

Periodismo de Datos. Explicando historias con Software. Conferencia con Al Shaw.

Karma:

Bienvenidos y bienvenidas a las terceras jornadas de periodismo de datos que celebramos. Estas terceras jornadas son especiales, porque por primera vez estamos ya en cuatro ciudades. Eso quiere decir que, en este momento, en Madrid, aquí en Barcelona, en Almería y en Iruña, Pamplona, hay gente interesada en periodismo de datos que está asistiendo a un programa en directo, como nosotros. Yo quería empezar enseñando esta imagen, que, no sé, muchos de vosotros la conoceréis, es la imagen de Florence Nightingale, que ya hace más de 100 años describió cómo la guerra de Crimea, demostró cómo en la guerra de Crimea había muchos soldados que no solamente morían en el campo de batalla, sino también por las enfermedades contagiosas que cogían en los batallones y en las enfermerías, ¿no? Ella era enfermera y explicó de esta manera estos datos, ¿no? Y quería empezar con esta gráfico así explicado muy de pasada, porque el periodismo de datos no es algo nuevo, sino que ya, como nos demostró Florence Nightingale, pero también Minar y otros estadistas de hace más de 100 años que se podían hacer infografías y representaciones con los datos, lo que sí es nuevo es la manera en que hoy estamos trabajando los datos desde el periodismo, la manera en que accedemos a las bases de datos, la manera en que accedemos a estas herramientas digitales de visualización, la manera en que medios como ProPublica con imágenes satélite nos muestran y nos explican qué está pasando en una zona de EEUU, etc., ¿no? Esta aventura que empezamos de las jornadas en 2013 Mar Cabra y yo, la ministra de investigación de referencia que tenemos en España. Me hace gracia después de 3 años echar la vista atrás porque en el 2013 apenas se hablaba, ni se escuchaba, de datos, no teníamos ley de transparencia en España, no la teníamos tampoco en Cataluña ni en otras CCAA.

Configurar el programa de aquel año, que no hace tanto, solamente de hace dos años atrás, fue un auténtico reto para las dos, porque convencer a los periodistas para que vinieran a unas sesiones era una tarea titánica; intentar encontrar algún ejemplo de un medio de comunicación que estuviera trabajando con datos era aún más difícil. Este año lo difícil ha sido todo lo contrario: crear una parrilla, un programa donde tuviera cabida tantísima gente que ya está trabajando con datos aquí en España y donde pudiéramos traer a gente de otros países para que nos dieran su experiencia. Adolfo Antón, que ha cogido este año el relevo de Mar Cabra en Madrid, hablábamos de, le decía yo, Adolfo, esto es como hacer un puzzle, ir encajando piezas para que todos los ponentes y expertos puedan estar aquí dentro.

Bien, si veis el programa de mano que tenéis, este año tenemos muchos más logos que nos acompañan, nuestro agradecimiento a todas las instituciones, centros, organizaciones y empresas que nos han apoyado para que tengamos un programa en cuatro ciudades de casi una semana de actividades. Nuestro especial agradecimiento al centro de cultura contemporánea, que nos acoge por tercer año, y al estupendo equipo de Juan Isúa con María en la coordinación, que hace posible que todo esto vaya adelante por supuesto. A la facultad de comunicación Blanquerna, donde estaremos mañana, sábado y el domingo haciendo talleres, y además al Col·legi de periodistas que este año ha sido la sede, está siendo la sede de la HAKATONJEN que se está celebrando en este momento y ayer hicimos un estreno de estas jornadas en Caixa Forum. Son unas jornadas, como veréis en el programa, con muchos expertos, gente que ha decidido venir desinteresadamente a estas sesiones. A todos ellos, nuestro especial agradecimiento porque les decimos que vengan de Madrid a Barcelona en un tiempo récord, el AVE esta funcionando espero que muy bien, tenemos ponentes que han llegado ahora mismo después de coger un tren a las 7 de la mañana y durante el día irán llegando ponentes que están en este momento en Madrid, Al Shaw y Elisabetta Tola ya han sufrido esto de Madrid-Barcelona, porque estrenaban anteayer en Madrid. A todos ellos, muchísimas gracias, por supuesto. Bien, este es un proyecto, como sabéis organizado por la *Open Knowledge Foundation*, es una organización internacional que está en mas de 30 países. Los que formamos parte de esta organización, esta es la OKFN Spain, lo que pretendemos es organizar sesiones para dar a conocer, eso, precisamente el conocimiento abierto, y organizar más sesiones para que más gente aprenda, más gente sepa sobre un tema en concreto. Y yo, como miembro interesada en el periodismo de datos soy la que siempre lío todas las sesiones, talleres, hackathon sobre los datos. Quisiera que, si tomáis el programa de mano, os voy a comentar un poco cuestiones prácticas y repasar lo que se nos viene encima, porque esto es una maratón, como otros años. Como os he dicho, en este momento se está realizando el hackathon, el hack day GEN. Tenemos en Barcelona 2 equipos, la Vanguardia y el Diario Ara, hay en Madrid otros 5 equipos, y los 7 equipos de periodistas, programadores, diseñadores, están compitiendo para ir a la final, que la GEN celebrará el día 17 aquí también en Barcelona, una final con todos los equipos finalistas de muchos países del mundo, y la verdad es que es un reto poder celebrar esta hack day dentro de las jornadas. Ayer tuvimos, como os he comentado, el debate con Elisabetta y con Al Shaw en Caixa Forum, y hoy es el día para que disfrutéis conferencias, charlas, tendremos un *Hacks and Hackers* con cervezas al final del día, esto lo digo ahora para que nadie se vaya de aquí hasta la tarde-noche... Tenemos un desdoblamiento de programa por la tarde, dedicado al Open Data, con expertos y profesionales que ya tienen empresas y están trabajando con los datos abiertos, y, bueno, hay *streaming*, en este momento se está grabando para que todos los que no puedan estar aquí puedan seguir las charlas por *streaming*. Y mañana va a ser un día de talleres a tope, el programa casi no cabe en este tríptico de mano con expertos que hoy vais a escuchar y van a estar mañana también haciendo talleres, además de otra mucha gente que vendrá especialmente para esas sesiones de ordenador. Todos tenéis que venir con ordenador a los talleres, eso sí, que nadie venga sin su ordenador. Bien, yo no sé si os atreveréis a llegar hasta el final y acompañarnos hasta el domingo a las 2 de la tarde, donde se acaba esta maratón

de jornadas, espero que sí que seáis muchos. Si teneis dudas sobre los talleres, el contenido de los talleres, o los ponentes que van a participar, lo teneis en la pagina de periodismo de datos de las jornadas. Dentro de la sección ponentes teneis todos los detalles de ellos. He pasado antes unas etiquetas, espero que queráis jugar y os acabéis relacionando entre vosotros y que conozcáis al menos a 5 personas durante el día de hoy y compartáis los perfiles y a qué os dedicáis. Bien, lo único que me queda decir es que este evento está organizado con la mejor de las intenciones, con la idea de compartir el conocimiento que tenemos entre todos, de hacerlo accesible, a cuanta más gente, mejor, porque es la única manera de ir avanzando, mientras más sepamos entre todos. Muchas gracias, espero que disfrutéis del evento, y ahora os voy a dejar con Al Shaw, que paso a presentarlo: Él viene de ProPublica, medio independiente de EEUU, es uno de los medios de referencia que hay en este momento. Los trabajos que ellos hacen no son trabajos que se hacen en un día ni en un mes, ni en 5 minutos, son trabajos de un año o año y medio, a veces 3 años, luego vais a tener oportunidad de hablar con él y que os cuente. Él tiene un perfil profesional que nos gustaría tener a muchos de nosotros, porque lo llamamos híbrido, tiene perfil periodista, pero también programador, pero también diseñador. Dentro del equipo, él se encarga de hacer lo que llaman las *News Applications*, las aplicaciones para hacer noticias, y últimamente ha trabajado mucho con imágenes satélite que nos va a explicar, sobre todo en su último, uno de los últimos reportajes en los que ha trabajado, *Losing ground*. Yo ya no quiero hablar más, Al, es tu turno, gracias por estar aquí.

Al Shaw:

¿Hola? ¿Sí, hola? Gracias, Karma, me alegro de estar aquí. Voy a hablar en inglés, si no lo entendéis, puedo hablar un poco más despacio. Levantad la mano y preguntad, lo intentaré repetir. Soy Al, trabajo en ProPublica, que es una organización de noticias de investigación sin ánimo de lucro en Nueva York, comenzamos en 2008 y nos gusta decir que hacemos periodismo por el interés público. No cubrimos exclusivas, solo noticias centradas en la responsabilidad, con fuerza moral, que pueden cambiar algo en el mundo. Todas las historias que publicamos tienen la intención de causar un impacto. Así es como recaudamos dinero, es en lo que les decimos a los donantes que están invirtiendo. Ese es el tipo de trabajo que hacemos en ProPublica. Una de las formas en las que lo hacemos, la que más, es con datos. Trabajo en un equipo llamado de Aplicacionistas de Noticias, podéis encontrar nuestro trabajo en la pestaña de Herramientas y Datos de nuestra web. Usamos datos para que los gobiernos y empresas sean responsables de maneras que no podrían conseguirse solo citando fuentes o con noticias tradicionales. Contamos historias, aunque esto es un poco genérico, porque todas son historias, pero nosotros las contamos con texto, gráficos, que son pequeñas aplicaciones de una página con un poco de datos, y con las Aplicaciones de Noticias, que son grandes e interactivas bases de datos en las que se puede buscar tu historia de entre los datos. En nuestro equipo también escribimos texto, hacemos fotos, pero nos gusta hacer gráficos y también este tipo de Aplicaciones, de las que os voy a hablar en breve.

El proceso para crear una aplicación de noticias, o incluso para hacer cualquier cosa con los datos, es bastante largo, intenso, y tiene muchos pasos. El primero es conseguir los datos, que no es fácil. Una forma es a través del sistema de libertad de información que tenemos en EEUU, que nos permite obtener información de funcionarios públicos. También hay muchos datos de libre acceso, por lo que puede variar, desde ser supersencillo, como entrar a la web del gobierno y descargarlo, hasta hacer una petición oficial, esperar un año, comprobar que los datos están en el formato adecuado, que muchas veces no es así y hay que intentar cambiarlo al correcto, y después continuar con el siguiente paso, que es limpiarlos.

Esto es algo muy complicado y también puede ir desde cambiar algunas columnas a variar el formato, de PDF a Excel, unir muchos tipos de datos, es una tarea difícil. El siguiente paso es analizar los datos. Solo se puede comenzar a analizar datos una vez se han limpiado, es aquí cuando se intenta encontrar la noticia dentro de ellos, que a veces puede no haber. Para cuando has llegado a la fase de análisis a veces es demasiado tarde, has llegado demasiado lejos obteniéndolos y limpiándolos hasta que lo presentas, que es lo que llamamos aplicación de noticias.

Perdón. Las aplicaciones de noticias son bases de datos grandes e interactivas en las que puedes encontrar tu propia historia dentro de un set de datos más grande. Os voy a enseñar un ejemplo de una aplicación que hemos hecho llamada *The Opportunity Gap*. Esta app trata la desigualdad en la educación en Estados Unidos, si los diferentes estados proveen de un acceso igualitario a la educación y a la oportunidad educativa. Se basa en un grupo de datos que obtuvimos del Departamento de Educación de EEUU. Venían con muchos centros y distritos distintos y el problema con estos datos fue que los distritos daban su propia información, así que los datos provenían del Ministerio de Educación, pero cada distrito lo había presentado ante el Gobierno Federal, por lo que tenías que confiar en que los datos de cada distrito fueran correctos. Comprobamos que no siempre era así; algunos distritos decían, por ejemplo, que tenían mil clases de nivel avanzado, cuando en realidad el máximo que podían tener era de 30. En esos ejemplos teníamos que llamar al distrito en cuestión y consultarles por qué decían esto si el máximo eran 30 clases de este tipo, por lo que tuvimos que hacer esto con 60.000... No llamamos a los 60.000 centros, pero tuvimos que buscar lo que llamamos 'el absurdo', que implica, a la hora de limpiar los datos, localizar qué está mal y cómo podemos resolverlo.

