



REVISTA DE PRIVACIDAD Y DERECHO DIGITAL



ENRIQUE ALONSO GARCÍA

"Interoperabilidad legal" y otros problemas jurídicos derivados del manejo de "datos científicos" como "big data" por servicios web e infraestructuras de la economía del conocimiento: el ejemplo de los datos sobre diversidad biológica.

AÑO I • NÚM. 3 • ABRIL 2016

<http://www.rdu.es/articulos/revista/RPDD>

rdu@rdu.es / +34 91 574 64 11



“INTEROPERABILIDAD LEGAL” Y OTROS PROBLEMAS JURÍDICOS DERIVADOS DEL MANEJO DE “DATOS CIENTÍFICOS” COMO “BIG DATA” POR SERVICIOS WEB E INFRAESTRUCTURAS DE LA ECONOMÍA DEL CONOCIMIENTO: EL EJEMPLO DE LOS DATOS SOBRE DIVERSIDAD BIOLÓGICA.

Enrique Alonso García^{1 2}

I.- INTRODUCCIÓN

El presente artículo, redactado por Don Enrique Alonso García ha sido publicado, con autorización expresa de su autor, en el número 3 de la Revista de Privacidad y Derecho Digital, accesible a través del siguiente enlace (<http://www.rdu.es/PruebaGratis/listado>) y licenciado bajo una licencia internacional Creative Commons Reconocimiento-NoComercial 4.0 Internacional License.

¹ Consejero Permanente de Estado. Investigador Honorífico del Instituto Franklin de la Universidad de Alcalá y del Middlebury Monterey Institute of International Studies de California. Co-chair del RDA-CODATA Legal Interoperability Interest Group y miembro de la Task Force on Data Management Principles del Global Earth Observation System of Systems (GEOSS).

² El presente artículo es la versión académica más completa, con argumentación y citas adicionales, y en idioma castellano, del que, con naturaleza de ponencia-marco sobre cuestiones jurídicas e interoperabilidad legal, se elaboró en diciembre de 2015 para el proyecto “**GLOBAL Infrastructures for Supporting Biodiversity research**” (GLOBIS-B), del programa marco H2020 de la Comisión Europea Grant Agreement nº 654003, está publicado con el título “Legal Background and legal policy issues for the determination of Essential Biodiversity Variables through open web services and e-science infrastructures”, que puede descargarse en http://www.globis-b.eu/files/intranet/Deliverables/Final/D4_1%20final.pdf. También incluye, por adelantado, elementos del artículo colectivo titulado “Introduction to the Law of Open Web Services based on Research Data” dedicado a datos tanto de biodiversidad como de biomedicina al que se refiere la nota 9.

El proyecto GLOBIS-B³ pretende articular los mecanismos de

³ En el proyecto “**GLOBAL Infrastructures for Supporting Biodiversity research**” (GLOBIS-B), www.globis-b.eu, participan (vi cuadro inferior en esta misma nota) seis institutos de investigación y universidades europeas (el Instituto Franklin de la Universidad de Alcalá es la institución española y lleva precisamente los trabajos relacionados con la estructura jurídica general y de gestión de la arquitectura técnica de la colaboración entre infraestructuras y con proveedores y usuarios de datos, así como con las instituciones internacionales que se ocupan de los indicadores de biodiversidad: la Convención de Diversidad Biológica, el IPBES, el Partnership for Biodiversity Indicators, el PNUMA, el GEOSS, etc). Y lo llevarán a cabo 12 “e-science research infrastructures” especializadas en datos de biodiversidad, tres de ellas internacionales (GBIF, GEOBON y WFCC-MIRCEN) y nueve supranacionales y nacionales de distintos continentes (2 de la UE –Elixir y LifeWatch-; 2 de la República Popular de China -Academia de Ciencias con sede central en Beijing e Instituto de Germoplasma de Kunming-; 2 de Estados Unidos –NEON y DataOne-; 1 de Australia; 1 de Brasil; y 1 de la República de Sudáfrica). España acaba de ratificar su tratado de incorporación a Elixir y la otra europea. LifeWatch, que tendrá su sede central (Core T) y administrativa en España, se pondrá en marcha en este mismo año de 2016 (vid. pg. 11 de la recentísima publicación el pasado 26 de febrero de descripción de las nuevas infraestructuras de investigación de la UE, ESFRI Landmarks (Part 2), http://people.iom.cnr.it/rossi/20160229_LANDMARKS.pdf, en su página 11, 5 del pdf para Lifewatch y 15, 7 del pdf, para ELIXIR).

Son las siguientes:

Acronym	Organisation	Geographic scope	Website
<i>Project partners</i>			
UvA	University of Amsterdam (Institute for Biodiversity and Ecosystem Dynamics)	Netherlands	http://ibed.uva.nl/
CU	Cardiff University (School of Computer Science and Informatics)	UK	http://www.es.cf.ac.uk/
GNUBILA	gnúbila France	France	https://gnubila.fr/
CNR	Consiglio Nazionale delle Ricerche (Institute of Biomembranes and Bioenergetics)	Italy	http://www.cnr.it/sitocnr/home.html
FI-UAH	Universidad de Alcalá (Instituto Benjamin Franklin)	Spain	http://www.institutofranklin.net/
MLU	Martin-Luther-Universität Halle-Wittenberg (German Centre for Integrative Biodiversity Research i-Div)	Germany	http://www.idiv-biodiversity.de/idiv/research/geo-bon/
<i>Supporting research infrastructures</i>			
Atlas	Atlas of Living Australia	Australia	http://www.ala.org.au/
BC-CAS	Biodiversity Committee of the Chinese Academy of Sciences	China	http://www.kepingma.com/index.html
CRIA	Brazilian Reference Centre on Environmental Information	Brazil	http://www.cria.org.br/
DataONE	Data Observation Network for Earth	USA	http://www.dataone.org/
ELIXIR	European infrastructure for biological information	Europe	http://www.elixir-europe.org/
GBIF	Global Biodiversity Information Facility	Global	http://www.gbif.org/
GEOBON	Group on Earth Observations Biodiversity Observation Network	Global	http://www.geobon.org
GBoWS	Germplasm Bank of Wild Species at Kunming Institute of Botany	China	http://english.kib.cas.cn/
LifeWatch	European Infrastructure for Biodiversity and Ecosystem Research	Europe	http://lifewatch.eu/
NEON	National Ecological Observatory Network	USA	http://www.neoninc.org/
SANBI	South African National Biodiversity Institute	South Africa	http://www.sanbi.org/
WDCM	World Data Centre of Microorganisms at WFCC-MIRCEN	Global	http://www.wdcm.org/

El equipo de la UAH, además de por el autor, está compuesto por los Profs. Drs. José Antonio Gurpegui (male; Prof. of Transatlantic European-United States International

cooperación global entre las más importantes infraestructuras establecidas con el objetivo de dar apoyo a la investigación puntera en materia de biodiversidad, entendida como base científica para la toma de decisiones públicas y privadas relacionadas con la gestión y los impactos potenciales en la biodiversidad. Pretende ser una contribución a la formulación de indicadores globales y otros sistemas de medición cuantitativa y objetiva, necesarios todos ellos para el análisis, seguimiento y gestión de los cambios, a cualquier nivel geográfico – planetario, continental o nacional regional y local- y atendiendo a la dimensión temporal, en el estado de la diversidad biológica. Más en concreto, el proyecto se centra en determinar los servicios web e infraestructuras de e-ciencia de apoyo para medir los cambios, y en particular a partir de Big Data, de las denominadas EBVs (*essential biodiversity variables*)⁴. Por ello, se trata de un proyecto que es también estratégico para la aplicación del GEOSS (Global Earth Observations System of Systems⁵) al campo de la observación de la biodiversidad (GEOBON⁶).

Tiene, pues, el proyecto como tal un muy alto componente tanto científico como técnico de manera que se ocupa, este artículo, sólo de describir la arquitectura jurídica tanto del entrado del servicio web interoperable entre las infraestructuras (*e-science research infrastructures*) internacionales continentales, nacionales, regionales y locales del proyecto como el flujo de datos necesarios para lograr el cálculo de las variables con fiabilidad y calidad suficientes como para

Relations); Juan M. Junoy (male: Prof of Marine Biology) y por los investigadores y técnicos Alfonso Herrera (male; biologist and GIS Master), Paula A. Recarte (female; Bioacoustics technician, Data manager and GIS Master), María Lloset (female; Lawyer) y Beatriz Ramos López (female:biologist, and Nature Conservation and GIS Masters).

⁴ Vid., Kissling, W.D., Hardisty, Alonso García, E., A., Santamaria, M., De Leo, F., Pesole, G., Freyhof, J., Manset, D., Wissel, S., Konijn, J. & Los, W. (2015) Towards global interoperability for supporting biodiversity research on essential biodiversity variables (EBVs). *Biodiversity*, **16**, 99–107; y Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M. & Wegmann, M. (2013) Essential Biodiversity Variables. *Science*, **339**, 277-278.

⁵ <http://www.earthobservations.org/geoss.php>

⁶ <http://geobon.org/>

que puedan cumplir sus fines a todas las escalas temporales y espaciales; es decir, lograr la “interoperabilidad legal” de datos y atender al resto de las resto de cuestiones jurídicas vinculadas a las medidas legales o contractuales que su puesta en marcha y gestión pudiera suscitar.

Aunque, por ello, se obvia en este artículo una descripción amplia de esos detalles científicos y técnicos, debe recordarse, sin embargo, para que el análisis jurídico no sea meramente abstracto, que se trata, como objetivo esencial, de: (1) identificar cuáles son, científicamente, los datos que son relevantes para poder calcular las EBVs; si existen ya o no (y habría que montar en ese caso infraestructuras ecológicas de sensores o medidores para obtenerlos) y si están disponibles ya o no en la web, con especial relevancia a la cuestión de si, de existir, hay series históricas (o subrogados que permitan reprocesamiento) y si las mismas tienen continuidad temporal o permiten modelizar esa secuencia temporal –para apreciar los cambios en esta dimensión–; (2) especificar los requisitos que deben cumplir los usuarios de estos servicios e infraestructuras para acceder, extraer, manejar, visualizar y analizar los datos de biodiversidad de todo tipo que puedan ser accesibles a través de la web o por medios siempre que estén digitalizados; (3) identificar en qué campos pueden las infraestructuras de investigación actualmente existentes o en construcción mejorar su capacidad de manejo de “Big Data” de biodiversidad o desarrollar nuevas capacidades; (4) consensuar soluciones realistas para prestar el máximo apoyo posible a los requisitos mínimos que deben cumplirse para que los usuarios puedan realizar las actividades listadas en el punto 1 previo, de manera que las infraestructuras puedan identificar y poner en marcha servicios web útiles para el cálculo de las EBVs; (5) identificar las mejores prácticas que deben adoptar las infraestructuras; y (6) identificar las cuestiones jurídicas y solucionar los problemas que esta colaboración global entre infraestructuras, a diseñar científica y técnicamente, no constituyan un obstáculo, asegurando la “interoperabilidad legal” de los flujos de datos y en particular diseñar qué tipos de permisos y licencias, términos y condiciones de uso, renunciaciones o declaraciones (*waivers, disclaimers, representations, warranties...*) deben utilizarse por los *virtual research environments* lab o labs (en adelante VREs) en sus relaciones con las propias infraestructuras participantes y con los proveedores de datos a dichas infraestructuras y usuarios tanto de aquéllas como de los servicios o productos derivados de la colaboración, flujos de datos que es claro que no fluyen adecuadamente en este campo, al haberse

identificado ya numerosos cuellos de botella en dichos flujos dadas las limitaciones en el acceso o uso de los datos que se producen por el hecho de estar protegidos mediante derechos de propiedad, sobre todo de propiedad industrial e intelectual, pese a que deberían estar sometidos a principios de “open science” y “data sharing”.

Es, pues, sólo este último aspecto, el punto (6), sin perjuicio de la participación del autor en los restantes campos, el que intenta abordar el presente artículo, debiendo tenerse presente que, si bien los datos de biodiversidad son especiales y por ello necesitan un análisis *ad hoc*, gran parte de los problemas abordados son comunes a toda cuestión tratable a través de metodologías de Big Data en general y, mucho más, a través de los actuales sistemas de servicios de “big data” de imágenes geoespaciales satelitales en general⁷ y, sobre todo, de los servicios de big data biomédicos –incluidos los del microbioma del cuerpo humano– que ya están modificando drásticamente no sólo los servicios públicos y privados sanitarios, y de salud en general, e incluso de la medicina misma como ciencia de precisión individualizada⁸. Tanto los unos como los otros están siendo tratados en estudios y artículos individuales o colectivos del autor de futura publicación, a los que desde ahora me remito⁹.

