

# Ciencia + Transparencia = Ciencia 2.0

Autor beu  
jueves, 21 de diciembre de 2006

Antinio Lafuente, Madri+d Tecnocidanos

Enviado el lunes, 18 de diciembre de 2006 7:11

La convergencia del open access, open data, open source, open standars y open lab notebook, entre otros movimientos en marcha, hace viable la noción de ciencia 2.0.

Por encargo de Patricio Morcillo, director de la revista madri+d, escribí un artículo que acaba de aparecer sobre la la noción de ciencia 2.0. Presento aquí un extracto que resume lo sustantivo. Las nuevas tecnologías, al igual que están modificando lo que entendemos por cultura, también nos obligan a repensar lo que es la ciencia. O, dicho en otros términos, ¿está la web 2.0 sacando a flote una ciencia 2.0? Miremos donde miremos se impone la respuesta afirmativa.

El próximo lanzamiento de la revista PlosOne, perteneciente al grupo Public Library of Science, es un buen ejemplo de lo que está pasando. A nadie se le escapa que el sistema de control de calidad de los artículos científicos, el peer review, está en crisis. No es que haya dejado de ser útil, sino que son tan agudos sus problemas que ya se discuten abiertamente.

El open peer review que postula PlosOne es expresión de un movimiento de mayor calado y que está en resonancia con las movilizaciones favorables al open access y al open data pues, en la práctica, el primero es una autopista que conduce necesariamente al segundo. El asunto es que hoy las nuevas tecnologías permiten no sólo hacer experimentos que producen/demandan gigantescas masas de datos, sino situarlos también en un repositorio a disposición de quien los necesite.

El hecho, sin embargo, es que históricamente (sin entrar aquí en el vertiginoso proceso de privatización) se han producido grandes pérdidas debido a las deficientes políticas de archivado de datos. Y no estamos hablando solamente de seguridad, sino sobre todo de interoperatividad, pues la diversidad de tecnologías de conservación y las discontinuidades institucionales (cierre, cambio o abandono de bases de datos) han convertido el uso de la información pública en una carrera de obstáculos difícil de superar. Y la pérdida de datos implica despilfarro de recursos y de conocimientos.

Mucho más en nuestro tiempo que podemos diseñar/personalizar herramientas capaces de extraer información a conveniencia de las bases de datos accesibles. Baste aquí con considerar la red como un ingente repositorio para percibir la importancia de que la información sea accesible, pues ha sido a partir de esta deriva que Internet ha permitido la creación de miles de productos, muchos de ellos desarrollados por usuarios anónimos, que no cesan de agregar valor añadido al conjunto.

El caso de Amazon es paradigmático pues, tras obtener, igual que sus competidores, la información de los libros del registro ISBN optó por un modelo de negocio en el que los clientes podían añadir comentarios, incrementando exponencialmente el valor de la información ofrecida a través de su registro ASIN. Lo mismo está sucediendo ahora con la información cartográfica y la continua amalgamación de datos (mashups) que hacen los usuarios, mezclando los planos de ciudades con los datos/noticias de criminalidad, gustos musicales o precios de alquiler de la vivienda para ofrecer unas cartografías inverosímiles, dinámicas y colaborativas que inventan las más extrañas formas de ?vecindad?.

Podríamos seguir con nuevos ejemplos, sin que mejore la claridad del argumento que nos traemos entre manos: abrir los datos, no sólo es un requerimiento derivado de la doble necesidad de que la ciencia se acerque al viejo modelo de una República de Sabios y al que exige una democratización del conocimiento, sino que implica apostar por la oportunidad difícilmente discutible de que aparezcan nuevas e imprevistas formas de usarlos y conectarlos o, en otros términos, de crear conocimiento.

Los datos, en consecuencia, deberían ser algo que se encontrase en la web, antes que en el laboratorio. La web 2.0 llevará el sello Data Inside, una analogía con el Intel Inside del PC que domina la cultura del escritorio y que será reemplazado por la noción de la red como una plataforma global de computación. La web del futuro, la llamada web semántica, (ver Strange Attractor), sentenció no hace mucho Tim Berners-Lee será una red de datos (alternativa a la actual red de documentos).

Y, en efecto, la combinación de open access y open data demanda iniciativas que favorezcan el open source y el open document, es decir el software de código abierto y los formatos estándar. Los motivos son obvios, como ha explicado entre otros muchos Glyn Moody y Ian Foster. Para que la web funcione como una plataforma es imprescindible asegurar la interoperatividad entre todas las máquinas y todos los programas o, dicho en otros términos, que todos los objetos que circulen por la red puedan ser correctamente interpretados, cualquiera que sea el hardware o el software

que empleemos. La red, en consecuencia, debe ser neutral desde el punto de vista tecnológico.

Que un documento sea estándar quiere decir que ha sido codificado y archivado según protocolos abiertos y que, en consecuencia, pueden ser modificados y adaptados a nuestras necesidades. También significa (ver la interesante entrevista a Peter Murray-Rust en Reactive Reports) que no dependemos de ninguna empresa para editarlo, formatearlo, enviarlo o archivarlo, pues tales funciones deberían poder ejecutarse con cualquiera de los procesadores de textos o clientes de correo existentes en el mercado. Y lo mismo tendría que ocurrir con las bases de datos, pues de otro modo no podríamos diseñar herramientas específicas (API) para interactuar con ellas y extraer los datos que necesitamos. Y esto es importante porque las bases de datos no debieran condicionar la forma en la que puedan ser consultadas, pues siempre es concebible una utilidad insospechada por sus creadores y que demandará nuevas API cuyo desarrollo es impensable sin tener acceso al código fuente.