Para encontrar 'el absurdo' hacemos varias cosas diferentes, como por ejemplo: buscar cuentas y totales, mala geocodificación, como que no incluyan direcciones, o sí las incluyan, pero no los códigos postales o países... Hay muchas estrategias que adoptamos para limpiar los datos antes de empezar a analizarlos. Se puede ver la metodología que usamos en este proyecto y leer estas guías sobre como asegurar los datos que uno de nuestros antiguos editores escribió.

Después de limpiar los datos llega mi parte favorita del proceso, que es crear estas aplicaciones de datos, estas bases de datos interactivas con una noticia dentro de un grupo de datos más grande. Decimos que hacemos las noticias con *software* en vez de con palabras y fotos. Estos grupos de datos interactivos, las aplicaciones de noticias, tienen muchos componentes, pero los principales son la vista lejana y la vista cercana.

Igual que cuando estás leyendo una noticia normal, al inicio de la misma hay un titular y un párrafo que te da lo que llamamos el resumen, y luego viene otro párrafo que te da el tema principal, la información importante de la noticia, y esto funciona igual en las aplicaciones. La vista lejana te da primero el título grande que explica qué está ocurriendo y luego el párrafo resumen. En esta aplicación, *The Opportunity Gap*, puedes buscar por todos los Estados del país y ver qué oportunidades educativas ofrecen sin importar el nivel de ingresos. Se puede ver que en este ejemplo el primer estado es Alabama, los mejores ofreciendo acceso a clases avanzadas; que el segundo estado es Alaska, y puedes seguir avanzando y consultar tu estado, en este macro nivel, para comprobar que cada estado es diferente del anterior al ofrecer acceso a estas clases. Lo siguiente que puedes hacer es entrar en la vista cercana. Cuando lees una noticia tradicional y avanzas por la página puede que llegues a una anécdota: alguien a quien los periodistas encontraron que ilustra muy bien la noticia. Ves citas de esa fuente, es el protagonista de la noticia y te guía por esta novedad nacional. En las aplicaciones de noticias tú eres esa anécdota protagonista, te permitimos buscarte, encontrarte en ella. Cuando buscas tu centro o tu ciudad ves datos sobre él y un párrafo con información. Trabajamos con una empresa llamada *Narrative Science* que escribió 60.000 historias, una sobre cada centro del país, y puedes leerlo, hay 60.000 de estas. Puedes pensar que esta aplicación de noticias tiene, entonces, 60.000 noticias distintas, pero de hecho tiene más, ya que puedes ir aún más profundo y comparar ese centro con otros cercanos, o con centros en los que hay más o menos pobreza. Puedes pensar: “Bien, yo voy a este centro, es intermedio, el 38% de alumnos tiene almuerzo a precio reducido, que es nuestra medida para la pobreza, por lo que el 38% se beneficia de ese descuento o gratuidad, pero el 74% aprueba las clases de nivel avanzado, por lo que van bastante bien”. Por otra parte, en este centro de la izquierda el 89% de los alumnos tiene descuento o gratuidad en el almuerzo, por lo que es bastante pobre, donde el 8% aprueba las clases de nivel avanzado, y por otra parte hay un centro muy rico en el que solo el 1% tiene descuentos o gratuidad en la comida y el 87% de los alumnos aprueban las clases de nivel avanzado. Así se puede ver dónde está tu centro en perspectiva.

También hay mapas, este es uno que me gusta bastante, porque muestra dos vistas opuestas del sur de California. Podéis ver que las partes que son rojas en un mapa son mucho más claras en el otro, porque las partes rojas de un mapa muestran pobreza más alta y las blancas del otro muestran el número de alumnos inscritos en clases de nivel avanzado. Se puede ver que Los Ángeles es un lugar donde la oportunidad educativa no es muy alta, si vives en un distrito rico probablemente vayas a acceder a una mejor educación que si vives en un distrito pobre, que es lo que esta aplicación de noticias intenta mostrar.

Aun así, nosotros vamos un paso más allá, nos gusta decir que las aplicaciones de noticias generan noticias, porque periodistas locales pueden descargar estos datos y redactar sus propias noticias basándose en ellos. Pueden descargar el archivo CSV de forma gratuita y republicarlo, pueden usarlo como base para sus propias noticias o pueden usar la misma aplicación para crear una historia. Eso es lo que muchos periodistas hacen, ya que la mayoría no saben usar datos o trabajar con ellos. Por eso

creamos la aplicación, para que los reporteros puedan buscar centros cercanos a su área de trabajo, encontrar la noticia y redactarla para sus medios locales, cosa que muchos hicieron. Estas organizaciones y estaciones de televisión locales usaron la App y redactaron noticias locales sobre esto. Es una de las formas en las que estos datos se conocieron, no solo por gente que entrase a nuestra app, sino también por los que leyeron las noticias en sus medios locales. También contamos con una tienda de datos, donde la gente puede comprar o incluso descargar datos de manera gratuita. Si pasamos mucho tiempo intentando limpiarlos y trabajando con ellos, cobramos un poco por ellos, porque nos ha llevado mucho tiempo pasarlo al formato correcto, pero si simplemente hicimos la petición al Gobierno y nos lo han facilitado, lo ofrecemos gratuitamente, porque no tendría sentido hacer que todo el mundo pase por el mismo proceso si no tienen que hacerlo.

Por esto, las aplicaciones de noticias generan noticias, pero a veces hay un tercer paso más allá en el que una noticia genera una aplicación que a su vez genera nuevas noticias. Un ejemplo de esto, que a mí personalmente me gusta mucho, es esta noticia que vimos en el New York Times. Nos gustó y pensamos que quizás habría un modo de programar una aplicación sobre ella. Es una noticia sobre supergrupos, que son independientes y que no están vinculados a campañas políticas, o que no deberían estarlo, y que precisamente porque son independientes, pueden gastar todo el dinero que quieran en campañas. Esta noticia del New York Times demostró que muchas de estas consultorías que los supergrupos estaban usando también estaban contratadas por campañas oficiales, por mucho que supuestamente no debieran estar trabajando juntas, y a veces trabajaban incluso en el mismo edificio.

Pensamos que era interesante y que quizás podríamos averiguar más analizando los datos de la comisión federal, para ver quién estaba contratando a qué consultorías, ya que estos supergrupos y estas campañas deben informar a la comisión federal de las elecciones. Entonces desarrollamos esta lista interactiva que permite ver las diferentes campañas, supergrupos y grupos independientes en un lado y a los sitios contratados, las consultoras en el otro. Con esto pudimos publicar, por lo que acabamos todo el ciclo: leímos una noticia, creamos una aplicación basándonos en su esquema y pudimos publicar otra noticia nosotros con los datos obtenidos sobre una empresa llamada *Targeted Victory* que localizamos usando nuestra aplicación de noticias y que había sido contratada por la Campaña de Mitt Romney y por un grupo independiente que lo apoyaba, pero que en teoría no podía trabajar con la campaña. Descubrimos muchos de estos grupos y eso es lo que publicamos.

Esa es otra forma de usar datos. Una cosa en la que probablemente no estéis pensando es en usar datos para encontrar a otras personas. Bien, ponemos los datos en la web, la gente la visita y se encuentra en la noticia, pero, ¿y si queremos escribir los datos de manera que podamos encontrar esa anécdota sobre la que queremos publicar la noticia? También podemos usar los datos. Hace un par de años estuve trabajando en una noticia sobre el huracán Sandy, una gran tormenta que hubo en Nueva York y que causó muchos daños. Gran parte de estos ocurrieron porque los mapas de inundaciones que estaban en vigor cuando llegó la tormenta no eran suficientemente buenos, no eran

precisos y no se habían actualizado. Los datos en los que estos mapas se basaban eran de 1983, pero los mapas en sí se habían actualizado en 2007, sin que mucha gente lo supiera, con los datos de 1983. Cambiaron algunas cosas, pero en general se trataban de los datos de 1983 aunque los mapas indicaban 2007. Creímos que mucha gente habría pensado que los mapas eran actuales cuando compraron sus casas, ya que no sabía que los datos que había detrás de esos mapas eran tan antiguos. Por esto, decidimos buscar gente que compró su casa después de 2007 pensando que estaba a salvo, pero cuyos hogares resultaron dañados por el huracán. Después del huracán, el gobierno publicó nuevas áreas de inundación, que son estas zonas marrones, que predicen de forma mucho más precisa dónde iba a golpear la tormenta, usando datos nuevos. Por lo que pensamos: “¿Quién habría estado a salvo si el Gobierno hubiera usado datos buenos en 2007 en vez de reutilizar los de 1983?”. Escribimos un algoritmo, este es el pseudocódigo, para encontrar a gente afectada, gente que vivía en la nueva zona de inundación, no en la antigua, y que hubiera construido o reformado su casa después de 2007.

Encontramos a esta pareja que vivía en un barrio llamado Sheepshead Bay y eran el ejemplo perfecto: siendo muy precavidos, habían instalado una bomba de agua pequeña porque pensaron que quizás les llegaría algo de agua al sótano, lo que los mapas de 2007 decían, pero su sótano se inundó totalmente, algo que, lógicamente, no les gustó. Por ello, se convirtieron en los personajes de nuestra noticia a través de este análisis de datos. Hicimos varios mapas para esta noticia, a la izquierda está el barrio en el que ellos vivían, Sheepshead Bay en Brooklyn, pero esto pasó por toda la ciudad; hicimos tres mapas que mostraban todas esas marcas negras, los edificios que sufrieron daños en el huracán.

También decidimos ir un paso más allá y desarrollar una aplicación de noticias, porque pensamos que si había afectado a esa pareja podría haberte afectado a ti también, y, a través de nuestro análisis, averiguamos que los mapas de Nueva York y Nueva Jersey que estaban en vigor cuando llegó el huracán eran diferentes entre sí a la hora de predecir mejor o peor dónde iba a golpear el huracán, porque se habían actualizado en épocas diferentes sin ningún orden lógico. Muchos se habían actualizado hacía muchísimo, algunos de los condados ni siquiera tenían mapas digitales, son los grises que aparecen ahí. Entonces puedes buscar tu condado y ver exactamente lo que te pasaba, si vivías en un área que los mapas de 2007 no predecían, pero ahora, con los nuevos, sí estabas en zona de inundación. Bien, pues este es un ejemplo de cómo usar los datos para encontrar personas y además desarrollar una aplicación a partir de ello. Nos gusta ceder nuestro código además de nuestros datos, tenemos la tienda de datos, con datos descargables que proceden de nuestras aplicaciones, pero también nos gusta ceder el código que hace funcionar nuestras aplicaciones. Por ejemplo, para los mapas de los que estaba hablando se puede usar el software que creamos que se llama *Simple Tiles*. También hemos creado fuentes para los estados de EEUU y tenemos mucho código distinto. Normalmente, cuando creamos una aplicación que puede tener un componente de código que podría ser valioso para más gente fuera de ese único uso intentamos extraerlo para que más gente pueda reutilizarlo con otro propósito en sus propias aplicaciones.