⁷ Vid nota 5.

⁸ Véanse, en general, por citar la literatura más reciente, Elizabeth Gibney, *The Body Electric: Researchers want to wire the human body with sensors that could harvest dreams of data – and transform health care*, en *Nature* Vol. 528, 3 December 2015, pgs 26-28; o los distintos artículos incluidos en el monográfico *Big Data in Biomedicine*, *Nature Putlook*, Vo. 527, Issue N° 7576, 5 November 2015.

⁹ Para el análisis conjunto de datos biomédicos y de biodiversidad, véase la futura publicación del autor con Anne Bowser del Wilson Institute, que comparará las experiencias al menos de Estados Unidos y la Unión Europea y, más próxima, en unos meses, Enrique Alonso García, Elizabeth Kirby Chris Morris Irene Schluender y Stephanie Suhr, probablemente bajo el título de *Introduction to the Law of Open Web Services based on Research Data* y como resultado de los trabajos en el seno del *BioMedBridges project*, financiado por la Unión Europea. Séptimo Programa Marco, en el seno del *Capacities Specific Programme*, Grant Agreement n° 284209. Para los de datos geoespaciales el futuro trabajo con Paul Uhler et al, desarrollado a partir *The Value of Open Data Sharing*, Noviembre 2015, elevado al GEO-XII Plenary celebrado dicho mes en México D.F. https://www.google.es/?gfe_rd=cr&ei=pajiVq26A6Gs8wefubto&gws_rd=ssl#q=white+paper+geoss+paul+Uhlir+open

El punto de partida es la noción de “interoperabilidad legal”. Ahorramos, dada la materia que suele ser objeto de esta revista, explicaciones acerca de lo que constituye la esencia de los sistemas de comunicaciones de datos, es decir, interoperabilidad en su sentido técnico¹⁰ -sintáctica¹¹ y semántica¹²- que la legal siempre presupone ya

¹⁰ La interoperabilidad técnica de datos es la “propiedad de un producto o sistema ... de trabajar con otros productos o sistemas, actuales o futuros sin ninguna restricción en su acceso o implementación”. <http://interoperability-definition.info/en>.

¹¹ Si dos o más sistemas se pueden comunicar entre sí, gozan de interoperabilidad sintáctica, lo que se logra con formatos de datos y protocolos y estándares específicos, tales como XML o SQL u otros de menor nivel.

¹² La Interoperabilidad Semántica se define como la capacidad que poseen los sistemas de información para intercambiar datos con significado unívoco y compartido. Es un requisito que permite a dichos sistemas ejecutar mecanismos de inferencia lógica, descubrimiento de nuevo conocimiento y federación de datos, con lo cual puede decirse que pasan a denominarse como “sistemas de información y del conocimiento”. Por lo tanto, la Interoperabilidad Semántica no se refiere tan sólo a la “sintaxis” y simple intercambio de los datos, sino también a la transmisión simultánea junto a éstos de su significado (de ahí lo de “semántica”). Esto se lleva a cabo mediante la inclusión de los así llamados “metadatos” (“datos de los datos”, descriptores de éstos en sí), enlazando cada dato a un determinado “vocabulario compartido” compuesto por dichos metadatos. Así pues, el significado de los datos es transmitido junto los datos en sí conformando un “paquete de información” que pueda ser interpretado de manera independiente por cualquier sistema de información y del conocimiento. Es este “vocabulario compartido”, y sus enlaces asociados a una determinada “Ontología” los que confieren las bases y la capacidad a las máquinas que integran dichos sistemas de información y del conocimiento de poder interpretar e inferir lógicamente a partir del significado de dichos datos. Una ontología es una especificación de una conceptualización, esto es, un marco común o una estructura conceptual sistematizada y de consenso no sólo para almacenar la información, sino también para poder buscarla y recuperarla, de manera que una ontología define los términos y las relaciones básicas para la comprensión de un área del conocimiento, así como las reglas para poder combinar los términos para definir las extensiones de este tipo de vocabulario controlado anteriormente referido. Véase, por todos, Juan Miguel González-Aranda, Rodríguez-Clemente Rafael, Lozano Sebastián (2009). “E-Research In International Cooperation Networks In Science & Technology Research”. Capítulo del libro “E-Research Collaboration: Frameworks, Tools and Techniques”. Berlin/Heidelberg: Springer-Verlag. ISBN 978-3-642-12256-9 e-ISBN 978-3-642-12257-6, pp. 167-199 y que se basa en gran parte también en la tesis doctoral del primero, Juan Miguel González-Aranda, “Creation and nurturing of Science & Technology cooperation tools by means of Virtual Communities of Practice by using Organizational Knowledge Management Systems and E-Research Collaboration Tools” (2012), a su vez resultado de la colaboración entre el CSIC (Prof. Rafael Rodríguez-Clemente) y el Departamento de Organización Industrial y Gestión de Empresas de la Escuela de Ingeniería de la Universidad de Sevilla (Prof. Sebastián Lozano-Segura).

que opera a partir de ella o como un plus a la misma. El concepto de “interoperabilidad legal”, pues, traducción literal del inglés (“*legal interoperability*”) ya que en castellano sería más correcto hablar de “interoperabilidad jurídica”, es entendido hoy universalmente (aunque desde hace poco tiempo¹³), desde una perspectiva amplia, como el lado jurídico o los aspectos jurídicos que necesitan ser tenidos en cuenta y solventados siempre que se pretende lograr esquemas de “interoperabilidad técnica” de datos digitales¹⁴, y es que, a partir del reconocimiento de que “el otro reto que tiene la interoperabilidad (global) de datos consiste en lograr la “interoperabilidad legal” que, de no lograrse, condiciona, si no impide, el procesado automático de datos basado en interacciones “machine-machine” llevados a cabo globalmente y, por tanto, la susceptibilidad de acceso y reutilización de datos entre operadores situados en distintas jurisdicciones nacionales que, como soberanas, regulan de manera diferente el derecho aplicable al carácter público o confidencial de los datos por distintos motivos, de interés privado o público, en cualquier caso ambos de interés general normalmente, la responsabilidad de los errores o el mal uso de los datos u de las bases que los contienen, la protección del software que permite acceder o reutilizar los datos, los controles de su calidad, o la propiedad misma de los datos y a los que controlan su uso a través de los de

¹³ Ni siquiera lo utilizan formalmente los autores del artículo que dio lugar, creo que por primera vez, a un avance inicial de los problemas que luego quedarían abarcados, para buscarles soluciones, por dicho concepto. Vid. J. H. Reichman & P.F. Uhler, A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment, en *Law and Contemporary Problems*, Vol.66, pgs 315-462 (2003).

¹⁴ Como ejemplo de la aplicación de la misma al campo que es objeto en concreto de consideración, el de los datos de biodiversidad, puede verse, González-Aranda, J.M., Sáenz-Albanés A.J., Marco de Lucas J., Sánchez-Gimeno B. "The role of e-Infrastructures: Linking biodiversity and ecosystems through the deployment of new water quality technologies in river basins" (2013). *Proceedings of the 4th SCARCE CONSOLIDER International Conference on "Assessing and predicting effects on water quantity and quality in Iberian rivers caused by global change 2009-2014"*.

A su vez, como ejemplo muy gráfico y claro de qué es interoperabilidad técnica y precisamente en este campo, puede verse el Real Decreto 556/2011, de 20 de abril, para el desarrollo del Inventario Español del Patrimonio Natural y la Biodiversidad, <https://www.boe.es/buscar/doc.php?id=BOE-A-2011-8228>, cuya disposición adicional tercera, por cierto, impone la interoperabilidad con plataformas de e-ciencia. Su valoración desde la perspectiva jurídica y como política pública de manejo de datos puede verse en el dictamen del Consejo de Estado nº 2735/2010, de 3 de marzo de 2011, <https://www.boe.es/buscar/doc.php?id=CE-D-2010-2735> .

propiedad intelectual o industrial que les protegen¹⁵”. Pero también tiene el concepto una noción más restringida limitada a unos aspectos más concretos, a veces más discutible, ya que todavía es un concepto jurídico indeterminado relativamente joven. La más restringida de las nociones la limita a tener derecho a través de información obtenible a través de los datos, de si hay o no límite legal alguno a su uso o reutilización y, si lo hubiera de algún tipo, a conocer cómo y cuáles son, o mejor dicho, cómo puede llegarse, a través de la información que acompaña a los datos o a los servicios web o infraestructuras (de momento dejamos de lado la cuestión de si un acuerdo universal de metadatos al respecto -“*rights statements metadata*”) no tanto como para tener claro las posibilidades de acceso y reutilización sino. Al menos y como mínimo donde puede obtenerse información en formatos al menos “*human readable*” acerca de cuáles son esos límites o como llegar a enterarse de ello.

En palabras de Puneet Kishor *et al* hablando de la interoperabilidad legal de servicios web, “en lo que a los datos se refiere, cuando dos o más bases de datos se combinan (“*are mashed up together*”) el usuario tiene siempre la responsabilidad de intentar averiguar qué licencias o permisos (o derechos) de uso va a heredar la nueva base de datos recién creada por el nuevo servicio. No hay una salida simple a ese dilema en los casos de estas licencias o permisos (o derechos) de uso porque distintas licencias pueden contradecirse entre sí con total facilidad y bloquear la interoperabilidad de datos en que se sustenta el nuevo servicio web pese al que la interoperabilidad técnica y semántica sea perfecta. En realidad la interoperabilidad sólo puede solucionarse al nivel de información casi *ad hoc* para cada dato o series de datos mediante la remisión a documentos que sean “*human readable*” puesto que la mayoría de las combinaciones de permisos y licencias son o bien incompatibles entre sí (por ejemplo, algo no puede simultáneamente “*All Rights Reserved*” y “*By Attribution*”) o son redundantes (por ejemplo, cuando algo es a la vez “*All Rights Reserved*” y “*Public Domain*” *de facto* obviamente es “*All Rights Reserved*”)¹⁶.

¹⁵ Principles on the Legal Interoperability of Research Data. Implementation Guidelines. Draft de 9 febrero de 2016. Definitions and Discussion of Terms. Interoperability of Data.

¹⁶ Puneet Kishor, Oshani Seneviratne y Noah Giansiracusa, Policy Aware Geospatial Data. Por cierto este artículo muestra como tendría que operar un modelo de algoritmo para poder llegar a seleccionar la mejor combinación posible de las licencias, permisos

Y en palabras del mismo experto, todavía más gráficas, casi anecdóticas (pero como la memoria de las pequeñas cosas a veces señala a las grandes”¹⁷, no me resisto a reproducirlas), emitidas mientras tratábamos en grupo la cuestión de si la interoperabilidad legal es un mínimo aplicable a todo manejo de datos digitalizados (incluso sin necesidad de que transiten por la web cuando los datos se llevan en un USB de un ordenador a otro), o si su concepto debe por el contrario ampliarse a otro tipo de problemas que surgen cuando se montan servicios web o infraestructuras basadas en un número mucho mayor de datos y proveedores y usuarios que por tanto hacen necesario en la búsqueda de datos una respuesta automatizada a preguntas (“*response to queries*”) que ya responden (eso sí, si hay ordenadores por medio, a una adecuada interoperabilidad ontológica y semántica¹⁸:

“Cuando me refería a que mediante la interoperabilidad legal debe poderse responder a las preguntas como se formalizan las búsquedas (“response to queries”) no me refería a una búsqueda como la siguiente: “He Enrique, ¿podría tener el artículo sobre la Gran Disconformidad”¹⁹ que publicaste el año pasado? La respuesta a esa

o derechos de acceso y uso para crear una nueva licencia que sería la base técnica y económica del nuevo servicio.

¹⁷ Amelia Valcárcel, Lectio de Doctora Honoris Causa de la Universidad de Valencia, 8 de marzo de 2016, pg.13.

