

Objeto y ámbito de estudio

Las bases de datos masivas de proyectos de arquitectura como riesgo para la privacidad

Arquitectura, Big Data y Protección de Datos

Marco teórico y conceptual

“[Los datos son el petróleo del siglo XXI](#)”, es un mantra que se repite en multitud de publicaciones, siempre asociado a lo que se conoce como *big data*. Éste no constituye por sí solo un cambio de paradigma todavía y [tiene sus limitaciones](#), pero es indudable que estamos ante una evolución tecnológica cuyo potencial amenaza con sacudir los cimientos de muchos sectores, entre ellos el de la arquitectura.

Los arquitectos oímos hablar del big data como algo que servirá para mejorar la gestión del tráfico o para la optimización del consumo eléctrico procedente de las luminarias de las calles.

Pero esto es **solo una mínima parte** de lo que este campo de conocimiento basado en el uso de algoritmos informáticos puede generar y que tendrá profundas implicaciones en aspectos **éticos** y relativos a **protección de datos** en arquitectura, a todos los niveles.

La arquitectura y el urbanismo medidos en tiempo real

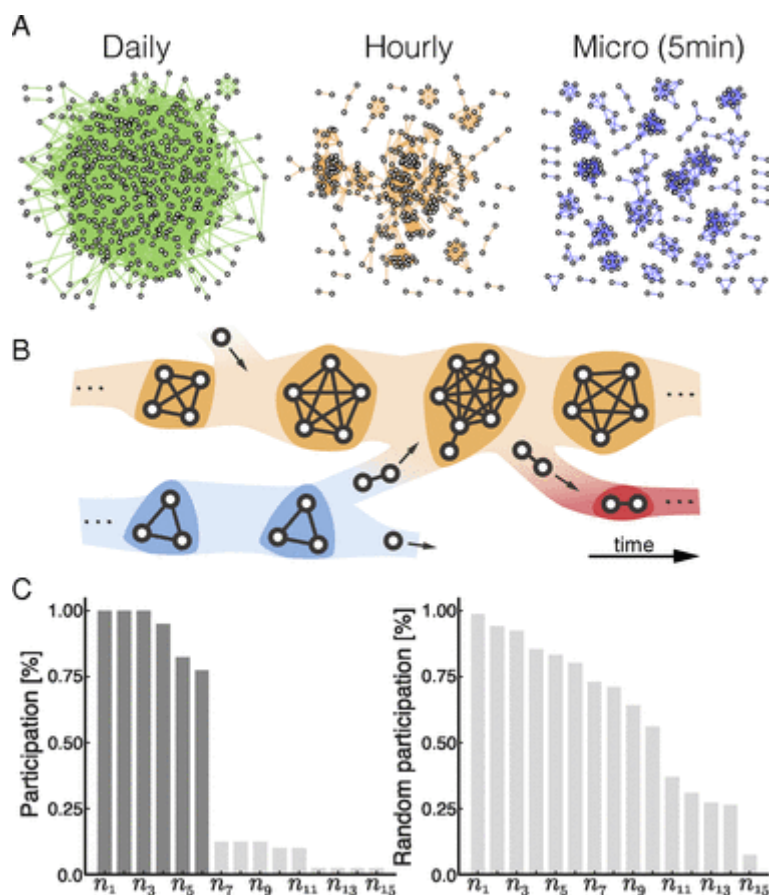
Los arquitectos hacemos habitualmente proyectos urbanos y edificios con datos reales muy escasos acerca del comportamiento de las personas. No se suelen decir cosas como "la creación de este barrio con esta tipología de vivienda en lugar de esta otra aumentará los niveles de satisfacción del ciudadano en un 5%", o "reducirá los niveles de obesidad de los residentes en un 6%".

Tampoco sabemos si los usuarios realmente usan “esa vivienda unifamiliar que les hemos diseñado” como esperábamos. Si usan la terraza en verano, o si las ventanas permanecen cerradas para optimizar la calefacción.

Una vez entregadas las llaves, es "su isla". Su intimidad.

Su **secreto**.

Los actuales sistemas de sensores distribuidos por el mundo en forma por ejemplo de teléfonos móviles permiten conocer [cómo se comportan los grupos de personas](#) de verdad, **sin necesidad de preguntarles**.