Estos son algunos de los ejemplos de las cosas que hacemos, pero quiero hablar más sobre un proyecto que hice el año pasado que prácticamente consumió todo mi año. Se llama *Losing Ground* . Es sobre cómo al sudeste de Luisiana, por una serie de factores, principalmente originados por el hombre, una gran parte del estado, aproximadamente un tercio, está desapareciendo y hundiéndose en el suelo, y la mayoría de los habitantes del país ni siquiera lo sabían, juzgando por la respuesta que tuvo la noticia. La gente estaba impactada a pesar de que los medios de Luisiana hubieran estado escribiendo sobre ello desde hacía bastante. Lo que me hizo investigar esto, además del hecho de que he estado viajando por Luisiana, fueron dos documentos distintos: uno, que fue el que más me asombró, fue este, que publicó una Agencia del gobierno estadounidense, creo que la Administración oceánica y atmosférica nacional. Este documento muestra toda la tierra que se ha perdido desde 1932 y que se perderá hasta 2050. Es una gran cantidad, y si esto ocurre, la ciudad de Nueva Orleans será casi una isla en esta península, rodeada de kilómetros y kilómetros de agua, en lugar de pantanos que protegían de los huracanes. El otro documento, también sorprendente, fue este mapa del sudeste de Luisiana que muestra todos los lugares que el gobierno ha quitado de los mapas oficiales por la pérdida de tierra. Son bahías, puertos, arroyos, ríos... que se han eliminado de los mapas porque esa tierra se ha perdido en los últimos 100 años.

Empecé a leer mucho, hay muy buenos libros sobre el tema, uno de mis favoritos se llama *Bayou Farewell* y trata de la cultura Cajún en Luisiana y sobre cómo la gente que vive allí se está adaptando a esta pérdida. Este libro me inspiró mucho a trabajar en esto y averiguar por qué estaba ocurriendo, ya que saber esto me permitiría contar la noticia y ver cómo la gente estaba lidiando con ello. Bien, esto es lo que está pasando: en los últimos 7.000 años, el río Misisipi, que pasa por Nueva Orleans, ha cambiado mucho, lo que ha creado esa forma de bota ahora icónica del estado. Si alguien conoce Luisiana, sabrá que es como una bota, y esto se debe a que el río ha cambiado su curso durante miles de años y ha depositado sus sedimentos de tal manera que se ha creado esta bota.

Bien; en los últimos mil años se ha asentado en la posición actual, pero a inicios del siglo XX hubo una inundación masiva. El río, que estaba comenzando su siguiente fase en la que iba a ir más allá de Nueva Orleans hasta el oeste, hacia el río Atchafalaya, se inundó y no llegó hasta Nueva Orleans porque habían colocado unos diques bajo la ciudad para aliviar la presión, pero querían que el río siguiese pasando por Nueva Orleans, que era un puerto muy importante. Si el río hubiera hecho lo que pretendía y se hubiera unido al Atchafalaya, Nueva Orleans no habría sido un puerto, lo que no podían permitir. Entonces, lo que hicieron fue crear un sistema de diques muy altos que mantenían al río en su lugar y le obligaban a bajar por su recorrido tradicional, por el que el río no quería ir, para evitar las inundaciones y permitir que los barcos accedieran. También establecieron algo llamado estructura de control del río antiguo, que permitía que algo de agua bajase por el Atchafalaya y la demás por el Misisipi; pudieron controlar cuánta agua bajaba por cada sitio. Debido a esto, el sedimento del río, que durante miles de años se había depositado creando esa forma de bota, ahora no estaba depositándose en los pantanos, sino en los diques y por el canal del Misisipi. Como estaba siendo forzado a desviarse y dado que los diques llegaban hasta bien entrados

en el océano, los sedimentos, que creaban los pantanos, estaban cayendo más allá de la plataforma continental y perdiéndose para siempre. Es decir, que todas estas zonas donde la gente vivía, esas áreas blancas, se vieron privadas de los sedimentos que formaban su tierra y estos se estaban perdiendo para siempre.

Ese es un factor. El segundo es que a inicios del siglo XX encontramos petróleo en el sur de Luisiana y empezamos a construir tuberías por los pantanos. En aquel momento nadie pensaba que los pantanos fueran valiosos, pensaban que eran básicamente lugares para los cocodrilos y un páramo, sin ninguna utilidad, así que mucha gente se enriqueció vendiendo sus tierras a petroleras que, sin ningún reparo para con la fauna o los pantanos empezaron a construir las tuberías. Así es ahora, estas tuberías cruzan todo el golfo, hay muchísimas. Y la forma en la que originalmente se construían era de la forma más destructiva para el medio ambiente: para poner la tubería creaban canales, que, como se puede ver en la imagen, eran mucho más anchos que la tubería en sí. Para hacer los canales cogieron tierra del suelo y la pusieron en las orillas para crear lo que se llaman orillas de desechos. No volvieron a colocar la tierra, la dejaron allí, y, dado que los pantanos son tan mullidos, la tierra que se colocó encima causó que se hundieran, puso presión en la tierra de fuera de los canales. Al cabo del tiempo descubrieron que había una forma mejor de realizar esto, que era solo excavar lo que necesitaban para poner la tubería y después volver a colocar la tierra al acabar. Para entonces, quizás unos veinte años después, era demasiado tarde. Podéis ver lo que ocurre cuando kilómetros y kilómetros de pantano han sido tratados así, porque la tierra se empezó a hundir; lo que originalmente eran pequeños canales ahora eran mucho más grandes y después se convirtieron en estanques, lagos, y finalmente mar abierto. Ahí se puede ver alguna infraestructura petrolera posada en ese trozo de tierra que apenas existe porque los canales se han convertido en mar abierto. No hay datos oficiales sobre los canales del sur de Luisiana, la única forma de saber dónde están es por unos pocos estudios y grupos pequeños de datos. Hablamos con un tipo que hizo su tesis sobre esto e hizo a mano los canales de este pequeño fragmento de bahía entre dos ríos. Al hablar con él nos percatamos de que este patrón era similar al resto del Golfo y región de Nueva Orleans y podíamos extrapolarlo. Podéis ver la densidad de canales que hay en esa zona.

Queríamos contar esta noticia y averiguamos que el Gobierno tiene muy buenos recursos de imaginería aérea y satélite, por lo que estas son mis fotos favoritas de las que encontramos disponibles gratuitamente a través de la USGS (La encuesta geológica de Estados Unidos). Lo llamamos La rueda del carro, es una zona muy al fondo del río Misisipi donde se encontró un domo salino bajo el que había muchísimo petróleo y alrededor del domo excavaron un canal inmenso para comenzar a perforar pozos a su alrededor. Al final, ese canal se agrandó y para 2013 se puede apreciar que es mayoritariamente agua que entra y sale de donde una vez hubo un pequeño canal. El patrón es sorprendente, al ver lo que ha pasado con el tiempo, se puede ver muy bien con estas imágenes aéreas. Los pantanos son tremendamente importantes, como decía antes. No solo por la fauna, la pesca, o por su belleza natural, sino porque también sirven como defensa contra los huracanes, así que por cada pantano que hay entre el mar abierto y la ciudad, los huracanes se ralentizan, se debilitan un poco antes de llegar al sistema de diques que rodea la ciudad, por lo que esto podría dañarla. Y, por esto,

una de las autoridades en el sistema de diques, las juntas locales que controlan los diques que rodean Nueva Orleans, denunciaron a 97 compañías petroleras y de gasóleo por dañar los pantanos, y el caso aún esta abierto en los juzgados.

Para poder redactar esta noticia confiamos en imágenes satélite porque era la mejor forma de hacerlo. Específicamente con el Landsat, un satélite del Gobierno estadounidense que cuenta con muchas bandas diferentes: las tradicionales rojas, verdes y azules y además infrarrojas, que recogen luz que no es visible al ojo humano. Hablaré sobre cómo trabajar con este tipo de imágenes en mi taller... creo que me estoy quedando sin tiempo.

Quiero enseñaros cómo hicimos el mapa que aparece en la primera página de la aplicación. No queríamos usar esta imagen, que es la que se encuentra, por ejemplo, en Google Earth, una imagen tradicional RGB, porque hay tantos sedimentos en el agua del río Misisipi que los sedimentos parecen agua, parece un montón de barro. Por eso, la iteramos varias veces, esta sería una segunda iteración, y al utilizar las diferentes bandas fuimos capaces de mostrar la diferencia entre tierra y agua claramente. Usamos RGB y también una máscara para el agua con una banda infrarroja para que se viera esa diferencia y también cómo se está perdiendo la tierra de un modo en el que no se vería con Google Earth. Hay muchos recursos que indican cómo procesar imágenes de este tipo, de los que os enseñaré algunos en el taller si decidís venir.

Os voy a enseñar rápidamente esto: así se ve hoy, pero obtuvimos mapas antiguos para que se viera cómo ha cambiado y, además, lo que es mi parte favorita del proyecto, que se viera todo el proceso intermedio, lo que ocurre entre los mapas de 1922, antes de los diques, del petróleo y de comenzar a perforar y hoy en día. LA USGS tiene un gran conjunto de datos algo difícil de entender, ya que parece confeti tirado en una hoja, pero conseguimos extraer los períodos de tiempo y convertirlos en una animación, para ver cómo ha cambiado la tierra a través del tiempo.

Y si se visita la aplicación se pueden ver todos estos factores, los mapas originales, los actuales, y todos los factores que contribuyeron a la pérdida de suelo (los diques, el sistema de tuberías, los canales...), y también qué pasaría si no hiciéramos nada para intentar resolver el problema antes de 2050. Una de mis cosas favoritas del proyecto es que también se puede escuchar a la gente afectada, oír sus voces, y que expliquen cómo esto les ha afectado personalmente. No sé si tenemos tiempo de escuchar a este señor, pero habla de cómo, cuando era pequeño, podía navegar por hasta 6 millas de arroyos, cenagales y naturaleza para llegar al Golfo y ahora puede ver el Golfo desde la puerta de su casa. Así que es muy sorprendente, y puedes avanzar por esta línea de tiempo para ver el cambio en cada período, desde los años 30 hasta ahora, así que este es el proyecto.

Hicimos un seguimiento sobre algunas de las cosas que el estado está intentando hacer para arreglarlo, pero no creo que tenga tiempo de hablar de ello ahora, quizás después. Bueno, esto es un ejemplo de a lo que me dedico. Si alguien tiene preguntas...

Karma:

Le he pedido a AI que vaya un poco acelerado porque quiero que vosotros le podáis hacer ahora preguntas, dejamos unos 10 minutos o 15 para preguntas, y, que nadie se apure, tenemos aquí a Concha Catalán, que hará, si queréis hacer las preguntas en

castellano, le va a hacer la traducción simultánea aquí al momento, bueno, ¿cómo se llama esto de la traducción...?

Concha:
Interpretación.

Karma:
No, no, traducción simultánea o... consecutiva, eso. Y si no, le podéis preguntar a Al directamente en inglés, directamente. Si tú quieres empezar haciendo preguntas...

Concha:
Yo tenía una pregunta: Le quería pedir que pusiera las primeras diapositivas otra vez y nos hablara de los problemas que se encuentra limpiando los datos y todo esto. ¿Podrías hablarnos de, creo, la tercera o cuarta diapositiva, en la que hablabas de limpiar los datos? Dijiste que a veces no hay una noticia, ¿podrías explicarnos sobre esto un poco?

Al Shaw:
Claro. Es distinto con cada noticia, pero cuando conseguimos datos lo hacemos desde distintas fuentes, a veces vienen de páginas web con datos de acceso libre que nos permiten descargarlo directamente y usarlo, lo que está muy bien, y otras veces están en formato PDF, o en otros formatos, y tenemos que limpiarlos. Esto también puede variar mucho: a veces solo tenemos que añadir la latitud y longitud a una dirección, otras hay campos sin rellenar o nulos y no sabemos por qué, o tenemos que diferenciar entre cero, nulo y vacío, que son cosas distintas.

Concha:
Le pido que vaya a la diapositiva porque había muchos puntos y ha ido muy rápido. ¿Podrías compartir esas diapositivas con nosotros? Las publicaremos en algún sitio.

Al Shaw:
Sí. Iba un poco rápido, pero conseguir y limpiar los datos son dos partes de un todo.

Concha:
¿Podrías hablar de estos apartados? Por ejemplo, ¿Límites de Excel...?

Al Shaw:
Bueno, pues viendo la lista: Hay que asegurarse del número de filas que tienes en tu grupo de datos es el que se espera que haya, y no la mitad ni el doble. Si la agencia te dice que te van a dar 5.000 filas y luego tienes 3.000 hay un problema. En cuanto a los límites de Excel, hay un número mágico que no recuerdo, 65.500 y algo, pero existe, y si tu grupo de datos es exactamente ese número, sabes que está mal. Porque sabes que es el Excel el que lo está cortando y no los datos, por lo que tienes que saber cuál es ese número y si es exactamente ese, está mal.

Máximos y mínimos absurdos, es de lo que estaba hablando en el grupo de datos de los centros escolares, si un centro solo puede ofrecer 30 clases de nivel avanzado pero dice que ofrecen mil, es imposible, por lo que buscas esos espacios vacíos en oposición a nulos, a veces un campo vacío es diferente a un valor que intencionadamente debería ser nulo, a veces la agencia dice “esto es un valor nulo”, lo que no es un valor en blanco, que puede no haberse rellenado de forma incorrecta, así que hay que buscarlos.

Concha:

¿A cuántas escuelas tuvisteis que llamar? Porque ha dicho que había 60.000.

Al Shaw:

Llamamos a cientos de distritos. Este fue un proyecto muy largo con mucho tiempo de limpieza. Después, errores de ortografía: si tienes dos médicos, por ejemplo, en tu base de datos, y el nombre de uno se ha escrito mal, puede que sean la misma persona, y si ha sido por un error de ortografía, eso es un problema.

Concha:

¿Cómo limpiáis eso?

Al Shaw:

Es difícil, la mayoría de esto trata de comprobaciones, hay que hacer revisiones aleatorias para ver si se encuentran este tipo de errores y hay que estar atento. Los tipos de datos son un problema mayor; por ejemplo, si necesitas algo con decimales, pero te lo dan con un número sin decimales porque el software los quitaba, esto puede suponer un problema. A veces consigues una *string*, que son caracteres en lugar de decimales, por lo que es obligatorio asegurarse de que los tipos de datos están bien. Te dan un diccionario de datos, o una plantilla de registro donde aparecen todos los datos y el tipo que tiene que tener cada uno, y si no coinciden tienes que decirles: “Dijisteis que este iba a ser un decimal, pero me estais dando una *string*”.

Algo habitual en EEUU ocurre con los códigos postales, que a veces pueden empezar por cero, pero si se cortan y se dejan enteros, se cae el cero inicial y obtienes un código postal acortado que comienza en el segundo número, lo que pasa a menudo. Así que si ves que es un número diferente, sabes que alguien lo convirtió a un número entero en vez de con una *string*, que preservaría ese cero inicial. La geocodificación es otro problema...

Concha:

Ah, creo que ya has hablado sobre este. Vale, hay más preguntas...

Al Shaw:

Aquí hay una guía que hizo un editor que solía trabajar con nosotros y que incluye todo esto, una guía para dejar los datos a prueba de balas, recomendando que se lea.

Concha:

Excelente, así que indicáis como hacerlo.

Al Shaw:

Sí, y todos los puntos que yo he mencionado están en esta guía.

Concha:

Vale, pues nos aseguraremos de poner una copia de esta diapositiva en algún sitio. Ahora una pregunta de Jesús Escudero, de El Confidencial.

Jesús Escudero:

Hola, ¿cómo empezáis una investigación en ProPublica, tenéis primero la pregunta y luego creáis la base de datos para responderla, o tenéis una base de datos y respondéis la pregunta? Creo que dependerá de cada investigación, pero, ¿cuál es la forma más habitual?

Al Shaw:

Pasa de las dos formas. Muchas veces tenemos unos datos, sabemos que van a salir, o los hemos pedido mucho antes y tenemos algunas hipótesis sobre lo que podríamos encontrarnos. No sabíamos lo que obtendríamos con los datos de los centros educativos, pero los conseguimos de forma exclusiva, teníamos una fuente en el Departamento de Educación que nos dijo que nos lo iban a dar a nosotros antes de hacerlo público, así que pensamos que quizás podríamos adelantarnos para ver qué podía tener de interesante. Nos arriesgamos un poco, porque al empezar no sabíamos que iba a haber una noticia ahí, desconocíamos si encontraríamos el tipo de desigualdad que finalmente encontramos. Es decir, muchas veces empezamos con los datos y tenemos la esperanza de que haya una noticia en ellos, otras veces nos han informado de algo y tenemos documentos, por ejemplo, que prueban algo, de la forma tradicional, cuando sabes que algo ha pasado y buscas los datos y documentos que lo secunden. Así que ocurre de las dos formas.

Concha:

Jesús está haciendo muchas peticiones de libertad de información porque tenemos problemas para obtener los datos en España. Yo he estado tuiteando sobre conseguir datos escolares en Cataluña porque envié una petición al Gobierno para decirnos cuántos nuevos puestos en las escuelas contaban con becas parciales y cuántos totales, y respondieron que no lo sabían y que debía rellenar un formulario para convertir 10.000 en cifras de dos dígitos. Dijeron que no tenían esa información, es totalmente diferente aquí, ¿verdad, Jesús?

Señora:

En cuanto a estos datos, ¿de dónde los sacáis? ¿Siempre son públicos, tenéis acceso preferente...?

Al Shaw:

Depende, a veces es público, como la tabla sobre las elecciones, con las campañas y los grupos independientes. A veces alguien nos lo da, otras lo pedimos y somos los

únicos que lo tienen porque fuimos los únicos que los han pedido. A veces lleva años que nos los den pero, cuando llegan, somos los únicos que los tienen. Otras veces creamos nuestros propios datos, tenemos algunos documentos y, a mano, los convertimos en datos. Esto lo hemos hecho mucho con registros de impuestos, que solo están disponibles como formularios, no como datos. Con grupos sin ánimo de lucro hemos tenido que coger los datos, ponerlos en una hoja de cálculo y crear el grupo de datos a mano. Podría decirse que los datos sobre los impuestos son de libre acceso pero que no son realmente datos, son formularios escritos, así que nadie hace esa noticia por resultar tan difícil obtener los datos desde los documentos, por lo que varía mucho, a veces son datos libres, otras veces los rebuscamos por páginas web, como nuestro proyecto *Dollars for Docs*, que trata de los pagos de las farmacéuticas a los médicos por conferencias, comidas, etc. Esos datos los obtuvimos buscando las páginas webs de las empresas, porque están obligadas a decir cuánto pagan a los médicos, pero no a ponerlo en un formato claro, así que acabamos revisando todo e intentando unirlos para que se viera cuánto se estaba llevando un médico determinado de cada empresa. Cada empresa tenía los datos en un formato distinto e intentamos combinarlos, por lo que creamos un grupo de datos a base de rebuscar y unirlos, no eran datos de acceso público.

Concha:

En este país esa información es alto secreto. Ni siquiera podríamos soñar con saber cuánto pagan las farmacéuticas a los médicos, con nombres y apellidos.

Al Shaw:

La verdad es que recientemente han cambiado muchas cosas que ahora hacen que sea más fácil obtener esa información en EEUU. Antes tenían que revelarlo en la forma que quisieran, pero ahora el gobierno está creando una base de datos para unirlos todo y que la gente pueda acceder, por lo que están mejorándolo en EEUU.

Señor:

¿Hacéis alguna estadística sobre el uso de estas aplicaciones? ¿La gente está buscando su propia noticia con ellas, o lo usan para ver una de las cosas y decir que es una buena noticia, por ejemplo: “Ah, qué app más buena, y tras dos clics ya está”?

Al Shaw:

De hecho, nuestras aplicaciones de noticias son mucho más populares que nuestras noticias. La de *Dollars for Docs* tiene miles de visitas al día. Publicamos noticias que quizás un día tienen muchas visitas, pero las aplicaciones viven mucho más tiempo, son lo más popular de toda nuestra página web. Nos gusta hacerlas, tanto porque ayudan a la gente a encontrarse a sí misma en la noticia como porque son resistentes, se convierten en recursos para algo posterior de un modo que una noticia no puede hacer. Son muy, muy populares.

Concha:

¿Podéis identificar si mayoritariamente se trata de gente escribiendo el nombre de su médico, o...?

Al Shaw:

Sí, podemos.

Concha:

¿Y es ese el usuario más habitual?

Al Shaw:

Sí.

Concha:

Eso es lo que haríamos todos.

Señora:

Yo tengo dos preguntas, la primera es: después del proyecto, ¿qué acciones tomó el gobierno? Y además, ¿podrías hablar un poco más del modelo de negocio que es ProPublica?

Al Shaw:

¿Qué proyecto, específicamente?

Señora:

El último del que has hablado.

Al Shaw:

¿El de Luisiana? No ha habido ninguna acción aún, es una pena. El gobierno tiene un plan de 50.000 millones de dólares para intentar arreglar esto que se supone que duraría 50 años, escribimos algo sobre ello en el seguimiento que hicimos, pero apenas hay dinero en Luisiana, que es un estado muy conservador, por lo que no hay mucho apoyo para causas medioambientales. Supongo que hizo que la gente conociera el problema, pero no creó un cambio, por desgracia. No todo lo que hacemos lo consigue; lo intentamos, pero no ha pasado en este caso, al menos por ahora, quizás más adelante. En cuanto al modelo de negocio, somos una organización sin ánimo de lucro, así que nos financia principalmente la filantropía, sobre todo fundaciones y algún donante individual. Todo comenzó con un filántropo principal que donó una gran cantidad para que pudiéramos comenzar en 2008 y que se ha ido retirando y ya otras fundaciones han aportado más, así que es una mezcla de fundaciones... [le preguntan algo inaudible] Algo de dinero, pero no lo suficiente para mantenernos a flote, es una cuenta pequeña.

Karma:

Gracias por tu presentación. Vais a tener oportunidad de hacer un taller con Al Shaw mañana por la mañana en Blanquerna. Al, brevemente, ¿qué vas a hacer en el taller, nos puedes explicar?

Al Shaw:

El taller va a tratar principalmente sobre cómo hacer mapas como el que habéis visto en la imagen satélite de Luisiana, cómo coger los datos del gobierno, porque en EEUU el gobierno lo da, lo puede descargar cualquiera; cómo crear una cuenta para conseguir esos datos y cómo procesarlos. Hablaremos de cómo hacer eso, de qué son los datos satélite y cómo entenderlos, porque hay bandas distintas, las RGB y las no visibles, cómo acceder a esos datos, buscarlos, descargarlos y procesarlos, usando tanto herramientas de acceso libre como Photoshop, así que es interesante, deberíais venir.

Karma:

Es un apasionado, como veis, de las imágenes satélite, así que mañana vamos a tener una hora por la mañana para que nos explique todo esto.