¹⁸ La conversación que reproduzco, con su autorización, y también reproducida ya en abierto en <http://punkish.org/Explaining-Legal-Interoperability>, proviene de un intercambio de chats simultáneo a una de las de las videoconferencias de la *National Academy of Sciences* de Estados Unidos en cuyo seno estamos teniendo gran parte de las discusiones del Grupo RDA-CODATA de Interoperabilidad legal que copresido. He cambiado el nombre del destinatario de su mensaje por el mío porque era la forma de Puneet de intentar explicar a un jurista experto en derecho digital pero escéptico acerca de que debiéramos precisar el concepto de interoperabilidad legal en que estábamos exactamente trabajando desde hacía varios años porque, efectivamente, dicho experto no parecía haberse enterado, lo que da sin embargo idea de la necesidad de extender los conocimientos de derecho digital más allá de lo que las firmas de abogados y cátedras nos tienen acostumbrados. El artículo sobre la Gran Disconformidad a que se refiere es, naturalmente, imaginario.

¹⁹ Una nota para juristas, ya que el Gran Sistema Educativo Español, que nos convierte en diletantes sistemáticos, nos impide enterarnos de qué va el mundo desde los 10 o 12 años de edad si queremos ser buenos juristas (!!): La Gran Disconformidad (*the Great Unconformity*) hace referencia a los enormes huecos o lagunas que quedan en el registro geológico de las capas o discos de roca a lo largo de todo el planeta, donde

búsqueda será, lógicamente, una copia o una clave de acceso a un artículo científico, pero eso nada tiene que ver con la interoperabilidad de datos científicos. Esta debe dar respuesta, más bien, a otra pregunta/búsqueda: “Hey, Enrique, por favor, ¿me puedes mandar o dar acceso al archivo netCDF de mediciones del modelo de temperatura para Groenlandia que utilizaste en Marzo pasado?” Si tengo suerte y obtengo ese netCDF de tu parte, lo que quiero es asegurarme de que puedo utilizar esos datos para mezclarlos con otros datos de isótopos del agua que tengo en formato ERDAS sin incurrir al hacerlo en algún problema jurídico porque los datos que me enviaste tienen restricciones de uso.

No importa cómo obtengo los datos –si llamándote, en un archive FPT, copiándolos usando el programa `scp`, en un floppy 5¼ via FedEx, a por medio de un servicio RESTful, quiera Dios que en un documento adjunto a un email tuyo de respuesta a mi petición – lo que si me gustaría y quiero que venga con los datos es una absoluta claridad acerca de si puedo usarlos o no. Es claro que en general unos resultados de esas preguntas/búsquedas serán susceptibles de copyright y otros no, algunos tendrán ciertas restricciones de acceso o uso y otros no. La clave está en tener absoluta claridad acerca de los derechos a que están sometidos los datos para optimizar la interoperabilidad-

Concluyo con dos afirmaciones:

1.- Mientras algunos datos pueden venir en un número incontable de formatos y modos, lo que no entiendo es que es lo que les hace tan esencialmente diferentes como para que todos ellos tengan interoperabilidad legal, Me gustaría saberlo.

rocas sedimentarias relativamente jóvenes, todavía de unos pocos cientos de millones se asientan justo encima de otros mucho más antiguos y que son un obstáculo (o más bien un reto científico) significativo para la comprensión de la historia de la Tierra, incluido un enigma para Darwin ya que como todo el mundo sabe, fue la geología (más que incluso la paleontología) lo que encendió la cerilla de su genio que llevó a la ciencia de la evolución. Por cierto, dadas las recientes vinculaciones de la explosión de diversidad de vida en el Cámbrico con una de esas Gran Disconformidades, aunque Puneet no lo sepa... ¡¡estaba avanzando datos muy significativos para la discusión de la escala temporal de las EBVs!!

2.-Si después de que llevamos trabajando tres años en este tema²⁰ las recomendaciones que hacemos sólo sirven para “simples bases de datos” o datos observacionales, quedaría decepcionado.”

De hecho, la falta de entendimiento de qué sea exactamente “interoperabilidad legal” ha hecho que ya cerrada la edición de este artículo haya tenido que resumirse la misma la madrugada (hora española) del 16 de marzo de 2016, para el debate internacional que va a tener lugar acerca de este tema en SciDataCon and International Data Week, que se celebrará el 11-16 de septiembre 2016 en Denver, Colorado, EE.UU...

La expresión más resumida, elaborada a partir de la definición del Group on Earth Observations, de 2014 y de un artículo clave, de próxima publicación en septiembre²¹, dice lo siguiente:

“La capacidad de la comunidad científica de investigadores de compartir, acceder, y reutilizar datos, así como de integrar datos de distintas fuentes para la investigación, educación, y otros fines, requiere reglas y prácticas de interoperabilidad efectiva tanto técnica como semántica y “legal”. Tiende a haber malentendidos y falta de conocimiento acerca de las cuestiones jurídicas que afectan en general a los datos científicos y en concreto sobre el concepto de “interoperabilidad legal”.

Existe interoperabilidad legal entre múltiples datos, listados y bases de datos cuando:

- Las condiciones de uso se pueden determinar claramente y con rapidez y facilidad para cada uno de los datos, listados de datos y bases de datos,
- Las condiciones jurídicas para su uso que impone cada dato, listado de datos y base de datos permiten la creación y uso de productos combinados o derivados, y
- Los usuarios pueden acceder legalmente y utilizar cada dato, listado de datos y base de datos sin tener que solicitar la autorización caso por caso de quienes tienen derechos de

²⁰ Se refería al trabajo de RDA-CODATA que copresido y sobre el que luego se habla más extensamente.

²¹ Doldirina, Eisenstadt, Onsrud, and Uhler, publication pending.

propiedad sobre los datos, asumiendo dichos usuarios, por tanto, que se dan las condiciones acumuladas para el uso de todas y cada una de las bases de datos y datos.

La interoperabilidad legal implica que la búsqueda y el rastreo de las condiciones jurídicas aplicables a los datos y bases de datos ocurrirán en línea a través de métodos automatizados. Cuando se combinan los datos de múltiples fuentes originarias, la base o el listado resultante incorporará las restricciones acumuladas impuestas por cada una de las fuentes. Cuanto menos restricciones existan en los datos, sus listados o bases de datos originarios o subsiguientes menos restricciones habrá en los productos derivados que utilizan aquéllos. La manera más simple de buscar y rastrear y de obtención de interoperabilidad legal tiene lugar cuando hay una afirmación clara e identificada de que no existe restricción jurídica alguna bien en los datos originarios bien en los subsiguientes y productos derivados”.

La ventaja relativa con que se parte en el campo de los datos de diversidad biológica por comparación con los de otros campos científicos es que “aunque hay diferentes puntos de vista acerca de qué quiere decir “interoperabilidad legal”, las infraestructuras del proyecto GLOBIS-B tienen ya un acervo común acerca de qué significa “asegurarse que los datos de dos o más bases de datos pueden ser combinados o reutilizados de cualquier otra manera por cualquier usuario sin comprometer los derechos de ninguna de las fuentes origen de dichos datos” (Ref: Legal Interoperability Subgroup of the Group on Earth Observations, Data Sharing Working Group)”²².

En realidad, las dificultades jurídicas a las que se enfrenta la interoperabilidad técnica derivan de susceptibilidad o no de los datos como objeto de propiedad (“*data ownership*”), sea pública o privada, de los derechos de propiedad intelectual o industrial de dichos datos una vez hechos públicos y ofrecido su acceso a cualquier tercero (posibilidad que, a diferencia de otros soportes –típicamente el papel- la digitalización y la web han permitido hasta extremos impensables hace

²² “*Flock together with CReATIVE-B: A roadmap of global research data infrastructures supporting biodiversity and ecosystem science*”, apartado 6.1, fruto también del trabajo del autor en otra investigación colectiva internacional, esta vez en el proyecto *Coordination of Research e-Infrastructures Activities Toward an International Virtual Environment for Biodiversity (CReATIVE-B)*, financiado por el FP7 de la Comisión Europea, Grant agreement n°: 284441.

25 años, lo que a su vez ha llevado a un re-análisis todavía confuso y no claramente articulado en los distintos ordenamientos nacionales, acerca de si se trata sólo de un cambio cuantitativo o, más bien, cualitativo debido a la naturaleza de los objetos digitales por comparación con los físicos)²³.

El problema es si los datos científicos, a diferencia de otros objetos susceptibles de copyright, por ejemplo la música en que el mp3 poco ha cambiado el sistema de atribución de derechos del vinilo, por esencia nunca debieron ser susceptibles de tales derechos porque si son realmente datos científicos, no hay creación alguna sino sólo descubrimiento de la realidad misma. Y a este debate, que subyace a la interoperabilidad legal incluso en su sentido más estricto, y que es un debate muy real y no decidido en el campo jurídico en ninguna jurisdicción nacional cuando se desciende al caso concreto (¿se trata de hechos o ideas no protegibles o de expresiones de esos datos o ideas susceptibles de copyright?), no son ajenos, sino todo lo contrario, los datos de biodiversidad.

Ello se debe a que la “cultura” de las comunidades científicas que se especializan en ella, a diferencia de lo que ocurre en otras comunidades científicas –típicamente la de los astrónomos²⁴ -lo que no impide sino todo lo contrario la innovación empresarial²⁵ - y, paradójicamente, por pasar de lo más grande a algo más pequeño, si bien en tiempo más cercano, la de los científicos de la física de partículas, a instancia precisamente de los científicos españoles del experimento CMS del

²³ Vid. Hoeren, Thomas (2014). “Big Data and the Ownership in Data: Recent Developments in Europe”, en *European Intellectual Property Review*, Bd. 2014, Nr. 12, S. 751-754. <http://www.uni-muenster.de/Jura.itm/hoeren/veroeffentlichungen>

²⁴ Vid, la presentación de Xavier Luri Torra, Gaia, Universidad de Barcelona (UB) en la 1ª reunión de la RDA en España, el 10 de julio de 2015, preparatoria de su IX Plenario mundial que se celebrará en Barcelona del 5 al 7 de abril de 2017. <https://rd-alliance.org/plenary-meetings/rda-ninth-plenary-meeting.html>.

²⁵ Vid., por ejemplo, la spin-off DAPCOM-DATA Services S.L., creada el 22 de febrero de 2013 participada por la UB y la Universidad Politècnica de Catalunya-BarcelonaTech (UPC), asesoradas por la Fundación Bosch i Gimpera (FBG) con sede en el Parc Científic de Barcelona para explotar la tecnología de compresión de grandes volúmenes de datos, útil para misiones espaciales y para otras aplicaciones del sector aeronáutico, de la supercomputación, y de las tecnologías de la información y las comunicaciones.

programa LHC del CERN²⁶, o los del proyecto Genoma Humano-, tienen la tradición de hacer pivotar sus carreras profesionales sobre datos obtenidos una vez, en investigaciones de campo que se llevaron su esfuerzo de juventud, por lo que muchas de ellas (cada vez menos) siguen padeciendo el irónicamente denominado síndrome quasi-psiquiátrico “*it’s my data syndrome*”²⁷.

Pero, al igual que el concepto de interoperabilidad legal debe tener en cuenta estos otros presupuestos jurídicos que le trascienden puesto que son aplicables a los datos como tales y también, por tanto a los datos derivados de artículos en revistas científicas no digitales, y es ineludible, por tanto, que en cuanto que se digitalizan, deben tenerse en cuenta, de la misma manera que otros muchos principios de los derechos de propiedad y de propiedad intelectual e industrial ya que no desaparecen estos sectores del ordenamiento por el mero hecho de hacerse los objetos digitales y publicarse en la web o ser transmitidos a través de ella aunque el destinatario no sea el público sino un concreto individuo por medio del correo electrónico, *drop boxes* u otros mecanismos de la nube (vid, por ejemplo el problema que plantea la digitalización de las colecciones de museos de historia natural; o las digitalizaciones de las bibliotecas, por no mencionar otros fenómenos tales como la Directiva de obras huérfanas), la interoperabilidad legal hace desaparecer parte de los problemas jurídicos de las bases de datos, servicios web o infraestructuras, pero ni mucho menos todos ellos.

²⁶ La investigadora del Instituto de Física de Cantabria (IFCA, CSIC-UC), Teresa Rodrigo quien, como presidenta del consejo de la colaboración CMS, impulsó la elaboración de la política de publicación de datos de CMS en abierto, en cuyo desarrollo e implementación han contribuido Jesús Marco (IFCA) y Alicia Calderón, investigadora Ramón y Cajal de la Universidad de Cantabria la investigadora del IFCA Ana Rodríguez ha colaborado en el desarrollo del portal y el software de análisis de varios ejemplos de física de CMS. El IFCA ha desarrollado también la web <https://cmsopendata.ifca.es> donde se ofrecen estos datos en abierto con fines educativos.

²⁷ El mejor estudio acerca de las actitudes de distintos investigadores y comunidades científicas acerca de los nuevos paradigmas Open Science y Data Sharing es el llevado a cabo en 2010 por DataOne, por encargo de la National Science Foundation de los EE.UU. cuyos resultados fueron presentados por Rebecca Koskela, en el primero de sus plenarios celebrados en EE.UU. (octubre 2012), de la RDA. Vid *Achieving Progress and Breaking Barriers in Data Access and Sharing*, descargable de la web de la RDA. Véase, analizando los resultados de ese estudio, Tenopir, C et. al, (2011) *Data Sharing by Scientists: Practices and Perceptions*, en <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>

Ciertamente ello escapa, por ahora, al menos, a los trabajos del Grupo de Interoperabilidad legal de la Research Data Alliance (RDA) y CODATA²⁸, pero no a los de otras instituciones y desde luego, no puede el proyecto GLOBIS-B de la misma manera que tampoco puede hacerlo ninguna de las bases de datos, servicios web o infraestructuras que manejan datos científicos, al análisis de esos otros problemas, aunque no sean de interoperabilidad legal *strictu sensu*.

Así, por ejemplo, uno de las principales infraestructuras de biodiversidad, participante en el proyecto, GEOBON, es parte del GEOSS y éste, además de principios de Data Sharing articula instrumentos y prácticas legales que deben respetarse, dentro de la debida flexibilidad que cada tipo de datos necesariamente conlleva y con el *caveat* que supone el debate científico sobre la utilidad real (en el fondo calidad de los datos) de la tecnología geoespacial para ofrecer datos sobre parte de la diversidad biológica (típicamente el mundo animal)²⁹. En concreto ha acordado ya *Data Management Principles* (sus *Implementation Guidelines* están también muy avanzadas e incluso han sido ya aceptadas para su incorporación “parcial”, en la medida de lo posible debido al tipo de datos que se manejan, por el U.S. Common Framework for Earth Observation el 22 de diciembre de 2015, para la revisión de su política de datos que comenzó el 15 de enero de 2016).

Además, finalmente, la praxis de los servicios web, basados o no en Big Data, y con o sin *crowdsourcing* y *citizen science*, ha dado lugar a todo un sector del ordenamiento que, no ya por tendencia natural a no apartarse de la misma, sino también porque las consecuencias jurídicas puede ser graves, a que los datos científicos no se aparten en exceso de ella.

A un ciudadano o ciudadana que mediante una app manda a la NOAA desde su barco (o un barco alquilado) una foto de su móvil de un ejemplar de una especie de cetáceo (dato científico) que naturalmente debe estar georeferenciada a través de la IP para poder ser validada en la infraestructura, no tiene que molestarle menos el que se trate de eso que de una simple consulta de la carta náutica que esté manejando (lo que no se recomienda por cierto, salvo que se sea experto en distinguir los

²⁸ <https://rd-alliance.org/group/rdacodata-legal-interoperability-ig.html>

²⁹ Vid. Andrew Skidmore, Natalie Pettorelli *et al*, Agree on Bio diversity Metrics to Track from Space, en *Nature*, Vol. 523, pgs 403-405, 23 de julio de 2015.

riesgos de su manejo en vez de manejar las oficiales) para conocer cuál es el puerto de la costa más cercano donde puede fondear esa noche (servicio comercial web que nada tiene que ver con la ciencia, ni tampoco los datos que se utilizan para montarlos) y que, naturalmente, también queda georeferenciada como información útil al prestador del servicio a efectos de agregar datos de rutas más o menos utilizadas en las que se consulta más o menos su producto y que naturalmente es también suya y puede vender como negocio o mecanismo de financiación adicional de su servicios, incluso en exclusiva, a otro negocio, etc.. Lo que le molesta es que su identificación personal a través de la IP se haga pública debido a qué ... ¿por qué razón tiene que saber nadie... o alguien en particular... que estaba precisamente allí esa noche de luna llena?

Los servicios web se basan en un entramado de negocios jurídicos más o menos regulados según cada jurisdicción nacional, que constituyen una auténtica red (conjunto o sistema de “clusters” de “legal issues”) que los especializados en biodiversidad científica no pueden ignorar porque su irresponsabilidad, de hacerlo, puede acarrear como consecuencia un desastre en un momento dado o el fracaso de toda una parte de la provisión de datos o de su uso (casi todos los proyectos científicos de *citizen science* acaban fracasando porque la ausencia de esas políticas y decisiones con implicaciones jurídicas acaba generando desconfianza en los voluntarios que entusiásticamente comenzaron a participar en ellos suministrando sus datos).

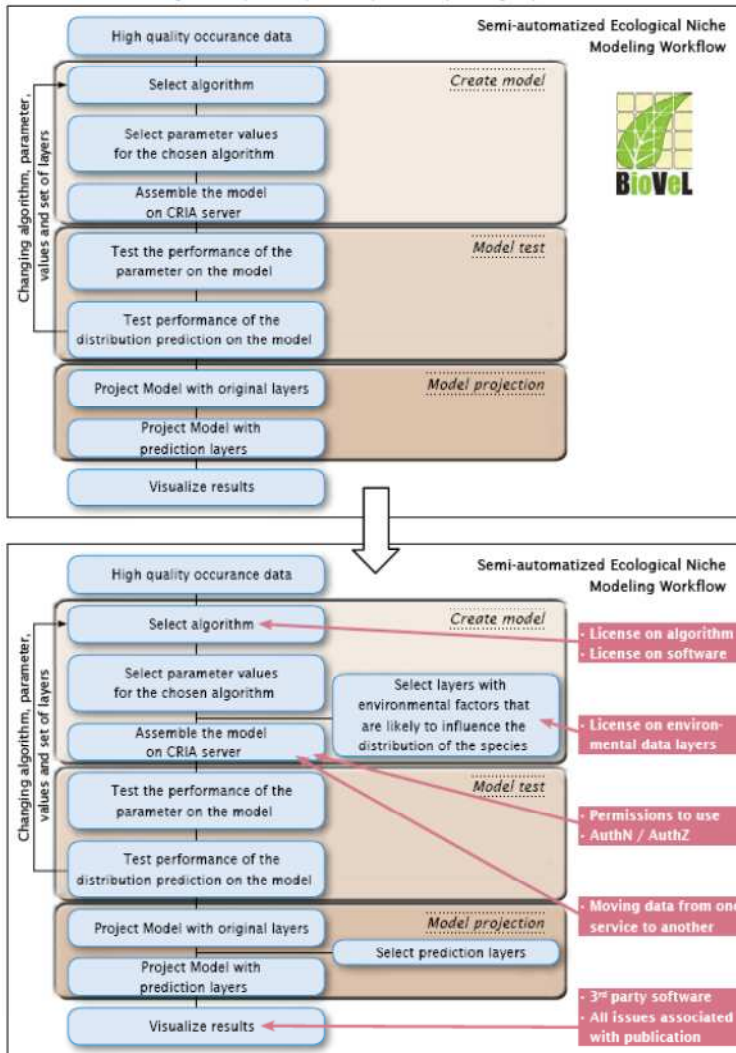
Por tanto, de ese último sistema debe también ocuparse este proyecto.

Tres puntualizaciones todavía necesarias en esta introducción.

1.- Muchos de los obstáculos y cuellos de botella que la ausencia de interoperabilidad legal puede generar en la construcción de servicios web y de infraestructuras, concretamente en el campo científico de la biodiversidad, también han sido abordados recientemente en el marco de otros proyectos de la Unión Europea. En concreto por el proyecto BIOVEL, como se pone de relieve en la hoja de ruta resultado de un proyecto previo al proyecto GLOBIS-B: “*Flock together with CReATIVE-B: A roadmap of global research data infrastructures supporting biodiversity and ecosystem science*”³⁰.

³⁰ Véase el punto 6 de dicha Hoja de Ruta, citada en la nota 16 anterior.

La figura, cortesía del proyecto BIOVEL y que se reproduce, por su importancia como visualización de lo que se quiere decir, tanto en el *Roadmap* como en la página siguiente de este artículo, muestran, con un ejemplo gráfico y derivado de un experimento real en el análisis de un flujo de datos de biodiversidad, la cantidad de problemas de interoperabilidad legal que generan los derechos de propiedad intelectual e industrial y la falta de formulación de licencias, renunciaciones, declaraciones, limitaciones de responsabilidad etc. Aquí también se parte con la ventaja del saber acumulado por comparación con otros campos científicos. Se pasa, pues, a analizar un poco más en profundidad cada una de estas materias, aunque sólo dese la perspectiva jurídica.



Además, puede consultarse la obra colectiva que permitió a la comunidad internacional de expertos algunos años antes identificar los grandes retos de la informática aplicada a la biodiversidad, lo que ha quedado plasmado en el *Global Biodiversity Informatics Outlook* (GBIO) que elaboramos en su día en 2012 un amplio número de expertos una semana de “encierro” para discusión abierta y centrada sólo en ello en

Copenhague y posterior serie de teleconferencias semanales que la siguieron³¹.

2.- La creación de valor económico a partir de largas series de datos (“Big Data”) es uno de los motores de la innovación empresarial producto de la economía del conocimiento³², que opera sobre la base de mejoras en la eficiencia y productividad que de la gestión de datos hacen las empresas (y los consumidores de sus productos). Dicha gestión a través de servicios web a las empresas (y al público en general) ha generado un interés muy importante en un nuevo subsector del derecho mercantil que intenta crear un subsistema coherente de normas aplicables a estos servicios y relaciones jurídicas que se entablan a través de los mismos, desde el simple re-análisis de qué es exactamente un “dato” (o mejor, un “dato digital”) y el derecho a través del cual adquiere valor (propiedad estricta del dato, propiedad intelectual e industrial, contratos de cesión de uso –*licensing*-, términos y condiciones de uso, renunciaciones unilaterales etc.) que tienen consecuencias directas e indirectas sobre todos los sujetos de estas relaciones jurídicas, es decir, tanto para los proveedores y usuarios de los datos como para los prestadores de los servicios web (o gestores de segmentos de su flujo de datos) y usuarios, por tanto de su hardware, middleware y software y apps, lo que se aborda más adelante al final del apartado II y en el apartado III de este artículo.

Pero no se entenderá ese subsistema jurídico, o más bien esa subcultura jurídica (es decir, “*legal environment*”), combinado con el que le sirve de base, que no son otros, para los datos mismos –el software es diferente- que los clásicos “secreto comercial” y “copyright” de las expresiones de otro tipo de datos que ha venido siendo aplicables a medios más tradicionales tales como las publicaciones y ediciones en papel, a las imágenes y fotografías, mapas, etc., incluso cuando la creación o su expresión procede de dinero pagado por los ciudadanos como contribuyentes, obra es pagada con dinero de los contribuyentes, también se ha venido aplicando, ¿por inercia?, a los datos científicos cuyas carreras profesionalizadas han hecho de ese derecho uno de los

³¹ <http://www.gbif.org/resource/80859>

³² Vid., acerca de ésta, Casas Baamonde M.A. & Alonso García E., 7ª edición revisada de Manuel Alonso Olea, Introducción al Derecho del Trabajo, Civitas, 2014, pgs 437 y ss; y Joel Mokyr, Los Dones de Atenea: los orígenes históricos de la economía del conocimiento, Marcial Pons Historia, Madrid 2008.

pilares de su promoción (de hecho los grandes consorcios o compañías editoriales son, muchos de ellos, más poderosos que los equivalentes de los de obras de ficción).

Y ello ha generado, hoy no reconocerlo es banal, la paralización, al menos en muchos sectores, de la ciencia, con supuestos nada anecdóticos de auténtico escándalo, como el del enorme retraso que en la investigación de remedios de enfermedades raras, en concreto la PXE, como consecuencia de práctica, por nadie contestada, de seguir operando con el derecho tradicional de datos hasta que los ciudadanos han decidido organizarse para controlar ellos mismos el manejo de datos frente a los actores tradicionales –grandes y célebres instituciones de investigación médica-³³. Y ese efecto simplemente paralizador de la ciencia se debe a que los excesos de la mano invisible (las aspiraciones del *homo economicus* en el fondo) de la economía de la ciencia, es decir de los científicos como profesionales individuales, ha dado lugar a que entrada la década presente el derecho haya afirmado exactamente los principios contrarios a aquéllos en los que se sustenta ese derecho de los datos tradicional, al menos –aunque no limitado a- el mercado de la investigación científica financiada con fondos públicos directa o indirectamente, al generarse ineficiencias que trascienden a la ciencia misma al convertirse en ineficiencia – a veces muy grave- en la gestión del pequeño pero nada desdeñable porcentaje del PIB dedicado a la misma. De la misma manera que la sociedad no puede permitirse el lujo de tener cinco o seis autopistas de Barcelona a Madrid porque a quienes construyeron las cinco primeras, aunque se le ha pagado todo su esfuerzo incluido el beneficio industrial, se le regala además el poder –cobrando- de decidir dejar pasar por ellas a quien les dé la gana, idea tan absurda que ni siquiera se sugiere por los expertos de la economía de las infraestructuras de transporte³⁴, la sociedad no puede permitirse el lujo que los datos que sirven de base a las nuevas infraestructuras claves en la economía del conocimiento (las *e-science research infrastructures*) tengan que comprar los datos a quienes no los tiene porque desde su origen mismo simplemente no son suyos.

³³ Sharon F. Terry. Life as a numerator: Putting the person in personal genomics, en *Applied and Translational Genomics*, 1 de febrero de 2016, <http://www.sciencedirect.com/science/article/pii/S2212066116300072>; o, más correctamente, Terry, S.F., Life as a numerator: Putting the person in personal genomics, *Appl. Transl. Genomic.* (2016), <http://dx.doi.org/10.1016/j.atg.2016.01.007>

³⁴ Véase, por todos, José Antonio Gómez Ibáñez, *Regulating Infrastructure: Monopoly, Contracts, and Discretion*, Harvard University Press (September 1, 2006)

Este exagerado mal funcionamiento del derecho regulador de los datos científicos es lo que ha generado que principios radicalmente contrarios al mismo en su versión tradicional se apliquen al derecho de los datos científicos cuando se han obtenido con dinero público. Y no se trata de una “moda” (no es meramente “trendy” como pretenden quienes intentan abortar el fenómeno en una vana esperanza de que se olviden esos nuevos principios en unos años), sino que se han convertido en uno de los pilares esenciales de la economía del conocimiento. Los principios de *Open Science* (o en otras formulaciones, de *Open Access*, o en EE.UU. de *Public Access*, *Data Sharing*, *Open Source* –para software, aunque aquí son más limitados en su alcance) que están siendo sólidamente establecidos entrada ya la década de 2010 por todo el mundo³⁵. Los datos producidos con recursos públicos deben ser públicos y de libre acceso y reutilización para todo el mundo, ese es el nuevo paradigma, y punto.

La *Open science*, que ya tiene incluso su mártir oficioso³⁶, y que no se limita a) a los “datos científicos” (*research data*³⁷) ni b) a la divulgación en abierto de los mismos (*openness of data*), no sólo responde a la imparable necesidad de asegurar el progreso científico y sacarle del

³⁵ Véanse entre otro muchos las políticas de Australia, EE.UU. y la Unión Europea. Para Australia esta política fue establecida por el Australian Research Council en enero de 2103, http://www.arc.gov.au/sites/default/files/filedepot/Public/Media%20&%20Resources%20Centre/Factsheets/ARC_OpenAccess_factsheet.pdf; en los EE.UU. la Casa Blanca lo hizo oficialmente el 22 de febrero de 2013, “Open Science and Innovation: Of the People, By the People, For the People”, <https://www.whitehouse.gov/live/open-science-and-innovation-people-people-people>; y en la UE el paquete de Open Access policies se estableció por la Comisión Europea en julio de 2012, http://ec.europa.eu/research/science-society/document_library/pdf_06/background-paper-open-access-october-2012_en.pdf

³⁶ Aaron Hillel Swartz (November 8, 1986 – January 11, 2013). Véase el espléndido documental *The Internet's Own Boy: The Story of Aaron Swartz*, estrenado en enero de 2014 en el Sundance Film Festival. Véase también, *Killswitch: The Battle to Control the Internet*, la película que recibió en octubre de 2014 su World Premiere en el Woodstock Film Festival, donde obtuvo el premio de Best Editing y ha continuado recibiendo múltiples más desde entonces.

³⁷ Podría mas bien hablarse de “datos de los procesos de investigación” por traducir más literalmente la reinante expresión de la lengua inglesa. En cualquier caso, debe quedar claro que abarca los datos tanto de la ciencias duras (y menos duras), las ciencias sociales y las humanidades, así como –lo más difíciles de hacer interoperables, lo de las ciencias inter y multidisciplinares.

fracaso de la aplicación sin sentido al mundo digital de principios tradicionales de los derechos de propiedad en sentido amplio, sino que trasciende a la economía de la ciencia y ha pasado a ser un principio general de la economía –en la medida en que por economía se entienda la del conocimiento y no la del ladrillo con la que ciertamente hay que convivir siempre que sea, esta última, un pequeño subsector y no pretenda convertirse en la reina de la economía viable.

Efectivamente, toda la economía sufre estagnación si los datos no se ofrecen en abierto ello se debe a que la competencia, esencial para el funcionamiento y la expansión de los mercados, o es tan buena, si se exagera, en la moderna economía del conocimiento donde “competitividad” debe convivir, y en muchos casos, desde luego en este, ceder terreno a la “generatividad”: a capacidad de producir cambios de manera no anticipada por medio de las contribuciones no filtradas de una audiencia amplia y variada (...) dando lugar al surgimiento de nuevos servicios o productos, impensables para quien hace la contribución [originaria o subsiguientes, añadido yo], construidos sobre la información (datos) que está en abierto y puede compartirse a través de web, de manera que pueda tener lugar una explosión de creatividad” (“*capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences (...) so that new services or products unthought of can be built upon information (data) which is open and sharable in the web, and so that the explosion of creativity may occur*”), en el sentido dado al concepto por Jonathan Zittrain y tal y como ha popularizado John Wilbanks³⁸. El conocimiento crece no por competitividad ni por casualidad, sino porque el que aportó el dato original o subsiguiente no puede pretender tener capacidad alguna para seguir eternamente añadiendo valor a esos datos. Otros con mejores conocimientos o competencias lo harán y nunca por casualidad. Añadirán así valor inimaginado por el contribuidor primero y subsiguientes simplemente porque el conocimiento a nivel individual es siempre limitado. Es obvio que si “cuatro ojos ven más que dos”, millones de ojos ven más que dos. O por utilizar un ejemplo muy fácil de visualizar que suelo usar en mis webinars de *fingerstyle guitar blues*, si a W.C.Handy se le hubiera ocurrido registrar como propiedad

³⁸<http://www.worldchanging.com/archives/011019.html>;
<http://blogs.law.harvard.edu/mediaberkman/2010/03/09/john-wilbanks-on-overcoming-systemic-resistance-to-generativity-in-science/> or
<http://www.slideshare.net/wilbanks/making-energy-data-generative>

intelectual el nombre de “blues” o, sobre todo, la estructura del blues de doce compases (*twelve-bar blues*), simplemente no sólo no habría ni blues, ni jazz –como todos los buenos músicos de jazz confiesan, aunque se pretendan que no es así-, ni folk and rock-and-roll... y el mundo probablemente sería un lugar mucho más triste³⁹ y... también más pobre económicamente hablando. En cualquier caso, volviendo a la economía, remito a las recientes conclusiones de un estudio de la OCDE⁴⁰.

A su vez, de la dos afirmaciones que se acaban de hacer, la primera implica que no debe haber, de hecho no hay, un claro consenso acerca de que es exactamente “dato científico” (*research data*) y con ello hay que convivir. Ciertamente el núcleo duro del concepto es claro: dato obtenido de la realidad física a partir del esfuerzo para descubrirlo llevado a cabo con metodología científica (y en su caso inversión tecnológica de apoyo). Pero todo dato o serie de datos creada, cuidada (*curated*) y transmisible electrónicamente por la web (o incluso en soporte no digital) puede convertirse y es rutinariamente también objeto de investigación científica, desde los datos del censo de población y el resto de la estadística pública, hasta los puros datos sobre los que se ha basado la investigación si protegible tales como los que fundamentan la solidez de estudios o publicaciones, informes etc encargados por los distintos poderes públicos o por ellos financiados (salvo cláusula en contrario que no debe ser de adopción discrecional sino estar basada en que su no adquisición abarata la inversión pública siendo urgente la obtención de la prestación y hacerlo a coste razonable). Este, la divulgación en abierto de estos datos, aunque no sean calificables de “datos científicos” (pero sin que nadie pueda impedir que se conviertan

³⁹ Difícilmente habrían compuesto en 1931 Harry Barris, Ted Koehler y Billy Moll el tema “*Wrap Your Troubles in Dreams (and Dream Your Troubles Away)*” que hizo salir de la depresión psicológica a todos los afectados por la Gran Depresión económica y cuya simple audición, sea quien quiera que sea el o la intérprete, sigue sirviendo a muchos para superar otras muchas crisis puntuales o crónicas.

⁴⁰ Data-Driven Innovation: Big Data for Growth and Well-Being, Published on October 06, 2015. <http://www.oecd.org/sti/data-driven-innovation-9789264229358-en.htm>

en tales⁴¹) es el presupuesto en el que se basa la denominada Directiva PSI⁴².

Dado que se pretende centrar este artículo en los datos de biodiversidad simplemente se deja todo ello apuntado. Es indudable que en este campo donde la inmensa mayoría de los datos han sido obtenidos con financiación pública, son estos nuevos principios los que deben matizar la aplicación del llamado “data law” o “Big Data law”.

3.- Por último, debe tener en cuenta el lector que las diferenciaciones entre datos, bases de datos, servicios web y, sobre todo, las e-infraestructuras de ciencia, la pieza esencial sobre la que pivota la economía del conocimiento de la misma manera que las infraestructuras “grises” primero y “verdes” después, han sido esenciales para la primera, segunda y tercera revolución industrial, también son relativamente claras.

Los científicos normalmente comparten entre sí datos que observan (y el software a través de los cuales se hace posible o más útil su acceso o reutilización). Pero, cada vez más, utilizan los servicios web, programas interactivos utilizados mediante una conexión web. Estos a veces son tan relevantes en la era digital que algunos gobiernos u organizaciones o regímenes internacionales los han institucionalizado debido al valor que añaden a la funcionalidad de sus propias instituciones, bien por prestar servicios “aislables” o incluso “externalizables” bien como puros brokers en el proceso de tratamiento de los datos como ocurre con los denominados “Data Discovery and Access Brokers“ (DAB)⁴³,

⁴¹ Vid. Erika Check Hayden, *The Big Peek: The data contained in tax returns, health and welfare records could be a gold mine for scientists – but only if they can protect people’s privacy*, in *Nature* Vol. 525, 24 September 2015 pgs. 440-442.

⁴² Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público, que entró en vigor el 31 de diciembre de 2003, y fue revisada por la Directiva 2013/37, de 26 de junio que entró en vigor el 17 de julio de 2013. Transpuesta en España en parte por la Ley 37/2007, de 16 de noviembre, modificada por la Ley 18/2015, de 9 de julio, y en parte por el Real Decreto 417/2006, de 7 de abril, luego el Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, todavía no modificado.

⁴³ Vid., por ejemplo, la figura de los Discovery and Access Brokers (DAB) del GEO, en Barbara J. Ryan, *Unleashing the Power of Earth Observations –Together*, RDA 6th Plenary, Paris, France 23 September 2015.

servicios, pues que aseguran que los datos llegan a sus potenciales usuarios de una manera técnicamente “generativa⁴⁴”, e incluso, lo más importante, han creado un nuevo tipo de infraestructura, la denominada “infraestructura del conocimiento” (“*knowledge infrastructure*” o “*e-science infrastructure*”), porque en términos de macroeconomía juegan el mismo papel que las clásicas infraestructuras “grises” jugaron –y siguen jugando- durante los siglos XIX y XX (autopistas, electricidad, aeropuertos...) o que juegan las más recientes infraestructuras verdes” (identificación de los elementos del territorio que aseguran los servicios ecosistémicos, tales como el carácter crítico de determinados hábitats o los paisajes ecológicos y corredores⁴⁵, cuyo reconocimiento empezó en los años 1970).

Pese a que podría dedicarse más tiempo para examinar estas nuevas infraestructuras, que tienen su propio ordenamiento jurídico, ello alejaría, todavía más de lo que lo ha hecho, del objeto esencial de este artículo. Me remito, pues, institucionalmente, a los tratados que las

⁴⁴ No es el mismo concepto de “generatividad” (*generativity*) antes explicado, sino el técnico (indudablemente vinculado con aquél) de “*generativeness*” que se usa para describir “el problema relacionado con la generación entendida como la habilidad de producir fácilmente una licencia nueva (“*the problem pertaining to generation, the ability to produce a new license easily*”), en el sentido que le ha atribuido Puneet Kishor, Oshani Seneviratne and Noah Giansiracusa (2013), Policy Aware Geospatial Data, <http://arxiv.org/ftp/arxiv/papers/1304/1304.5755.pdf>.

⁴⁵ Vease Paul N. Edwards, Steven J. Jackson, Geoffrey C. Bowker & Cory P. Knobel, *Understanding Infrastructures: Dynamics, Tensions and Design: Report of a Workshop on “History & theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures”* (National Science Foundation, January 2007), http://cohesion.rice.edu/Conferences/Hewlett/emplibary/UI_Final_Report.pdf; Brett Frischman, “Infrastructure: The Social Value of Shared Resources” (OUP 2012) and “Governing Knowledge Commons” (OUP 2014). On the functioning of “green” (or “blue” when the ecosystems are marine) see “Green Infrastructure (GI) — Enhancing Europe’s Natural Capital”, Communication of the European Commission COM(2013) 249 final, Brussels, 6.5.2013; acerca de las políticas de paisaje como infraestructura, véase Ana Recarte Vicente-Arche & Enrique Alonso. July 2009 Case study.- “Landscape Protection Policies: The Case of Vermont”, descargable en <http://www.institutofranklin.net/en/investigation/groups/Friends%20of%20Thoreau/projects/case%20studies>; acerca del geodiseño como servicio ideal de la tecnología IT de SIG para la construcción y manejo de la infraestructuras verdes, véase Carl Steinitz, GEOdiseño: Métodos de Planificación Integral del Territorio (ESRI-UAH 2014). O acerca de como diseñar la infraestructura verde a escala simultáneamente continental, estatal y local, puesto que está diseñando la estadounidense, el trabajo de Arancha Muñoz Criado y otros, bajo la esponsorización de ESRI (California) la National Geographic Society, recientemente presentado el 25 de febrero de 2016. Vd en Youtube :<https://www.youtube.com/watch?v=YD5HCtt10ps&feature=youtu.be>

crean internacionales o entre miembros de la Unión Europea (en los países de la Unión sólo la propia Unión tienen potencial para montar las realmente significativas) aunque para facilitar y agilizar su creación y gestión hay un marco en el que se mueven muchas de ellas (las que no están sujetas a cartas fundacionales aprobadas por tratados internacionales): el Reglamento que creó un nuevo tipo de persona jurídica reguladas por el derecho internacional y de la Unión (y el nacional pero sólo en la medida en que se remite al mismo): los ERIC (European Research Infrastructure Consortium)⁴⁶.

II.- DE LOS PROBLEMAS MÁS TÍPICOS DE INTEROPERABILIDAD LEGAL A LOS MÁS COMPLEJOS DE FUNCIONAMIENTO DE LOS VIRTUAL RESEARCH ENVIRONMENTS (VRES) DE INFRAESTRUCTURAS DE E-CIENCIA, PASANDO POR LOS SERVICIOS WEB.

Los principales problemas de interoperabilidad legal a los que un proyecto como el GLOBIS-B se enfrenta son los siguientes:

1.- ¿Cuáles son los mejores enfoques (*top-down and/ o bottom-up approaches*) en los que se basa la necesaria armonización de las políticas o *mandates* de Open Science y Open Access de los datos de biodiversidad dado que la casi totalidad de las infraestructuras participantes son de países que las han adoptado ya y son aplicables, al menos a priori, al tipo de datos necesarios para calcular las EBVs (incluidas las series de datos históricos y las de los distintos lugares geográficos), en especial las de abundancia u ocurrencia de poblaciones de las distintas especies?

¿Cómo lo hacen el GBIF con los datos que maneja en la actualidad y otras infraestructuras internacionales y nacionales? ¿están armonizadas las exigencias de interoperabilidad legal a día de hoy?

Por cierto, para comprender mejor los distintos tipos de datos que este área científica necesita manejar, se reproducen, por un lado, cortesía Donald Hobern, Secretario Ejecutivo del GBIF, una *slide* por él elaborada, porque una imagen vale más que mil palabras: se trata de

⁴⁶ Reglamento (CE) n° 723/2009 relativo al marco jurídico comunitario aplicable a los Consorcios de Infraestructuras de Investigación Europeas (ERIC), modificado por el Reglamento (UE) n° 1261/2013 del Consejo, de 2 de diciembre de 2013.

datos provenientes de fuentes tan variadas como la genómica (donde los desarrollos tecnológicos denominados eDNA están realmente desplazando a otras tecnologías a escalas temporales incluso de años o meses⁴⁷) o de *citizen science*, además de las clásicas de colecciones de museos de historia natural (cuyo proceso de digitalización está siendo tremendamente lento) y de publicaciones de todo tipo⁴⁸.

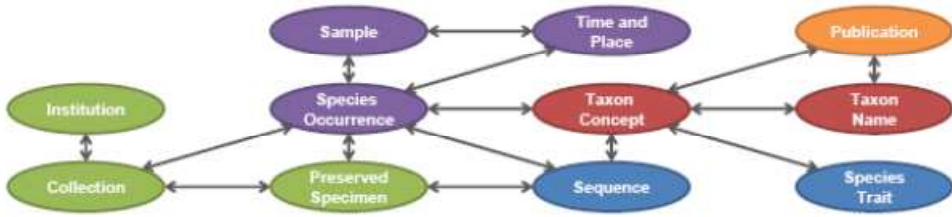
Por otro, otra *slide*, cortesía del centro de GEOBON de Leipzig⁴⁹, que representa el cubo en el cual deben recogerse la multiplicidad de datos para que las dimensiones temporales y geográficas puedan calcularse de manera efectiva.

⁴⁷ Ryan P. Kelly *et al.*, *Harnessing DNA to improve environmental management*, en *Science* Vol. 344 27 de junio de 2014, pgs 1455-56.

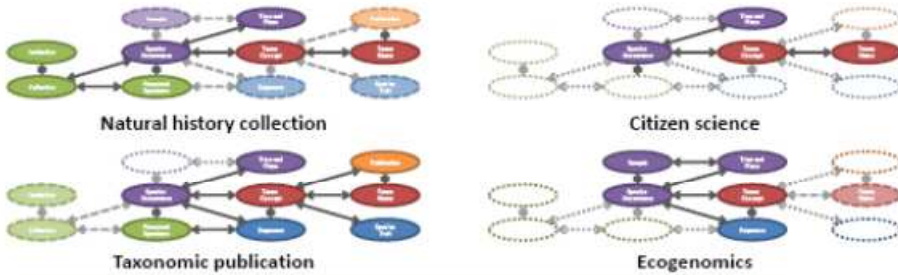
⁴⁸ Acerca del control de la calidad y de la reutilización de datos primarios indexados en el GBIF, desde la perspectiva técnica, aunque no sólo tal, véase la tesis doctoral de Javier Otegui Tellechea, *Quality and fitness-for-use assessments on the primary data indexed at the Global Biodiversity Information Facility (GBIF) / Evaluación de la calidad y usabilidad de los datos primarios indexados en la Infraestructura Global de Información sobre Biodiversidad (GBIF)*, dirigida por Arturo H. Ariño Plana y leída en la Universidad de Navarra en junio de 2012.

⁴⁹ Se trata de un gráfico que fue preparado como resultado de un taller de GEO BON en Leipzig y se incluirá en artículo de próxima publicación: Jetz W., Ferrier S., Pereira H., Fernández M., et al. (IN PREP.) *Species Distribution EBV Workshop*. GEO BON, 2015.

Complex Fragmented Data



Generalised biodiversity data



Essential Biodiversity Variables for species populations

- Defined by the GEO BON / EU BON workshop in 2014 as *"the relative abundance of a taxon (species) in a place and a time, measured repeatedly with a consistent methodology"*.
- **Big data** forms its own methodology.
- Uses n-dimensional hypercube, so called OLAP cube (on-line analytical processing).

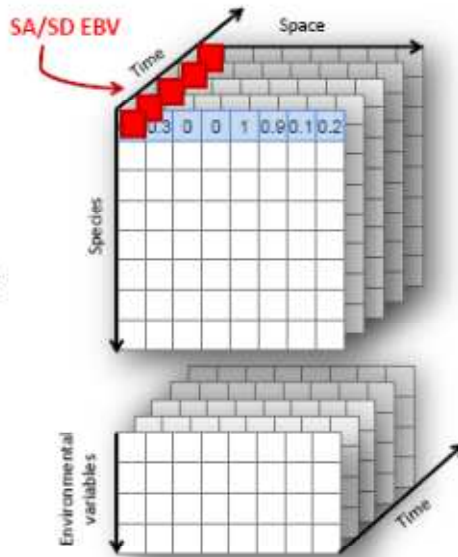


Illustration courtesy of Miguel Fernández, GEO BON

2.- Determinación de estándares y protocolos que vinculen a los proveedores de datos con los usuarios, o a las propias infraestructuras, que sean comunes y si hay o no algunos de obligatoria utilización por imponerlo alguna norma, bien totalmente vinculante (internacionales – mediante tratado- supranacionales o nacionales), por imponerse indirectamente (como condición de los entes financiadores del servicio o la infraestructura, o derivados de costumbre o practica basada en convenios o contratos (o prácticas) de las comunidades científicas o servicios web.

3.- Determinación si todas las infraestructuras y sus proveedores y usuarios de datos entienden lo mismo por “dedicación de los datos al dominio público”.

4.- Determinación si los límites legítimos a la publicidad y compartición de datos obedece en todas las infraestructuras a los mismos intereses generales (por ejemplo, recursos genéticos sometidos a reparto equitativo de los beneficios derivados de su utilización según el artículo 15 de la Convención de Diversidad Biológica y el Protocolo de Nagoya; localización exacta de especímenes de especies en peligro de extinción; confidencialidad de datos de pesquerías etc.).

5.- En particular, si existen límites a acceso o reutilización de datos derivados de intereses vinculados a las políticas científicas nacionales o supranacionales tales como a) obligaciones de confidencialidad (& secretos comerciales) y “periodos de embargo” (el tiempo en que el investigador puede reservarse los datos bien para poder publicar en revistas bien para solicitar patentes; o incluso cesión de datos públicos a investigadores individuales para que puedan publicar en revistas de prestigio su estas, opuestas al *Open Access* y a la *Open Science*, les piden traspaso de la propiedad de los datos⁵⁰, b) prácticas en comunidades concretas del “*it’s my data syndrome*” o tradiciones de carreras profesionales en el sector de la investigación basadas por entero en el mantenimiento en secreto de los datos...; o filosofías distintas de las diferentes comunidades científicas cuando manejan datos sobre la

⁵⁰ Vid, para España, por ejemplo, de esta más que discutible práctica, la primera parte del apartado 6 del artículo 37 de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación: “ Lo anterior [difusión en acceso abierto] se entiende sin perjuicio de los acuerdos en virtud de los cuales se hayan podido atribuir o transferir a terceros los derechos sobre las publicaciones,...”.

misma realidad y ambos pueden ser importantes para la determinación de algún indicador de la EBVs...

6.- Uso por investigadores de la biodiversidad de licencias CC en sus distintas seis variantes (y en particular de las licencias CC0 para los datos) [Por ejemplo, para investigadores españoles, a través de la web https://creativecommons.org/choose/?lang=es_ES].

7.- Técnicas de gestión de los derechos de propiedad intelectual e industrial de las expresiones de datos y de los datos mismos por las bases de datos de las que principalmente se nutren las infraestructuras o los servicios web que proporcionan y ofrecen.

8.- Gestión de la “*attribution*” (reconocimiento de autoría o descubrimiento originario de los datos, cada vez de más difícil solución cuando el resultado combina miles de datos), que ni es un derecho de propiedad sino, allí donde existe (no en los Estados Unidos ni en Gran Bretaña) sino algo muy distinto y típicamente salvaguardado por esquemas “normativos” no jurídicos, salvo en los casos de plagio⁵¹, y de las agregaciones a datos ofrecidos en abierto⁵².

9.- Análisis de si las infraestructuras existentes usan o no y en ambos casos si estiman que es útil y efectiva la utilización de alguna de las variantes de expresión de derechos sobre los datos en los metadatos (*rights statements metadata*) que existen ya⁵³.

⁵¹ Vid. las Implementation Guidelines de la RDA-CODATA al respecto. Para el concepto de normativo *versus* jurídicamente vinculante, véase el punto 11 del texto, inmediatamente posterior.