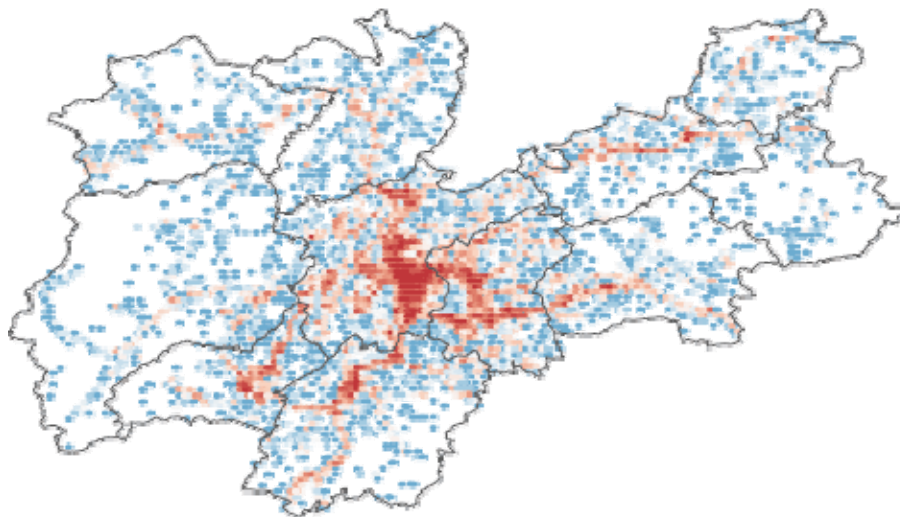


[Fuente](#)

Es un hecho contrastado científicamente que las personas tenemos comportamientos bastante predecibles la mayor parte del tiempo. Disponiendo de las bases de datos adecuadas, manipular éstas para obtener información del comportamiento es relativamente **barato** (se puede hacer con software libre), aunque **técnicamente complejo**.

En lo que respecta al diseño del territorio y al espacio urbano el big data supone de facto una revolución. Decisiones sobre si hay que peatonalizar tal o cual calle, tener información sobre cuáles son las "zonas muertas" de una ciudad y a qué horas y días exactos **es hoy más fácil que nunca**.

El big data permite generar mapas con información de la [capa relativa a la realidad social](#). Estos mapas era impensable crearlos hace solo unos años atrás.

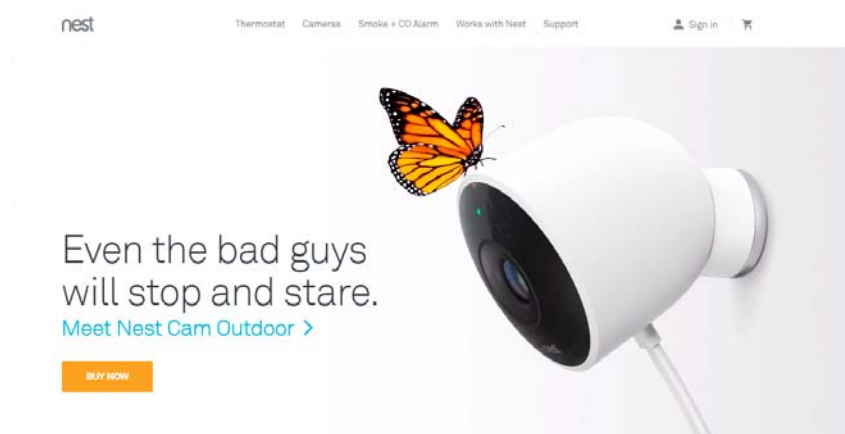


[Fuente](#)

Y no solo hablamos de información "sociológica". Las redes eléctricas en países emergentes como Senegal se diseñan ya [atendiendo a la actividad de las redes de telefonía móvil](#). También el análisis territorial en base a los datos procedentes de los teléfonos móviles está sirviendo para optimizar la [gestión de eventos](#), o las [estrategias turísticas de los territorios](#). Esto es solo el principio. El “**Urbanismo del big data**” ya trabaja con el análisis de cómo parámetros más específicos, [como el estado financiero de las personas](#), afectan a los patrones de movilidad en las ciudades.

La transformación de la gestión del hecho urbanístico no solo está ocurriendo desde el lado público. [Empresas privadas](#) están utilizando el big data para ofrecer herramientas que sirvan a los profesionales del comercio para la mejor toma de decisiones, afectando a aspectos como la implantación del negocio. También en el **mercado de la vivienda** están desarrollando herramientas que influyen en la fijación de [precios](#).

En lo relativo a los **edificios**, la dificultad de obtener datos de posición precisos de los teléfonos móviles en interiores hace más difícil el uso de estos datos como fuente de big data. Para saber lo que pasa dentro de casa las grandes tecnológicas tratan de obtener la información por otros medios, a través de información por ejemplo del comportamiento energético de la vivienda o incluso de la voz de los usuarios, con dispositivos como [Echo](#) o [Nest](#).