(Fin de transcripción)

Periodisme de Dades. Explicant històries amb *Software*. Conferència amb Al Shaw.

Karma:

Benvinguts i benvingudes a les terceres jornades de periodisme de dades que celebrem. Aquestes terceres jornades són especials, perquè per primera vegada estem ja a quatre ciutats. Això vol dir que, en aquest moment, a Madrid, aquí a Barcelona, a Almeria i a Iruña, Pamplona, hi ha gent interessada en el periodisme de dades que està assistint a un programa en directe, com nosaltres. Jo volia començar ensenyant aquesta imatge, que, no sé, molts de vosaltres la coneixereu. És la imatge de Florence Nightingale, que ja fa més de 100 anys va descriure com la guerra de Crimea, va demostrar com a la guerra de Crimea hi havia molts soldats que no només morien en el camp de batalla, sinó també per les malalties contagioses que agafaven als batallons i a les infermeries, no? Ella era infermera i va explicar d'aquesta manera aquestes dades, no? I volia començar amb aquest gràfic així explicat molt de passada, perquè el periodisme de dades no és una cosa nova, sinó que ja, com ens va demostrar Florence Nightingale, però també Minar i altres estadistes de fa més de 100 anys, es podien fer infografies i representacions amb les dades. El que sí que és nou és la manera amb què avui estem treballant les dades des del periodisme, la manera amb què accedim a les bases de dades, la manera amb què accedim a aquestes eines digitals de visualització, la manera amb què mitjans com ProPublica a través d'imatges per satèl·lit ens mostren i ens expliquen què està passant en una zona dels Estats Units, etc. Aquesta aventura que vam començar durant les jornades del 2013 la Mar Cabra i jo, la ministra d'investigació de referència que tenim a Espanya. Em fa gràcia després de 3 anys mirar enrere perquè el 2013 pràcticament no es parlava, ni se sentia a parlar, de dades, no teníem llei de transparència a Espanya, no la teníem a Catalunya ni en altres comunitats autònomes.

Configurar el programa d'aquell any, que no fa tant, només dos anys enrere, va ser un autèntic repte per les dues, perquè convèncer els periodistes perquè vinguessin a unes sessions era una tasca titànica; intentar trobar algun exemple d'un mitjà de comunicació que estigués treballant amb dades encara era més difícil. Aquest any el més difícil ha estat tot el contrari: crear una graella, un programa que donés cabuda a tantíssima gent que ja està treballant amb dades aquí a Espanya i que ens permetés portar a gent d'altres països perquè ens donessin la seva experiència. Amb l'Adolfo Antón, que ha agafat el relleu de la Mar Cabra a Madrid, parlàvem. Jo li deia: "Adolfo, això és com fer un trencaclosques, anar encaixant peces perquè tots els ponents i experts puguin ser a dins."

Bé, si veieu el programa de mà que teniu, aquest any tenim molts més logos que ens acompanyen, el nostre agraïment a totes les institucions, centres, organitzacions i empreses que ens han donat suport perquè tinguem un programa a quatre ciutats amb

gairebé una setmana d'activitats. El nostre especial agraïment al Centre de Cultura Contemporània, que ens acull per tercer any, i al magnífic equip d'en Juan Isúa amb la Maria com a coordinadora, que fa possible que tot això vagi endavant per descomptat. A la Facultat de comunicació Blanquerna, on serem demà, el dissabte i el diumenge fent tallers, i també al Col·legi de periodistes que aquest any ha estat la seu, està sent la seu de la hackatò que s'està celebrant en aquest moment i ahir vam estrenar aquestes jornades al Caixa Fòrum. Són unes jornades, com veureu al programa, amb molts experts, gent que ha decidit venir desinteressadament a aquestes sessions. A tots ells, el nostre especial agraïment perquè els vam demanar que vinguessin de Madrid a Barcelona en un temps rècord. L'AVE està funcionant, espero que molt bé. Tenim ponents que han arribat ara mateix després d'agafar un tren a les 7 del matí i durant el dia aniran arribant ponents que en aquest moment són a Madrid. L'Al Shaw i l'Elisabetta Tola ja han patit això de Madrid-Barcelona, perquè estrenaven abans-d'ahir a Madrid. A tots ells, moltíssimes gràcies, per descomptat. Bé, aquest és un projecte, com sabeu, organitzat per l'Open Knowledge Foundation, una organització internacional que es troba en més de 30 països. Els que formem part d'aquesta organització, aquesta és l'OKFN Spain, el que pretenem és organitzar sessions per donar a conèixer això, precisament, el coneixement obert, i organitzar més sessions perquè més gent aprengui, més gent sàpiga sobre un tema concret. I jo, com a membre interessada en el periodisme de dades, soc la que sempre embolico totes les sessions, tallers, hackatons sobre dades. Voldria que, si agafeu el programa de mà, us comentaré una mica qüestions pràctiques i repassaré el que ens ve a sobre, perquè això és una maratò, com altres anys. Com us he dit, en aquest moment s'està realitzant la hackatò, el *hack day* GEN. A Barcelona tenim 2 equips, *La Vanguardia* i el diari *Ara*. A Madrid hi ha 5 equips. I tots 7 equips de periodistes, programadors, dissenyadors, estan competint per anar a la final, que la GEN celebrarà el dia 17 aquí, també, a Barcelona. Una final amb tots els equips finalistes de molts països del món. I la veritat és que és un repte poder celebrar aquesta *hack day* dins de les jornades. Ahir vam tenir, com us he comentat, el debat amb l'Elisabetta i amb l'Al Shaw a Caixa Fòrum, i avui és el dia perquè gaudiu de conferències, xerrades, tindrem un *Hacks and Hackers* amb cerveses al final del dia. Això ho dic ara perquè ningú se'n vagi d'aquí fins a la tarda-nit... Tenim un desdoblament de programa a la tarda, dedicat a l'*Open Data*, amb experts i professionals que ja tenen empreses i estan treballant amb les dades obertes i, bé, hi ha *streaming*, en aquest moment s'està gravant perquè tots aquells que no poden ser aquí puguin seguir les xerrades per *streaming*. I demà serà un dia de tallers al màxim. El programa gairebé no cap en aquest tríptic de mà amb experts que avui escoltareu i que demà també hi seran fent tallers, a més de molta altra gent que vindrà especialment per aquestes sessions d'ordinador. Tots heu de venir amb ordinador als tallers, això sí, que ningú vingui sense el seu ordinador. Bé, jo no sé si us atrevireu a arribar fins al final i ens acompanyareu fins al diumenge a les 2 de la tarda, quan s'acabarà aquesta maratò de jornades, espero que sí que sigueu molts. Si teniu dubtes sobre els tallers, el contingut dels tallers, o els ponents que hi participaran, ho teniu a la pàgina sobre periodisme de dades de les jornades. Dins de la secció "ponents" hi ha tots els detalls sobre tots ells. Abans he passat unes etiquetes, espero que vulgueu jugar i us acabeu relacionant entre vosaltres i que conegueu almenys a 5 persones durant el dia d'avui i compartiu els perfils i a què

us dediqueu. Bé, l'únic que em queda per dir és que aquest esdeveniment està organitzat amb la millor de les intencions, amb la idea de compartir el coneixement que tenim entre tots, de fer-lo accessible a quanta més gent millor, perquè és l'única manera d'anar avançant, mentre més sapiguem entre tots. Moltes gràcies, espero que gaudiu de l'esdeveniment. I ara us deixaré amb l'Al Shaw, a qui passo a presentar. Ell ve de ProPublica, un mitjà independent dels Estats Units, un dels mitjans de referència que hi ha en aquest moment. Els treballs que fan no són treballs que es fan en un dia ni en un mes, ni en 5 minuts, són treballs d'un any i d'un any i mig, a vegades 3 anys. Després tindreu l'oportunitat de parlar amb ell i que us ho expliqui. Ell té un perfil professional que ens agradaria tenir a molts de nosaltres, perquè és el que en diem híbrid: té perfil de periodista, però també de programador, però també de dissenyador. Dins de l'equip, ell s'encarrega de fer el que en diuen les *News Applications*, les aplicacions per fer notícies, i últimament ha treballat molt amb imatges per satèl·lit que ara ens explicarà, sobretot en el seu últim, un dels últims reportatges en què ha treballat, *Losing ground*. Jo ja no vull parlar més. Al, és el teu torn. Gràcies per ser aquí.

Al Shaw:

Hola? Sí, hola? Gràcies, Karma, m'alegro de ser aquí. Parlaré en anglès. Si no m'enteneu, puc parlar una mica més a poc a poc. Aixequen la mà i pregunteu, ho intentaré repetir. Soc l'Al, treballo a ProPublica, que és una organització de notícies d'investigació sense ànim de lucre de Nova York. Vam començar el 2008 i ens agrada dir que fem periodisme per l'interès públic. No cobrim exclusives, només notícies centrades en la responsabilitat, amb força moral, que poden canviar alguna cosa en el món. Totes les històries que publiquem tenen la intenció de causar un impacte. Així és com recaptem diners, és en el que els diem als donants que estan invertint. Aquesta és el tipus de feina que fem a ProPublica. Una de les maneres amb què ho fem, la que més, és amb dades. Treballo en un equip que en diem d'aplicacionistes de notícies, podeu trobar els nostres treballs a la pestanya Eines i Dades de la nostra web. Utilitzem dades perquè els governs i les empreses siguin responsables d'una manera que no es podria aconseguir només citant fonts o amb notícies tradicionals. Expliquem històries, encara que això és una mica genèric, perquè tot són històries, però nosaltres les expliquem amb text, gràfics, que són petites aplicacions d'una pàgina amb una mica de dades, i amb les Aplicacions de Notícies, que són grans i interactives bases de dades on es pot buscar la teva història entre les dades. En el nostre equip també escrivim text, fem fotos, però ens agrada fer gràfics i també aquest tipus d'aplicacions, de què us parlaré de seguida.

El procés per crear una aplicació de notícies, o inclús per fer qualsevol cosa amb les dades, és bastant llarg, intens i té molts passos. El primer és aconseguir les dades, que no és fàcil. Una manera és a través del sistema de llibertat d'informació que tenim als Estats Units, que ens permet obtenir informació de funcionaris públics. També hi ha moltes dades de lliure accés, però això pot variar, des de ser súper senzill, com entrar a la web del govern i descarregar-ho, fins a fer una petició oficial, esperar un any, comprovar que les dades estan en el format adequat, cosa que moltes vegades no és

així i s'ha d'intentar canviar-lo al correcte, i després continuar amb el següent pas, que és netejar-les.

Això és una cosa molt complicada i també pot anar des de canviar algunes columnes fins a variar el format, de PDF a Excel, unir molts tipus de dades, és una tasca difícil. El següent pas és analitzar les dades. Només es pot començar a analitzar dades un cop s'han netejat. És aquí quan s'intenta trobar la notícia dins de les dades, que a vegades pot no haver-n'hi cap. Quan has arribat a la fase d'anàlisi a vegades és massa tard, has arribat massa lluny obtenint i netejant les dades fins que ho presentes, que és el que anomenem aplicació de notícies.