⁵² Véase, en particular, Lee Belbin, Joanne Daly, Tim Hirsch, Donald Hobern & John La Salle, “A specialist’s audit of aggregated occurrence records: An ‘aggregator’s’ perspective”, *ZooKeys* 305: 67–76. 2013

⁵³ Vid., por ejemplo, el “*Green Paper on Recommendations for the Technical Infrastructure for Standardized International Rights Statements of the International Rights Statements Working Group –Digital Public Library of America (DPLA) & European*” (<http://www.infodocket.com/2015/05/11/a-dpla-europeana-and-creative-commons-working-group-release-white-papers-on-establishing-international-interoperable-rights-statement/>) y la “*Recommended Practice of the National Information Standards Organization on Access License and Indicators*” NISO RP-22-2015 (http://www.niso.org/apps/group_public/download.php/14226/rp-22-2015_ALI.pdf). Pueden verse otros ejemplos en áreas más especializadas en la Implementation Guidelines de la RDA-CODATA.

10.- Si el software de necesaria utilización⁵⁴ para recibir sus servicios web (o las aplicaciones que necesitan los captadores de datos en el caso de la *citizen science*) limitan la capacidad de utilización de los datos que ofrecen; en particular si la reutilización de datos visualizados en mapas sólo puede hacerse utilizando licencias de sistemas de información geográfica no libres y gratuitos (y lo mismo para el funcionamiento de las plataformas de *virtual research environments* –VREs- que pongan en la nube).

11.- Dado que los datos de biodiversidad son de múltiples orígenes y necesitan compilarse tanto bióticos como abióticos, a todos los niveles (del genético a los paisajes, pasando por la diversidad en microbiomas, poblaciones, especies y ecosistemas), y además abarcan no sólo las ciencias de la vida sino también las sociales y las humanidades⁵⁵ siendo interdisciplinarios por definición, lo cual es garantía de innovación⁵⁶... ¿no debería centrarse la gestión de la atribución de autoría y de los derechos sobre los datos desde una perspectiva “normativa” –en el sentido anglosajón del término⁵⁷ - en vez de jurídica estricta para

⁵⁴ Acerca de las opciones de software libre que hoy en día tienen los investigadores para ofrecer en abierto sus datos de manera operativa a otros colegas, véase, por todos, Teresa Gómez-Díaz. CNRS. Laboratoire d'informatique Gaspard-Monge. “Free software, Open source software, licenses: A short presentation including a procedure for research software and data dissemination”, September 14th 2014

⁵⁵ Como ejemplo muy claro puede citarse el de la fijación del nivel de degradación y el de los parámetros a lograr en la restauración de la biodiversidad en la huerta de Valencia por Arancha Muñoz Criado, a partir de los poemas de la literatura árabe clásica de Ibn Al Zayqqaq o Ibn Khafaja (siglo XII) dedicados a la ciudad cuando casi era la capital del comercio mundial, rivalizando con Damasco y de la literatura de viajes de Jerónimo Münzer (1494-95), textos de Gauna de los viajes de Felipe III (1599), el Barón de Burgoing (1777-95), Cavanilles (1795), Teófilo Gautier (1840) además de la Barraca de Blasco Ibáñez (1898). Vid. Arancha Muñoz Criado, Paisaje, humanidades y restauración de ecosistemas: el caso de la huerta de Valencia, Ponencia recogida en los Proceedings of the Third EASLCE (European Association for the Study of Literature, Culture and Environment) Conference *Cultural Landscapes: Heritage and Conservation*, más completo en cuanto a estos datos en Paln de la Huerta de Valencia: un paisaje cultural milenario, Vol. 1, Estrategias de Preservación y Gestión, 2009, pgs. 25-30.

⁵⁶ Vid. Phillip A. Sharp, Meeting Global Challenges: Discovery and innovation through convergence, en Science Vol. 346, 1 de diciembre de 2014, pgs. 1468-1671.

⁵⁷ Aunque sea difícil de entender para juristas continentales europeos, el concepto de lo “normativo” (“*normative*”) se utiliza en teoría general del derecho y técnicas de gestión de políticas públicas en Norteamérica y Gran Bretaña como opuesto a “normativo legal” o como equivalente a normas de grupos sociales cuyo incumplimiento no es sancionable aplicando el derecho sino con otras presiones

facilitar la aplicación de las mismas reglas a comunidades científicas de muy distintas tradiciones? ¿no tiene más sentido realmente alejar el derecho de estas formas de compartir datos en aras de un bien público superior, la ciencia en abierto, aunque reforzando la ética del cumplimiento de acuerdos formales o informales “normativos” y que en el fondo son más eficaces?

12.- En la era de la economía del conocimiento basada en la tecnología digital los datos se han convertido en un patrimonio empresarial muy importante que normalmente es explotado a través de licencias o venta a terceros. Por ello, la protección de los datos se ha exacerbado en los últimas décadas –y todavía más en los últimos años- como un objeto de transacciones per se o como un negocio accesorio de otros servicios tecnológicos a través de múltiples fórmulas contractuales. Ello como ya se vió en los párrafos finales de la introducción ha generado una reacción contraria a su patrimonialización en el caso de los datos científicos por el bien no sólo de la ciencia sino también de la propia economía ya que en este caso los “*property rights*” ahogan mercados, no los producen, al limitar la “generatividad”. En cualquier caso sea como patrimonio apropiable, sea como bien (“*asset*”) difundido en abierto, lo que sí ha generado es toda una cultura de gestión y jurídica de gestión de datos o si se prefiere, ya que en realidad de ello se trata, de planificación de la gestión de datos y de su gestión posterior conforme a dicho plan (“*data management plans*”) hasta el punto de que puede esta gestión ser objeto en sí mismo al ser externalizado a terceros como servicio propio a prestar a otros servicios o procesos industriales de producción. Y se trata de una gestión que tiene sus propias normas jurídicas y prácticas.

sociales (por ejemplo, el ostracismo, el desprestigio, obligación de retractarse, etc. de quien incumple la norma pero sin cuestionar jurídicamente la validez de o que se ha hecho. Por ejemplo, el acuerdo de Genoma Humano de que el descubridor, aunque tiene obligación de revelar inmediatamente sus datos a toda la comunidad, tiene automáticamente concedido por el resto de sus colegas (a nivel global) un plazo de embargo para ser el primero que publique el ,o acerca del, descubrimiento (solo roto una vez por un indeseable que con ello ha logrado ser el más desprestigiado de su profesión, lo que le ha hecho más desafortunadamente famoso que publicar el primero sobre la base del descubrimiento de otro, alcanzando un grado de desprestigio casi universal, por la vía rápida. Acerca de ejemplos concretos de instancias normativas que son más eficaces que acuerdos supuestamente vinculantes de *data sharing*, vid. John Wilbanks “Public domain, copyright licenses and the freedom to integrate science”, *Journal of Science Communication - JCOM* 7 (2), June 2008

El hecho de que el acceso en abierto (*Open Access*), como principio general del ordenamiento aplicable al sector de los datos científicos, se haya afirmado e implantado como principio casi de orden público exactamente contrario al de protección de los derechos de propiedad de los datos con un crecimiento exponencial en los últimos años al menos para el campo de la investigación financiada por fondos públicos, no puede sin embargo entenderse que llegue hasta el extremo de hacerla desaparecer toda la cultura jurídica que se ha ido generando por la *praxis* y la jurisprudencia acerca de cómo deben planificarse la gestión y gestionarse los datos, cultura que puede ser, muchas veces es, aplicable independientemente de que se trate de un servicio público un negocio privado mercantil o del tercer sector (*non-profit*). Lo que en términos generales es conocido como “*data licensing agreements*” es un subsector del derecho digital que opera con toda naturalidad bajo el paraguas de normas legales (pocas pero muy importantes), reglamentos, jurisprudencia y costumbre creada por la *praxis* contractual y de la responsabilidad contractual (a modo de moderna *lex mercatoris*) y extracontractual.

Así que no sólo coexisten toda una serie de normas nuevas regulando la propiedad en sentido general de los datos incluyendo las propiedades especiales intelectual e industrial- (“*data ownership*”)- y el mundo de las obligaciones contractuales “*data licensing*”. La gran diferencia obviamente radica en que, en el primer caso, “*data ownership*”, las normas pueden imponerse a terceros, mientras que en el segundo, “*data licensing*”, sólo vinculan a las partes objeto de la relación contractual (o a) quasi-contractual dado el *caveat* que supone el cómo deben calificarse la interacción directa con los clientes web mediante miles o millones de “contratos” que se celebran mediante un simple un simple click o incluso simplemente accediendo, sin estar obligado a clicar nada, a una página web a través de internet, como ocurre con los denominados “*clickwrap*” or “*browsewrap agreements*”; o b) incluso a veces incluso extracontractual –de ahí lo ambiguo de las diferenciaciones entre el acto de *licensing* y el de aprobar y publicar *terms of use, waivers, disclaimers* etc, muchos de los cuales son unilaterales y no exigen expresión de voluntad del tercero-).

III.- LA GESTIÓN DE DATOS POR SERVICIOS WEB COMPLEJOS E INFRAESTRUCTURAS DEL CONOCIMIENTO.

El último punto del apartado anterior ha puesto de relieve cómo estos servicios, en el fondo y más allá de las cuestiones de interoperabilidad legal en sentido estricto y de gestión de los datos con calidad suficiente (principios de *data management*) dan lugar a un entramado de relaciones formalizadas que se apoyan en la previa experiencia de los servicios web de Big Data como servicio que constituye una actividad empresarial y que ha generado un derecho relativamente complejo mezcla de normas de todo tipo (vid. la consideración final de la introducción del apartado I).

No es cuestión ahora de exponer ni siquiera mínimamente ese derecho, ni siquiera sus principios elementales ya que incluso ha dado lugar al fenómeno sociológico de formas de abogados especializadas por entero o parcialmente en lo que podría denominarse “Data Law” o “Big Data Law”. Se remite para ello al magnífico resumen que ha elaborado la Practical Law Company⁵⁸.

1.- Los datos pueden estar protegidos (“owned”) por fórmulas del ordenamiento clásico de los derechos de propiedad que “dan la vuelta” (*circumvent*) al principio elemental de que los hechos no son susceptibles de apropiación. Estas fórmulas van desde considerar los datos propiedad intangible de la que su propietario o poseedor no puede ser desposeído o privado sin su consentimiento; sería una invasión ilegítima –*trespass* del *common law*- un hurto o robo o un supuesto de quasi-delito de abuso sancionado en las leyes que regulan los ordenadores y sitios web protegiendo del acceso no consentido la propiedad de su contenido en forma de archivos digitales almacenado o una violación de la confidencialidad protegida por el secreto comercial, y los otros derechos de propiedad intelectual o industrial –típicamente el copyright de los datos, las patentes o copyright del software y apps, o los derechos *sui generis* sobre las bases de datos en las jurisdicciones

⁵⁸ La lista que sigue es una racionalización de lo que considero la mejor descripción, o al menos una de las mejores, de qué supone cada uno de los principios que en ella aparecen y las implicaciones y principios adicionales a que cada uno de ellos da lugar, la de Daniel Glazer, Henry Lebowitz & Jason Greenberg, Fried, Frank, Harris, Shriver & Jacobson LLP and Practical Law Intellectual Property & Technology, “Data as IP and Data License Agreements”, Practical Law Company, 2013.

nacionales que los admiten (la Unión Europea, Méjico y Corea del Sur), o los creados artificialmente no se sabe bien con qué intereses TDM license (*text and data mining license*) que es de esperar que sólo acabe siendo la frustrada *non-nata* como por fortuna a veces pasa con las ocurrencias de los Comisarios y Directores Generales de la Comisión Europea, o los subrogados equivalentes en la praxis del derecho en los Estados que no tienen ese derecho *sui generis*⁵⁹.

2.- Esta protección legal (es decir, basada en las leyes y reglamentos) ha dado lugar a la creación de un derecho consuetudinario mercantil (*lex mercatoria*) basado adicionalmente en contratos, regulación (legal o contractual) de licencias y permisos de uso y sus retribuciones y limitaciones, que aseguran más allá de la ley pero sólo frente a quienes se contrata o se licencia el control (la propiedad en sus distintas firmas) de los datos regulando su acceso, uso y reutilización por parte del servicio web que accede a ellos, con variables en cuanto a la exclusividad o no, la subcontratación (*sublicensing*) o no, la agregación y combinación o no de datos sobre esos datos, las finalidades a las que se puede dedicar el acceso y el uso, la existencia o no de limitaciones territoriales, o incluso la obligación de tener situado el servicio web que puede acceder y usar los datos en determinado lugar o país o limitar o no los equipos y herramientas que debe necesariamente utilizarse cuando se usan los datos...

3.- Regulación de la propiedad (contractual o legal) de los resultados del uso de los datos o sus derivados.

4.- Regulación de los detalles de la provisión de los datos (causa del cibtrato y precio) y de la migración y transición de los mismos.

5.- Obligaciones de seguridad en el manejo de los datos por el servicio web impuestas por los proveedores de datos (*data security control*), incluidos el seguimiento y control (inspecciones y auditorías) de los servicios web y de las compañías que hacen uso de los datos.

6.- Controles del respeto a la privacidad y aplicación de las leyes reguladoras de la confidencialidad de los datos personales; y

⁵⁹ Vid., al respecto, las obras de Glazer *et al* y de Michael Carroll citadas posteriormente.

7.- Asunción o imputación a otros de los riesgos de errores en los datos con limitaciones unilaterales o en cláusulas contractuales (o incluso con simples clicks o ni siquiera eso –los ya citados “*clickwraps*” y “*browsewraps agreements*”-) de responsabilidades (a través de *waivers* y *disclaimers*, así como de *understandings*, *representations* y/o *warranties e indemnity provisions*).

Además, y ello es lo trascendental, el contenido real de cada uno de estos principios (o más bien “contenidos” usuales de este “Derecho de Datos Digitales” (*Data Law* o *Big Data Law*) y sus excepciones (*fair use*, *fair dealings*, *limitations and exceptions* ...según se trate de derecho norteamericano, de otros países o europeo) o limitaciones por razones de interés público o general, depende tanto en su formulación como en su aplicación de las distintas jurisdicciones nacionales, siendo prácticamente inexistentes los instrumentos de armonización (incluso interna en los estados descentralizados), supranacional o internacional. Luego su valor como tales principios es muy limitado y exige un análisis caso por caso.

¿Están los servicios web y las infraestructuras de datos de biodiversidad sometidos a este derecho?

Pues bien, el nuevo principio estructural de orden público que supone la *Open Science* en sus distintas variantes (*Data Sharing*, *Open/Public Acces*, *Open Source*...) sólo matiza y deroga (o prevalece) sobre la regulación del punto de partida (forzando un estatuto de dominio público o equivalente de los datos *versus* sus formas de apropiación -y con limitaciones-), pero desde luego, no hace desaparecer, ni siquiera en las *e-science research infrastructures* y los servicios web, incluidos los VRE labs que crean, la necesidad de ajustarse en su gestión a esos principios en función del país donde se sitúan (o sometidas a derecho internacional que es lo que suele ocurrir, como se ha visto), por mucho que todo este derecho esté en formación y, por tanto, tenga grandes vacíos a fecha de hoy.

Por eso precisamente conviene (cada vez más se aprecia su necesidad) ajustarse en su conducta a las directrices de todo tipo que están elaborando, en procesos abiertos y transparentes, instituciones más o menos prestigiosas y de autoridad, apoyadas por los Gobiernos, lo que no obsta a que realmente sean elaboradas mediante procesos *bottom-up*, es decir, con participación de los propios actores (incluidos los

investigadores y sus instituciones –“principio de responsabilidad” de los formulados por la RDA-CODATA-)

No es necesario reproducir en lengua castellana lo que ya está accesible en la web en inglés, al menos por ahora, mientras no esté finalizada y adoptada por múltiples instituciones de todo tipo su formulación, y se conviertan en los principios que rigen del “Research Data Law” o el “Big Research Data Law”.

La publicación de la que este artículo se hace eco sí los enuncia.

Y, pese a que quienes se han adentrado un poco en él desde una perspectiva académica no encuentran muchas variaciones con el antes descrito, lo que hemos denominado principios del “Big Data Business Law”, ni siquiera donde reina el principio contrario de *Open Science*⁶⁰, ello no es del todo cierto tampoco y desde luego no lo es en su punto de partida porque la esencia del “Research Data Law” es la no susceptibilidad de apropiación ni *ex lege* ni mediante confidencialidad o derecho de propiedad intelectual de base legal o contractual de datos sometidos al estatuto de acceso y reutilización en abierto de todo dato obtenido mediante financiación pública.

Por tanto resulta capital someter toda nueva infraestructura pública de investigación y sus servicios web al menos a cuatro grandes bloques de principios:

1.- Los de múltiples instituciones sobre Open Access (los más recientes, del 27 de enero de 2016 y además referidos a datos ambientales absolutamente relevantes para la biodiversidad son los Belmont Forum Open Data Principles for Environmental Change Research⁶¹)

2.- Los “Principles on the Legal Interoperability of Research Data” y sus “Implementation Guidelines”, del RDA-CODATA Interest Group on the Legal Interoperability of Research Data., que se pueden seguir en la web en general y participando en ese grupo⁶².

⁶⁰ Vid., por todos, Carroll M.W. (2015). Sharing Research Data and Intellectual Property Law: A Primer. PLoS Biol 13(8): e1002235. doi:10.1371/journal.pbio.1002235

⁶¹ <https://www.belmontforum.org/announcements/2016/belmont-forum-adopts-open-data-principles-environmental-change-research>

⁶² <https://rd-alliance.org/ig-rdacodata-legal-interoperability.html>

3.- Para datos (no sólo geoespaciales) del GEOSS los Data Sharing Principles⁶³ y sobre todo, para gestión de datos de su sistema, los diez Data Management Principles y sus Implementation Guidelines⁶⁴, éstas en proceso de revisión final.

4.- Justo al cierre de esta edición se ha recibido el texto adicional que, para datos de biodiversidad acaba de aprobar, EU BON⁶⁵, el proyecto de la Comisión Europea establecido de 2012 a 2017 que, en el marco de GEO BON y para los países de la UE, intenta avanzar en la definición en general de qué características técnicas (y legales), las Recomendaciones que deben seguirse para los datos de biodiversidad. Estas Recomendaciones⁶⁶ intentan armonizar los criterios para las publicaciones de los datos abordando: i) la movilización de datos de biodiversidad; ii) la remoción de obstáculos legales; iii) promoción del cambio de actitudes; y iv) actores que deberían adoptarlas (investigadores e instituciones de investigación, reguladores, agregadores de datos, entidades de financiación de la investigación y editores).

Además, para la puesta en marcha, en lo que a la biodiversidad se refiere, del *cluster* de políticas y derecho de gestión de datos que como infraestructuras deben poner en marcha ellas mismas como tales y sus servicios web, el panorama no está huérfano de precedentes ya que muchas de las doce Research Infrastructures participantes están operativas y tienen VREs en funcionamiento⁶⁷ y también hay ejemplos de VRE globales similares como el de la FAO para pesquerías⁶⁸.

⁶³ Su post-2015 status puede seguirse en <http://www.earthobservations.org/dswg.php>

⁶⁴ https://www.earthobservations.org/documents/geo_xii/GEO-XII_10_Data%20Management%20Principles%20Implementation%20Guidelines.pdf

⁶⁵ <http://www.eubon.eu/>

⁶⁶ Egloff W, Agosti D, Patterson D, Hoffmann A, Mietchen D, Kishor P, Penev L (2016) Data Policy Recommendations for Biodiversity Data. EU BON Project Report. Research Ideas and Outcomes 2: e8458. doi:10.3897/rio.2.e8458, descargable en <http://rio.pensoft.net/articles.php?id=8458>

⁶⁷ Por citar los de la europea LifeWatch pueden verse el de biodiversidad marina http://www.lifewatch.eu/Virtual_Research_Environments y para el de biodiversidad terrestre y aguas continentales puede verse la presentación de Julien Radoux <http://slideplayer.com/slide/9861587/>

⁶⁸ El de BlueBridge. Véase Pasquale Pagano. CNR. “Infrastructure Policies”. EGI Forum, Bari, November 2015, <http://www.slideshare.net/PasqualePagano/virtual-research-environments-as-a-service>

Pero lo más importante es que el modelo de análisis de los *clusters* de temas jurídicos que deben abordar, incluso en el supuesto más complejo, el de incorporación de datos de *citizen science*, su estudio y análisis lidera al resto de los sectores, hasta el punto de que las infraestructuras de datos biomédicos a nivel global y la política de *citizen science* de las grandes agencias federales norteamericanas estamos abordando una estrategia parecida, que está en fase inicial (la primera) pero que se prevé hacer muy rápido, a través de una adaptación de las elaboradas para la biodiversidad.

Se remite al excelente Primer (aunque algo necesitado ya de revisión pese a ser de agosto de 2013) elaborado por Anne Bowser et al, “Data Policies for Public Participation in Scientific Research: A Primer”⁶⁹, y por cortesía de DataOne y la National Science Foundation, se rehace ligeramente a efectos de reproducirlo de manera sintética el esquema que organiza de manera magistral (haciendo el puzzle a partir de sus piezas) el conjunto y subconjunto de reglas jurídicas vinculantes y “normas” (en el sentido antes expuesto) que a manera de red organizan la gestión de la complejidad que todo este grupo de principios de “*Big Research Data Law*” requieren en lo que a datos de biodiversidad se refiere.

⁶⁹ Anne Browser, Andrea Wiggins & Robert D. Stevenson, Data Policies for Public Participation in Scientific Research: A Primer, DataOne Public Participation in Scientific Research Working Group, August 2013
<http://www.birds.cornell.edu/citscitolkit/toolkit/policy/Bowser%20et%20al%202013%20Data%20Policy%20Guide.pdf>

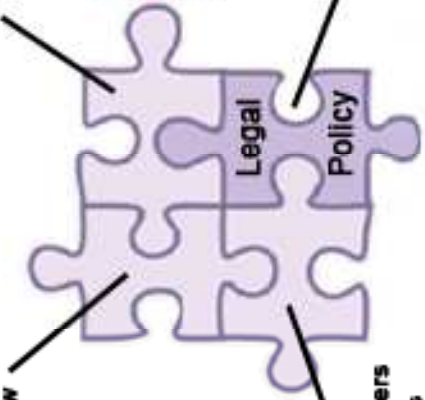
Legal Compliance

- With International Law
- With National Law
- With Local Law

Non- Legal Compliance

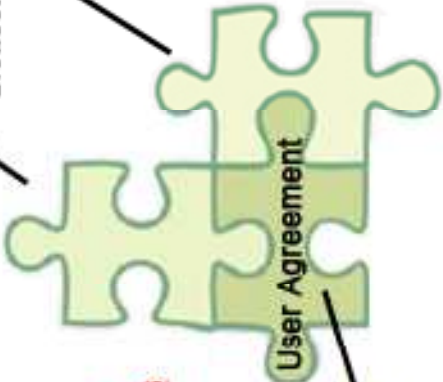
- With non-legal Principles

MANAGEMENT OF COMPLEXITY
NETWORK OF RULES AND NORMS
THE DATA POLICY PUZZLE



Liability

- Liability disclaimers
- Other disclaimers



Agreement Type

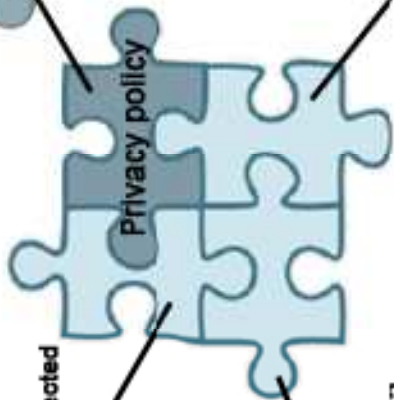
- Clickwrap
- Browsewrap

Accessability

- Location
- Point of access

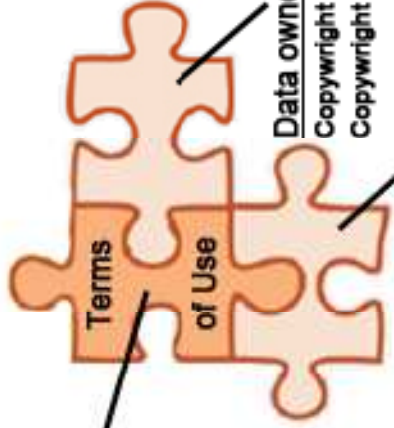
Usage Data

- Automatically collected



Volunteer data

- Data collected through registrations
- Data submissions



Data ownership

- Copyright of individual data
- Copyright of data sets

Data use

- Data access
- Data use