Pero hay más. Siempre se ha dicho que la replicabilidad es una de las características insoslayables del trabajo científico. Si esta condición debe ser mantenida en las ciencias experimentales, tenemos que preguntarnos por la pertinencia de programas informáticos de código oculto (cerrado) que funcionan como una caja negra en donde se introducen datos para que sean procesados y luego devueltos como un output o solución que no puede verificarse dada la imposibilidad de conocer las operaciones a las que han sido sometidas las cifras iniciales que le suministramos. No es que tengamos pruebas que justifiquen alguna sombra de desconfianza. Lo que aquí importa es la cuestión de fondo de si podemos usar herramientas científicas que están construidas para que no podamos saber cómo funcionan.

El asunto es delicado, pues no cuesta mucho imaginar instalaciones (centrales nucleares, laboratorios científicos o bases de datos fiscales, por ejemplo) que, siendo estratégicas para la seguridad nacional, deberían ser gestionadas por programas informáticos de código abierto, no sea que el país/corporación donde se redactó el software que los regula se enemigue o quiebre y deje de cedernos las actualizaciones necesarias para el correcto funcionamiento de todos los procesos. Así, soberanía (nacional), replicabilidad (científica), interoperabilidad (tecnológica) y código (lengua) son conceptos de mucha envidia política que se entrelazan de diversas e intrincadas maneras. La Ciencia 2.0 tiene que ver entonces con todas las tecnologías que favorecen la cultura de la transparencia y de la colaboración. Si le agregamos los dígitos 2.0 que aluden, como sucede con los programas informáticos, a la existencia de versiones antiguas, estables o en construcción, es porque queremos reivindicar no sólo el desideratum de una forma de saber en Beta, sino también la viabilidad de una forma de organización desjerarquizada, desnacionalizada, desinteresada, despatronizada y descorporativizada o, en otros términos, verdaderamente acéfala, cosmopolita, altruista, comunitarista y pública.

Usamos el meme (o marca, en el sentido más comercial del término) lanzado por O'Reilly por la provocación que supone conceptualizar la web como algo editable y siempre en construcción por los mismos internautas. Podríamos hablar, y de hecho se hace, de Biblioteca 2.0, Ciudad 2.0 y Futuro 2.0. Lo que estas denominaciones comparten es la voluntad de que converjan todas las tecnologías y prácticas favorables a la horizontalidad, la transparencia, la participación o cualquier otra forma de implicación de los usuarios en su gestión.

La ya citada PlosOne, por ejemplo, admite comentarios (como también alertas, etiquetas o enlaces a otros datos, textos o críticas) sin preguntar por la especialidad, cargo, institución o edad del crítico y, al igual que hay estudiantes que son editores de revistas de prestigio o hackers que son excelentes programadores sin pertenecer a ninguna estructura académica, también cabe esperar más de una sorpresa de estos experimentos incipientes.

La ciencia 2.0 seguirá necesitando investigadores, datos, publicaciones y gestores, pero quizás pueda contribuir a evitar las muchas y disparatadas formas de nepotismo, ocultamiento y privatización, por no citar los cada día más frecuentes casos de despilfarro, duplicación o fraude. No estamos refiriéndonos a una utopía inalcanzable, la deriva que está tomando la ciencia, prosperará en la medida en que entren en circulación proyectos y tecnologías que la hagan viable.

Ya hemos hablado de la eficacia del broadcasting y del open peer review, ahora vamos a detenernos en el OpenWetWare (OWW, nacido como Endipedia en mayo de 2005), una aventura que promueve un grupo de investigadores del MIT. Quien haya llegado hasta aquí entenderá perfectamente de qué se trata. Consiste en apostar a fondo por la cultura abierta (openness) y compartir en el ámbito de las biotecnologías ideas, protocolos, datos y prácticas.

Quienes sospechen que algo así sólo puede funcionar en un mundo ideal deberían leer las páginas en donde se responde a este tipo de preocupaciones. Y sí, aunque parezca mentira, se argumenta que cuando alguien se apropia de una idea y lo oculta, la mejor solución es cambiar de asunto y a otra cosa, pues para los promotores es más probable que los visitantes incluyan comentarios breves y pertinentes que sirvan de ayuda. Desde la perspectiva del bien común, está claro que estamos ante una muy buena iniciativa pues en el peor de los casos (cuando hay plagio) el conocimiento crece. En los demás, se avanza hacia formas de colaboración amenazadas.

Los perjudicados saben que se trata de una apuesta a medio plazo y que siendo wiki la tecnología de intercambio empleada, queda registrado y a la vista el historial de todas las contribuciones, lo que permitirá en todo momento asignar méritos y restablecer la reputación. Hace unos días, el 10 de noviembre de 2006, había 265 usuarios registrados,

29 administradores, 2725 páginas, 15 laboratorios implicados y contaba con unas 1000 visitas diarias. No es mucho para las cifras a las que nos tiene acostumbrados la web, pero nadie sabe todavía si estamos al comienzo de otras de esas iniciativas que introducirán cambios irreversibles.

Los que se apresuran a decir que eso no ocurrirá en ciencia, deben saber que hay mucha gente competente movilizándose para que los cuadernos de laboratorio se hagan públicos o que los papers colectivos se escriban con wiki, desde el brainstorming inicial hasta su envío a un repositorio público o revista. Iniciativas que, huelga decirlo, solo serían plenamente operativas si se avanza en la estandarización de los documentos, las bases de datos y los cuadernos de notas.