Perdó. Les aplicacions de notícies són bases de dades grans i interactives on pots trobar la teva pròpia història dins d'un set de dades més gran. Us ensenyaré un exemple d'una aplicació que fem fet que es diu *The Opportunity Gap*. Aquesta app tracta la desigualtat en l'educació als Estats Units, si els diferents estats proveeixen d'un accés igualitari a l'educació i a l'oportunitat educativa. Es basa en un grup de dades que vam obtenir del Departament d'Educació dels Estats Units. Venien amb molts centres i districtes diferents i el problema amb aquestes dades va ser que els districtes donaven la seva pròpia informació, així que les dades provenien del Ministeri d'Educació, però cada districte les havia presentat davant el Govern Federal, per la qual cosa havies de confiar en què les dades de cada districte fossin correctes. Vam comprovar que no sempre era així; alguns districtes deien, per exemple, que tenien mil classes de nivell avançat, quan en realitat el màxim que poden tenir era de 30. En aquests exemples havíem de trucar al districte en qüestió i consultar-los per què deien això si el màxim eren 30 classes d'aquest tipus. Ho vam haver de fer amb 60.000... No vam trucar als 60.000 centres, però vam haver de buscar el que en diem 'l'absurd', que implica, a l'hora de netejar les dades, localitzar què està malament i com podem resoldre-ho.

Per trobar 'l'absurd' fem diverses coses diferents, com per exemple: buscar comptes i totals, mala geocodificació, com que no incloguin adreces o que sí les incloguin però sense el codi postal o el país... Hi ha moltes estratègies que adoptem per netejar les dades abans de començar a analitzar-les. Es pot veure la metodologia que utilitzem en aquest projecte i llegir aquestes guies sobre com assegurar les dades que va escriure un dels nostres antics editors.

Després de netejar les dades arriba la meua part preferida del procés, que és crear aquestes aplicacions de dades, aquestes bases de dades interactives amb una notícia dins d'un grup de dades més gran. Diem que fem les notícies amb *software* en comptes de fer-les amb paraules i fotos. Aquests grups de dades interactives, les aplicacions de notícies, tenen molts components, però els principals són la vista llunyana i la vista propera. Igual que quan estàs llegint una notícia normal, al principi hi ha un titular i un paràgraf, que et fa el que en diem el resum, i després ve un altre paràgraf que t'explica el tema principal, la informació important de la notícia, i això funciona igual en les aplicacions. La vista llunyana et dona primer el títol gran que explica què està passant i després el paràgraf de resum. En aquesta aplicació, *The Opportunity Gap*, pots buscar

per tots els estats del país i veure quines oportunitats educatives ofereixen sense importar el nivell d'ingressos. Es pot veure que en aquest exemple el primer estat és Alabama, els millors oferint accés a classes avançades; que el segon estat és Alaska, i pots seguir avançant i consultar el teu estat, en aquest macro nivell, per comprovar que cada estat és diferent de l'anterior a l'hora d'oferir l'accés a aquestes classes. El següent que pots fer és entrar a la vista propera. Quan llegeixes una notícia tradicional i avances per la pàgina pot ser que arribis a una anècdota: algú a qui els periodistes van trobar que il·lustra molt bé la notícia. Veus cites d'aquesta font, és el protagonista de la notícia i et guia per aquesta novetat nacional. En les aplicacions de notícies tu ets aquesta anècdota protagonista, et permetem buscar-te, trobar-t'hi. Quan busques el teu centre o la teva ciutat veus dades sobre això i un paràgraf amb informació. Treballem amb una empresa que es diu *Narrative Science* que va escriure 60.000 històries, una sobre cada centre del país, i pots llegir-les, n'hi ha 60.000. Pots pensar que aquesta aplicació de notícies té, per tant, 60.000 notícies diferents, però, de fet, en té més, perquè pots anar més endins i comparar aquest centre amb altres de propers, o amb centres on hi ha més o menys pobresa. Pots pensar: "Bé, jo vaig a aquest centre, és intermediari." El 38% d'alumnes té el preu del dinar reduït, que és la nostra mesura per la pobresa, per la qual cosa el 38% es beneficia d'aquest descompte o gratuïtat, però el 74% aprova les classes de nivell avançat, per la qual cosa van força bé. Per altra banda, en aquest centre de l'esquerra el 89% dels alumnes té descompte o gratuïtat en el dinar, per la qual cosa és força pobre, i el 8% aprova les classes de nivell avançat. I per altra banda hi ha un centre molt ric on només l'1% té descomptes o gratuïtats pel menjar i on el 87% dels alumnes aproven les classes de nivell avançat. Així es pot veure on és el teu centre en perspectiva.

També hi ha mapes, aquest és un dels que m'agrada força, perquè mostra dues vistes oposades del sud de Califòrnia. Podeu veure que les parts que són vermelles en un mapa són molt més clares en l'altre, perquè les parts vermelles d'un mapa mostren pobresa més alta i les blanques de l'altre mostren el número d'alumnes inscrits en classes de nivell avançat. Es pot veure que Los Angeles és un lloc on l'oportunitat educativa no és gaire alta, si vius en un districte ric probablement accediràs a una millor educació que si vius en un districte pobre, que és el que aquesta aplicació de notícies intenta mostrar.

Tot i això, nosaltres anem una mica més enllà. Ens agrada dir que les aplicacions de notícies generen notícies, perquè els periodistes locals es poden descarregar aquestes dades i redactar les seves pròpies notícies basant-se en la informació que n'extreuen. Poden descarregar-se l'arxiu CSV de manera gratuïta i publicar-lo de nou, poden utilitzar-lo com a base per les seves pròpies notícies o poden usar la mateixa aplicació per crear una història. Això és el que fan molts periodistes, ja que la majoria no saben utilitzar dades o treballar-les. Per això vam crear l'aplicació, perquè els reporters poguessin buscar centres propers a la seva àrea de treball, trobar la notícia i redactar-la pels seus mitjans locals, cosa que molts van fer. Aquestes organitzacions i estacions de televisió locals van utilitzar la nostra app i van redactar-ne notícies locals. És una de les maneres a través de les quals es van conèixer aquestes dades, no només per gent

que entrés a la nostra app, sinó també per les persones que van llegir les notícies en els seus mitjans locals. També tenim una botiga de dades, on la gent pot comprar i inclús descarregar dades de manera gratuïta. Si passem molt de temps intentant netejar-les i treballar amb les dades, cobrem una mica per les dades, perquè ens ha portat molt de temps passar-les al format correcte, però si simplement vam fer la petició al Govern i ens les han facilitat, les oferim gratuïtament, perquè no tindria sentit fer que tothom passi pel mateix procés si no cal que ho facin.

Per això, les aplicacions de notícies generen notícies, però a vegades hi ha un tercer pas més enllà en què una notícia genera una aplicació que a la vegada genera noves notícies. Un exemple d'això, que a mi personalment m'agrada molt, és aquesta notícia que vam veure al *New York Times*. Ens va agradar i vam pensar que potser hi hauria una manera de programar-ne una aplicació. És una notícia sobre supergrups, que són independents i que no estan vinculats a campanyes polítiques, o que no haurien d'estar-ho, i que precisament perquè són independents poden gastar tots els diners que vulguin en campanyes. Aquesta notícia del *New York Times* va demostrar que moltes d'aquestes consultories que els supergrups estaven usant també estaven contractades per campanyes oficials, per molt que suposadament no haurien d'estar treballant juntes, i a vegades treballaven fins i tot en el mateix edifici.

Pensem que era interessant i que potser podríem esbrinar més coses analitzant les dades de la comissió federal, per veure qui estava contractant a quines consultories, ja que aquests supergrups i aquestes campanyes han d'informar a la comissió federal de les eleccions. Llavors, vam desenvolupar aquesta llista interactiva que permet veure les diferents campanyes, supergrups i grups independents en un costat i els llocs contractats, les consultories, a l'altre. Amb això vam poder publicar, per la qual cosa acabàvem tot el cicle: vam llegir una notícia, vam crear una aplicació basant-nos en el seu esquema i vam poder publicar una altra notícia nosaltres amb les dades obtingudes sobre una empresa anomenada *Targeted Victory* que vam localitzar utilitzant la nostra aplicació de notícies i que havia estat contractada per la campanya de Mitt Romney i per un grup independent que li donava suport, però que en teoria no podia treballar a la campanya. Vam descobrir molts d'aquests grups i això és el que vam publicar.

Aquesta és una altra manera d'utilitzar dades. Una cosa en què probablement no esteu pensant és a fer servir les dades per trobar a altres persones. Bé, posem les dades a la web, la gent la visita i es troba en la notícia, però i si volem escriure les dades de manera que puguem trobar aquesta anècdota sobre la qual volem publicar la notícia? També podem usar les dades. Fa un parell d'anys vaig estar treballant en una notícia sobre l'huracà Sandy, una gran tempesta que hi va haver a Nova York i que va causar molts danys. Gran part dels danys van tenir lloc perquè els mapes d'inundacions que estaven en vigor quan va arribar la tempesta no eren prou bons, no eren precisos i no s'havien actualitzat. Les dades en què es basaven aquests mapes eren del 1983, però els mapes en si s'havien actualitzat el 2007, sense que gaire gent ho sabés, amb les dades de 1983. Van canviar algunes coses, però en general es tractava de les dades de 1983 encara que els mapes indicaven 2007. Vam creure que molta gent hauria pensat que

els mapes eren actuals quan van comprar les seves cases, ja que no sabien que les dades que hi havia darrere d'aquells mapes eren tan antigues. Per això, vam decidir buscar persones que havien comprat casa seva després del 2007 pensant que estaven fora de perill, però les llars de les quals van resultar malmeses per l'huracà. Després de l'huracà, el Govern va publicar noves àrees d'inundació, que són aquestes zones de color marró, que prediuen de manera molt més precisa on colpejaria la tempesta, fent servir dades noves. Per la qual cosa vam pensar: "Qui hauria estat fora de perill si el Govern hagués utilitzat dades bones el 2007 en lloc de reutilitzar les de 1983?" Vam escriure un algoritme, aquest n'és el pseudocodi, per trobar gent afectada, gent que vivia en la nova zona d'inundació, no en l'antiga, i que hagués construït o reformat casa seva després de 2007.

Vam trobar a aquesta parella que vivia en un barri anomenat Sheepshead Bay i eren l'exemple perfecte: sent molt previnguts, havien instal·lat una bomba d'aigua petita perquè van pensar que potser els arribaria una mica d'aigua al soterrani, el que deien els mapes de 2007, però el seu soterrani es va inundar totalment, cosa que, lògicament, no els va agradar. Per tot això, es van convertir en els personatges de la nostra notícia a través d'aquesta anàlisi de dades. Vam fer diversos mapes per aquesta notícia, a l'esquerra hi ha el barri on vivien, Sheepshead Bay, a Brooklyn, però això va passar per tota la ciutat; vam fer tres mapes que mostraven totes aquestes marques negres, els edificis que van patir danys durant l'huracà.

També vam decidir anar un pas més enllà i desenvolupar una aplicació de notícies, perquè vam pensar que si havia afectat a aquella parella et podria haver afectat a tu també. I a través del nostre anàlisi, vam esbrinar que els mapes de Nova York i Nova Jersey que estaven en vigor quan va arribar l'huracà eren diferents entre si a l'hora de predir millor o pitjor on colpejaria l'huracà, perquè s'havien actualitzat en èpoques diferents sense cap ordre lògic. Molts no s'havien actualitzat des de feia molt, alguns dels comtats ni tan sols tenien mapes digitals, són els grisos que apareixen aquí. Llavors, pots buscar el teu comtat i veure exactament què et passava si vivies en una àrea que els mapes de 2007 no predeïen, però ara, amb els nous, sí que et trobaves en una zona inundable. Bé, doncs aquest és un exemple de com fer servir les dades per trobar persones i, a més, desenvolupar una aplicació a partir d'això.