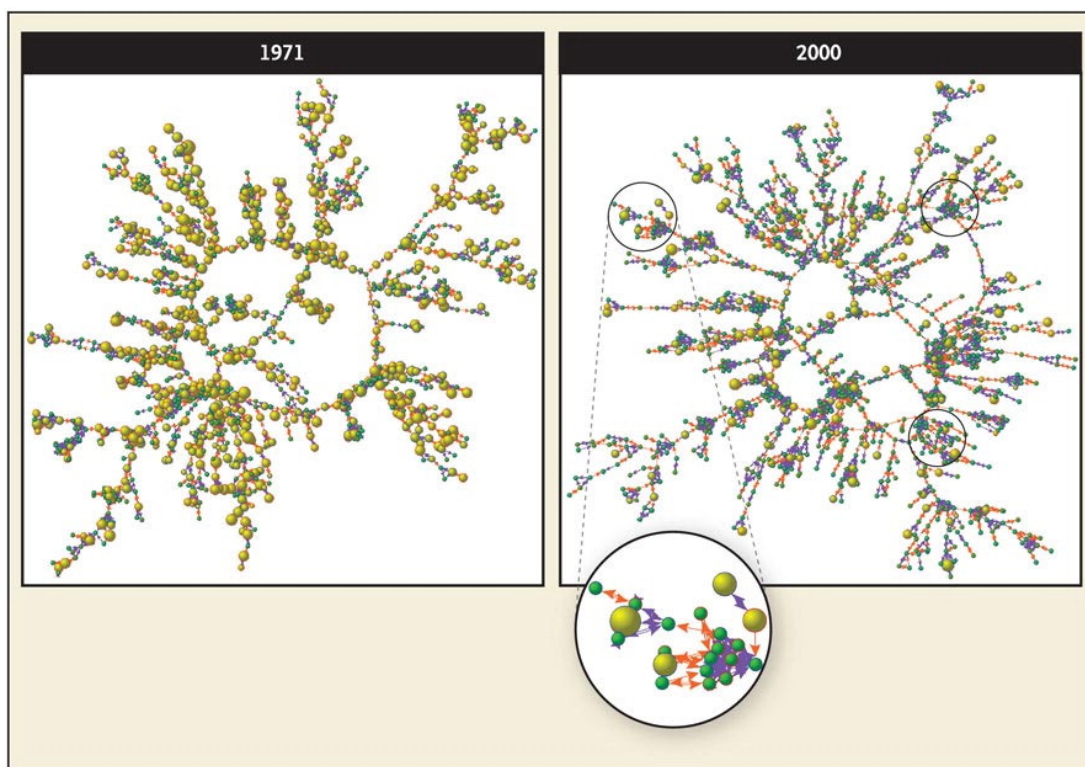


[Fuente](#)

Con el móvil y la generalización del **IoT** (*Internet of Things*, o Internet de las Cosas) casi cualquier aparato acabará convirtiéndose en un potencial **proveedor de datos de comportamiento dentro de las viviendas**, como [recientes noticias](#) se encargan de alertar.

Desde una perspectiva en positivo esto **puede servir a los arquitectos** para implementar **círculos virtuosos de aprendizaje continuo** y de testeo de las soluciones arquitectónicas. El "secreto" de si una arquitectura funciona bien está más cerca de ser desvelado, quizás para horror de algunos arquitectos que verán sacada a la luz pública las ineficacias de sus propuestas.

El big data también sirve para medir cambios en el comportamiento de las personas pero [también para generarlos](#). Gracias a lo que se conoce en el campo de la psicología como **influencia social** muchos comportamientos humanos funcionan [como si fuesen agentes contagiosos](#). Las personas dejan de fumar en grupo. Engordan en grupo.



[Fuente](#)

Podríamos imaginar una especie de **influencia del entorno construido** y podríamos medirla y testarla usando big data. La **arquitectura como motor de cambio de comportamiento medible** está muy cerca si no está ya aquí.

Del grado de control que podamos tener sobre esta influencia dependerá en gran medida el éxito de la implantación del big data en arquitectura. Para que podamos decir con precisión cosas como “use esta tipología de vivienda que reducirá la obesidad en un rango del 3 al 6% en un 80% de los usuarios”.

Datos de arquitectura en Internet. Riesgos ocultos

Aparentemente una planta de una vivienda es algo inofensivo. Cuando vemos una en un portal de arquitectura en internet pensamos: "La posibilidad de un uso perverso es remota. Podría ser útil a un ladrón para robar pero si no indico la dirección postal no pasa nada".

En la mayoría de los casos es así. Y de hecho no suelen conocerse casos de un uso delictivo directo de este tipo de información arquitectónica en la red. Sin embargo **el uso inadecuado puede darse por otras vías**, más "difusas".

En la [Política de Privacidad](#) de Architizer ("the largest database for architecture and sourcing building products", con alrededor de 40.000 proyectos según información en su portal) podemos leer:

"We may also share your information with third parties in order to inform you about products or services that may be of interest to you, **to supplement their information about you**, or for other business purposes".

Traducido al castellano sería algo así como:

"Podemos también compartir tu información con terceros para informarte sobre productos y servicios que pueden interesarte, **para complementar su información sobre ti**, o para otros propósitos del negocio".

Este tipo de frases están en la base de [conflictos de gigantes como Facebook](#) o Google [con las autoridades europeas](#). De un lado la protección de las personas, de otro las realidades de comportamiento y la aparente poca importancia que damos a estos asuntos. