.

(3) Peticiones.

La primera petición es preguntaros que os parece el planteamiento, que os gusta y que no; que echáis en falta y no entendéis, etc.

Con el planteamiento de desarrollo presentado, hay varios frentes abiertos y en los que vuestra participación es requerida y puede ser crucial para tener al final "un producto" --que también conlleva metodología de trabajo—útil y que de verdad contribuya a que la gestión que hacemos de las especies sea más sencilla, nos resulte útil para más cosas y ganemos en interoperabilidad.

Estos frentes se pueden presentar como sigue:

Casos de uso	Esto es ejemplos de cómo estas listas de especies se usan y combinan. Los más básicos son URLs con fichas de especies
Metadatos	Aquí vendría muy bien saber que sistemas tenéis para estos catálogos y si usáis estándares, cuales. La opción por defecto es la de GBIF, que as u vez se basa en el "Dublin Core (DC)", pero por una parte DC se nos va a quedar corto y por otra eso no significa que sea la solución final
Capacidades informáticas	Estaría bien empezar a saber hasta qué punto podemos implicarnos cada uno en esto. De momento, estoy preparando el borrador de una propuesta al Ministerio de Ciencia e Innovación sobre todo para el tema del portal y de los LSIDs.
Datos de pruebas	Tengo para cacharrear fichas de catálogo de Nacional de Especies Amenazadas, y fichas de algunos grupos de "Fauna ibérica", "Flora iberica" y "Flora Micológica Ibérica"... puede venir bien tener alguna de, por ejemplo, nombres vernáculos, manejo o legislación

(4) Creo que de aquí a noviembre podemos discutir y aclarar dudas (no en mucha profundidad, porque en remoto y con nuestras agendas, soy consciente de que cuesta, pero que por lo menos sirva para construir un entendimiento común) y, por ejemplo, llegar un día antes a Pamplona para la reunión (en primicia y en pruebas: http://www.gbif.es/administraciones_ambientales_2010.php), juntarnos este grupo y echarlo en acordar cómo y por donde tirar para hacer un prototipo funcional unos pocos meses después.

En las próximas dos semanas no voy a ver email todos los días per si me propongo aclararos --y aclárame-- de "lo que va aquí" y de lo que hace falta.

Por favor todo lo que os provóquelo que aquí os planteo, mandadlo a todos los destinatarios de este mensaje. Aunque sea para decir "vaya patata". Es la única manera de que todos sepamos que terreno pisamos.

Aquí vamos pues.

Un saludo muy cordial,

Paco"

Francisco Pando
GBIF.ES, Nodo Nacional de Información en Biodiversidad
MICINN-CSIC
pando@gbif.es
www.gbif.es