Ens agrada cedir el nostre codi a més de les nostres dades, tenim la botiga de dades, amb dades descarregables que provenen de les nostres aplicacions, però també ens agrada cedir el codi que fa funcionar les nostres aplicacions. Per exemple, pels mapes de què estava parlant, es pot usar el *software* que vam crear, que es diu *Simple Tiles*. També hem creat fonts pels estats dels Estats Units i tenim molt de codi diferent. Normalment, quan creem una aplicació que pot tenir un component de codi que podria ser valuós per més gent fora d'aquell únic ús, intentem extreure'l perquè més gent pugui reutilitzar-lo amb un altre propòsit en les seves pròpies aplicacions.

Aquests són alguns dels exemples de les coses que fem, però vull parlar més sobre un projecte que vaig fer l'any passat que pràcticament va consumir tot el meu any. Es diu

Losing Ground. És sobre com el sud-est de Louisiana, per una sèrie de factors principalment originats per l'home, una gran part de l'estat, aproximadament un terç, està desapareixent i s'està enfonsant a terra, i la majoria dels habitants del país ni tan sols ho sabien, si jutgem la resposta que va tenir la notícia. La gent estava impactada a pesar que els mitjans de Louisiana haguessin estat escrivint sobre el tema des de feia bastant de temps. El que em va fer investigar això, a més del fet que he estat viatjant per Louisiana, van ser dos documents diferents. Un, que va ser el que més em va sorprendre, va ser aquest, que va publicar una agència del govern nord-americà, crec que l'Administració Oceànica i Atmosfèrica Nacional. Aquest document mostra tota la terra que s'ha perdut des de 1932 i que es perdrà fins al 2050. És una gran quantitat de terra i, si això passa, la ciutat de Nova Orleans serà gairebé una illa en aquesta península, envoltada de quilòmetres i quilòmetres d'aigua, en lloc dels pantans que la protegien dels huracans. L'altre document, també sorprenent, va ser aquest mapa del sud-est de Louisiana que mostra tots els llocs que el govern ha tret dels mapes oficials per la pèrdua de terra. Són badies, ports, rierols, rius... que s'han eliminat dels mapes perquè aquesta terra s'ha perdut en els últims 100 anys.

Vaig començar a llegir molt, hi ha llibres molt bons sobre el tema. Un dels meus preferits es diu *Bayou Farewell* i tracta de la cultura cajun a Louisiana i sobre com la gent que hi viu s'està adaptant a aquesta pèrdua. Aquest llibre em va inspirar molt a treballar en això i a esbrinar per què estava passant, ja que saber-ho em permetria explicar la notícia i veure com la gent estava trampejant la situació. Bé, això és el que està passant. Durant els últims 7.000 anys, el riu Mississipí, que passa per Nova Orleans, ha canviat molt, cosa que ha creat aquesta forma de bota ara icònica de l'estat. Si algú coneix Louisiana, sabrà que és com una bota, i això es deu a què el riu ha canviat el seu curs durant milers d'anys i ha dipositat els seus sediments de tal manera que ha creat aquesta bota.

Bé, en els últims mil anys s'ha assentat en la posició actual, però a inicis del segle XX hi va haver una inundació massiva. El riu, que estava començant la seva següent fase en què aniria més enllà de Nova Orleans fins a l'oest, cap al riu Atchafalaya, es va inundar i no va arribar fins a Nova Orleans perquè havien col·locat uns dics sota la ciutat per alleugerir la pressió, però volien que el riu seguís passant per Nova Orleans, que era un port molt important. Si el riu hagués fet el que pretenia i s'hagués unit a l'Atchafalaya, Nova Orleans no hauria estat un port, cosa que no podien permetre. Llavors, el que van fer va ser crear un sistema de dics molt alts que mantenien el riu al seu lloc i l'obligaven a baixar pel seu recorregut tradicional, per on el riu no volia anar, per evitar les inundacions i permetre que els vaixells accedissin a la ciutat. També van establir una cosa anomenada "estructura de control del riu antic", que permetia que una mica d'aigua baixés per l'Atchafalaya i la resta pel Mississipí; van poder controlar la quantitat d'aigua que baixava per cada lloc. A conseqüència d'això, el sediment del riu, que durant milers d'anys s'havia dipositat creant aquesta forma de bota, ara no s'estava dipositant en els pantans, sinó en els dics i pel canal del Mississipí. Com que estava sent forçat a desviar-se i com que els dics arribaven fins ben endins de l'oceà, els sediments que creaven els pantans queien més enllà de la plataforma continental i es perdien per sempre. És a dir, que totes aquestes zones on la gent vivia, aquestes àrees

blanques, es van veure privades dels sediments que formaven la seva terra, que a més s'estaven perdent per sempre.

Aquest és un factor. El segon és que a inicis del segle XX trobem petroli al sud de Louisiana i comencem a construir canonades pels pantans. En aquell moment ningú pensava que els pantans fossin valuosos, pensaven que eren bàsicament llocs per als cocodrils i un ermot sense cap utilitat, així que molta gent es va enriquir venent les seves terres a petrolieres que, sense cap remordiment per la fauna o pels pantans, van començar a construir les canonades. Així és ara, aquestes canonades creuen tot el golf, n'hi ha moltíssimes. I la manera en què originalment es construïen era la manera més destructiva pel medi ambient: per posar la canonada creaven canals, que, com es pot veure a la imatge, eren molt més amples que la canonada en si. Per fer els canals van agafar terra del sòl i la van posar a les ribes per crear el que s'anomenen "ribes de rebuig". No van tornar a col·locar la terra al seu lloc, la van deixar allà, i com que els pantans són tan tous, la terra que es va col·locar a sobre va fer que s'enfonsessin, va fer pressió a la terra de fora dels canals. Al cap del temps van descobrir que hi havia una manera millor de fer tot això, que era només excavar el que necessitaven per posar la canonada i després tornar a col·locar la terra en acabar. Però llavors, potser uns vint anys després, era massa tard. Podeu veure què passa quan quilòmetres i quilòmetres de pantà han estat tractats d'aquesta manera, perquè la terra es va començar a enfonsar; el que originalment eren petits canals ara eren molt més grans i després es van convertir en estanys, llacs i, finalment, mar obert. Aquí es pot veure alguna infraestructura petroliera posada en aquest tros de terra que gairebé no existeix perquè els canals s'han convertit en mar obert. No hi ha dades oficials sobre els canals del sud de Louisiana, l'única manera de saber on són és per uns pocs estudis i petits grups de dades. Vam parlar amb un home que va fer la seva tesi sobre això i va fer a mà els canals d'aquest petit fragment de badia entre dos rius. En parlar amb ell ens vam adonar que aquest patró era similar a la resta del golf i regió de Nova Orleans i podíem extrapolar-lo. Podeu veure la densitat de canals que hi ha en aquesta zona.

Volíem explicar aquesta notícia i vam esbrinar que el Govern té molt bons recursos d'imatgeria aèria i satèl·lit, per la qual cosa aquestes són les meves fotos preferides de les que vam trobar disponibles gratuïtament a través de la USGS (l'enquesta geològica dels Estats Units). En diem "La roda del carro", és una zona molt al fons del riu Mississipí on es va trobar un dom salí sota el qual hi havia moltíssim petroli. Al voltant del dom van excavar un canal immens per començar a perforar pous al seu voltant. Al final, aquest canal es va fer gran i el 2013 es pot apreciar que és majoritàriament aigua que entra i surt d'on una vegada hi va haver un petit canal. El patró és sorprenent, quan es veu què ha passat amb el temps, es pot veure molt bé amb aquestes imatges aèries. Els pantans són tremendament importants, com deia abans. No només per la fauna, la pesca o per la seva bellesa natural, sinó perquè també serveixen com a defensa contra els huracans, així que per cada pantà que hi ha entre el mar obert i la ciutat, els huracans s'alenteixen, es debiliten una mica abans d'arribar al sistema de dics que envolta la ciutat, per la qual cosa això podria danyar-la. I, per això, una de les autoritats en el sistema de dics, les juntes locals que controlen els dics que envolten Nova Orleans, van denunciar 97

companyies petrolieres i de gasoil per danyar els pantans, i el cas encara està obert als jutjats.

Per poder redactar aquesta notícia vam confiar en imatges per satèl·lit perquè era la millor manera de fer-ho. Específicament amb el Landsat, un satèl·lit del Govern nord-americà que disposa de moltes bandes diferents: les tradicionals vermelles, verdes i blaves i, a més, infraroges, que recullen la llum que no és visible a l'ull humà. Parlaré sobre com treballar amb aquest tipus d'imatges al meu taller... Crec que m'estic quedant sense temps.

Vull ensenyar-los com vam fer el mapa que apareix en la primera pàgina de l'aplicació. No volíem fer servir aquesta imatge, que és la que es troba, per exemple, a Google Earth, una imatge tradicional RGB, perquè hi ha tants sediments a l'aigua del riu Mississipí que els sediments semblen aigua, sembla un piló de fang. Per això, la vam iterar diverses vegades, aquesta seria una segona iteració, i en utilitzar les diferents bandes vam ser capaços de mostrar la diferència entre terra i aigua clarament. Vam fer servir RGB i també una màscara per l'aigua amb una banda infraroja perquè es veiés aquesta diferència i també com s'està perdent la terra d'una manera que no es veuria amb Google Earth. Hi ha molts recursos que indiquen com processar imatges d'aquest tipus, us n'ensenyaré alguns al taller si decidiu venir.

Us ensenyaré ràpidament això: així es veu avui, però vam obtenir mapes antics per tal que es veiés com ha canviat i, a més, la que és la meva part preferida del projecte, que es veiés tot el procés intermedi, el que passa entre els mapes de 1922, abans dels dics, del petroli i de començar a perforar i avui en dia. La USGS té un gran conjunt de dades una mica difícil d'entendre, ja que sembla confeti tirat en un paper, però vam aconseguir extreure els períodes de temps i convertir-los en una animació per veure com ha canviat la terra a través del temps.

I si es visita l'aplicació es poden veure tots aquests factors, els mapes originals, els actuals i tots els factors que van contribuir a la pèrdua de sòl (els dics, el sistema de canonades, els canals...) i també què passaria si no féssim res per intentar resoldre el problema abans de 2050. Una de les meves coses preferides del projecte és que també es pot escoltar la gent afectada, sentir les seves veus, i que expliquin com això els ha afectat personalment. No sé si tenim temps d'escoltar a aquest senyor, però parla de com, quan era petit, podia navegar per fins a 6 milles de rierols, fangars i naturalesa abans d'arribar al golf i ara pot veure el golf des de la porta de casa seva. Així que és molt sorprenent, i pots avançar per aquesta línia de temps per veure el canvi en cada període, des dels anys 30 fins ara, així que aquest és el projecte.

Vam fer un seguiment sobre algunes de les coses que l'Estat està intentant fer per arreglar-ho, però no crec que tingui temps de parlar-ne ara, potser després. Bé, això és un exemple de la feina que faig. Si algú té preguntes...

Karma:

Li he demanat a l'AI que vagi una mica accelerat perquè vull que vosaltres li pugueu fer ara preguntes, deixem uns 10 minuts o 15 per a preguntes, i que ningú es preocupi, tenim aquí la Concha Catalán que farà, si voleu fer les preguntes en castellà, li farà la traducció simultània aquí al moment, bé, com es diu això de la traducció...?

Concha:

Interpretació.

Karma:

No, no, traducció simultània o... consecutiva, això. I si no, podeu preguntar a l'AI directament en anglès. Si tu vols començar fent preguntes...

Concha:

Jo tenia una pregunta. Li volia demanar que posés les primeres diapositives una altra vegada i que ens parlés dels problemes que es troba netejant les dades i tot això. Podries parlar-nos de, crec, la tercera o quarta diapositiva, en què parlaves de netejar les dades? Has dit que a vegades no hi ha una notícia, podries explicar-nos això una mica?

AI Shaw:

És clar. És diferent amb cada notícia, però quan aconseguim dades ho fem a partir de diferents fonts, a vegades venen de pàgines web amb dades d'accés lliure que ens permeten descarregar-les directament i fer-les servir, cosa que està molt bé, i altres vegades estan en format PDF o en altres formats i les hem de netejar. Això també pot variar molt: a vegades només hem d'afegir la latitud i la longitud a una direcció, altres hi ha camps sense omplir o nuls i no sabem per què, o hem de diferenciar entre zero, nul i buit, que són coses diferents.

Concha:

Li demano que vagi a la diapositiva perquè hi havia molts punts i ha anat molt ràpid. Podries compartir aquestes diapositives amb nosaltres? Les publicarem en algun lloc.

AI Shaw:

Sí. Anava una mica ràpid, però aconseguir i netejar les dades són dues parts d'un tot.

Concha:

Podries parlar d'aquests apartats? Per exemple, límits d'Excel...?

AI Shaw:

Bé, doncs veient la llista. Cal assegurar-se que el número de files que tens en el teu grup de dades és el que s'espera que hi hagi, i no la meitat ni el doble. Si l'agència et diu que et donaran 5.000 files i després en tens 3.000 hi ha un problema. Quant als límits d'Excel, hi ha un número màgic que no recordo, 65.500 i alguna cosa, però existeix, i si el teu grup de dades és exactament aquest número, saps que està

malament. Perquè saps que és l'Excel el que ho està tallant i no les dades, per la qual cosa has de saber quin és aquest número i si és exactament aquest, està malament.

Màxims i mínims absurds, és del que parlava en el grup de dades dels centres escolars. Si un centre només pot oferir 30 classes de nivell avançat però diu que n'ofereix mil, és impossible. Per la qual cosa busques aquests espais buits en oposició als nuls, a vegades un camp buit és diferent a un valor que intencionadament hauria de ser nul, a vegades l'agència diu "això és un valor nul", però és un valor en blanc, que pot haver-se omplert de manera incorrecta, així que s'han de buscar.

Concha:

A quantes escoles vas haver de trucar? Perquè ha dit que n'hi havia 60.000.

Al Shaw:

Vam trucar a centenars de districtes. Aquest va ser un projecte molt llarg amb molt de temps de neteja. Després, errors d'ortografia: si tens dos metges, per exemple, en la teva base de dades, i el nom d'un d'ells s'ha escrit malament, pot ser que siguin la mateixa persona, i ha estat un error d'ortografia. Això és un problema.

Concha:

Com ho netegeu això?

Al Shaw:

És difícil. La majoria d'això tracta de comprovacions. S'han de fer revisions aleatòries per veure si es troben aquest tipus d'errors i s'ha d'estar atent. Els tipus de dades són un problema més gran; per exemple, si necessites alguna cosa amb decimals, però te la donen amb un número sense decimals perquè el *software* els treia, això pot suposar un problema. A vegades aconsegueixes una *string*, que són caràcters en lloc de decimals, per la qual cosa és obligatori assegurar-se que els tipus de dades estan bé. Et donen un diccionari de dades o una plantilla de registre on apareixen totes les dades i el tipus que ha de tenir cadascuna, i si no coincideixen els has de dir: "Vau dir que aquesta dada seria decimal, però m'esteu donant una *string*."

Una cosa habitual als Estats Units passa amb els codis postals, que a vegades poden començar amb un zero, però si es tallen i es deixen enters, cau el zero inicial i obtens un codi postal escurçat que comença amb al segon número, cosa que passa sovint. Així que si veus que és un número diferent, saps que algú el va convertir en un número enter en lloc de en una *string*, que preservaria aquest zero inicial. La geocodificació és un altre problema...

Concha:

Ah, crec que ja n'has parlat sobre aquest. D'acord, hi ha més preguntes...

Al Shaw:

Aquí hi ha una guia que va fer un editor que solia treballar amb nosaltres i que inclou tot això, una guia per deixar les dades a prova de bales. En recomano la lectura.

Concha:

Excel·lent, així que indiqueu com fer-ho.

Al Shaw:

Sí, i tots els punts que he mencionat són a la guia.

Concha:

D'acord, doncs ens assegurarem de posar una còpia d'aquesta diapositiva en algun lloc. Ara una pregunta d'en Jesús Escudero, de *El Confidencial*.

Jesús Escudero:

Hola, com comenceu una investigació a ProPublica? Teniu primer la pregunta i després creeu la base de dades per respondre-la o teniu una base de dades i responeu la pregunta? Crec que dependrà de cada investigació, però quina és la forma més habitual?

Al Shaw:

Passa de les dues maneres. Moltes vegades tenim unes dades, sabem que sortiran o les hem demanat molt abans i tenim algunes hipòtesis sobre el que podríem trobar-nos. No sabíem el que obtindríem amb les dades dels centres educatius, però les vam aconseguir de manera exclusiva. Teníem una font al Departament d'Educació que ens va dir que no ens les donarien a nosaltres abans de fer-les públiques, així que vam pensar que potser podríem avançar-nos per veure què podia haver-hi d'interessant. Ens vam arriscar una mica, perquè en començar no sabíem que hi hauria una notícia allà, desconeixíem si trobaríem el tipus de desigualtat que finalment vam trobar. És a dir, moltes vegades comencem amb les dades i tenim l'esperança que hi hagi una notícia. Altres vegades ens han informat d'alguna cosa i tenim documents, per exemple, que proven alguna cosa, de la manera tradicional, quan saps que ha passat alguna cosa i busques les dades i els documents que ho secundin. Així que passa de les dues maneres.

Concha:

En Jesús està fent moltes peticions de llibertat d'informació perquè tenim problemes per obtenir les dades a Espanya. Jo he estat piulant sobre aconseguir dades escolars a Catalunya perquè vaig enviar una petició al Govern per demanar quantes noves places a les escoles comptaven amb beques parcials i quantes totals, i van respondre que no ho sabien i que havia d'omplir un formulari per convertir 10.000 en xifres de dos dígit. Van dir que no tenien aquesta informació. És totalment diferent aquí, oi Jesús?

Senyora:

Quant a aquestes dades, d'on les traieu? Sempre són públiques, teniu accés preferent...?

Al Shaw:

Depèn, a vegades és públic, com la taula sobre les eleccions, amb les campanyes i els grups independents. A vegades algú ens ho dona, altres ho demanem i som els únics que ho tenen perquè hem estat els únics que ho hem demanat. A vegades porta anys que ens les donin, però quan arriben som els únics que les tenim. Altres vegades creem les nostres pròpies dades, tenim alguns documents i, a mà, els convertim en dades. Això ho he fet molt amb registres d'impostos, que només estan disponibles com a formularis, no com a dades. Amb grups sense ànim de lucre hem hagut d'agafar les dades, posar-les en una fulla de càlcul i crear el grup de dades a mà. Podria dir-se que les dades sobre els impostos són de lliure accés però que no són realment dades, són formularis escrits, així que ningú fa aquesta notícia perquè resulta molt difícil obtenir les dades a partir dels documents. Per la qual cosa varia molt, a vegades són dades lliures, altres vegades les rebusquem per pàgines web, com al nostre projecte *Dollars for Docs*, que tracta dels pagaments de les farmacèutiques als metges per conferències, àpats, etc. Aquestes dades les vam obtenir buscant les pàgines web de les empreses, perquè estan obligades a dir quant paguen als metges, però no a posar-ho en un format clar, així que vam acabar revisant-ho tot i intentant unir-ho perquè es veiés quant s'estava emportant un metge determinat de cada empresa. Cada empresa tenia les dades en un format diferent i vam intentar combinar-les, per això vam crear un grup de dades a base de rebuscar i unir-ho, no eren dades d'accés públic.

Concha:

En aquest país aquesta informació és secreta. Ni tan sols podríem somiar en saber quant paguen les farmacèutiques als metges, amb noms i cognoms.

Al Shaw:

La veritat és que recentment han canviat moltes coses que ara fan que sigui més fàcil obtenir aquesta informació als Estats Units. Abans ho havien de revelar en la forma que volguessin, però ara el Govern està creant una base de dades per ajuntar-ho tot i que la gent hi pugui accedir. S'està millorant als Estats Units.

Senyor:

Feu alguna estadística sobre l'ús d'aquestes aplicacions? La gent hi busca la seva pròpia notícia o les usen per veure una de les coses i dir que és una bona notícia, per exemple: "Ah, quina app més bona", i després de dos clics ja està?

Al Shaw:

De fet, les nostres aplicacions de notícies són molt més populars que les nostres notícies. La de *Dollars for Docs* té milers de visites al dia. Publiquem notícies que potser un dia tenen moltes visites, però les aplicacions viuen molt més temps, són el més popular de tota la nostra pàgina web. Ens agrada fer-les, tant perquè ajuden la gent a

trobar-se a si mateixa en la notícia com perquè són resistents, es converteixen en recursos per alguna cosa posterior d'una manera que una notícia ni pot fer. Són molt, molt populars.

Concha:

Podeu identificar si majoritàriament es tracta de gent escrivint el nom del seu metge o...?

Al Shaw:

Sí, podem.

Concha:

I aquest és l'usuari més habitual?

Al Shaw:

Sí.

Concha:

Això és el que faríem tots.

Senyora:

Jo tinc dues preguntes. La primera és: després del projecte, quines accions va prendre el Govern? I també, podries parlar una mica més del model de negoci que és ProPublica?

Al Shaw:

Quin projecte, específicament?

Senyora:

L'últim del què has parlat.

Al Shaw:

El de Louisiana? No hi ha hagut cap acció encara, és una llàstima. El Govern té un pla de 50.000 milions de dòlars per intentar arreglar això que se suposa que duraria 50 anys, vam escriure sobre això en el seguiment que vam fer, però amb prou feines hi ha diners a Louisiana, que és un estat molt conservador, per la qual cosa no es dona gaire suport a les causes mediambientals. Suposo que va fer que la fent coneguéssim el problema, però no va crear un canvi, lamentablement. No tot el que fem ho aconseguim; ho vam intentar, però no ha passat en aquest cas, almenys per ara, potser més endavant. Quant al model de negoci, som una organització sense ànim de lucre, així que ens finança principalment la filantropia, sobretot fundacions i algun donant individual. Tot va començar amb un filantrop principal que va donar una gran quantitat perquè poguéssim començar el 2008 i que s'ha anat retirant i ja altres fundacions han aportat més, així que és una mescla de fundacions... [Li pregunten alguna cosa inaudible] Una mica de diners, però no el suficient per tirar endavant, és un compte petit.

Karma:

Gràcies per la teva presentació. Tindreu l'oportunitat de fer un taller amb l'Al Shaw demà al matí a Blanquerna. Al, breument, què faràs al taller? Ens ho pots explicar?

Al Shaw:

El taller tractarà principalment sobre com fer mapes com el que heu vist en la imatge satèl·lit de Louisiana; com agafar les dades del Govern, perquè als Estats Units el Govern les dona, les pot descarregar qualsevol; com crear un compte per aconseguir aquestes dades i com processar-les. Parlarem de com fer això; de què són les dades satèl·lit i com entendre-les, perquè hi ha bandes diferents, les RGB i les no visibles; de com accedir a aquestes dades, buscar-les, descarregar-les i processar-les, fent servir tant eines d'accés lliure com Photoshop. Així que és interessant, hauríeu de venir.

Karma:

És un apassionat, com veieu, de les imatges per satèl·lit, així que demà tindrem una hora al matí perquè ens expliqui tot això.

(Fi de transcripció)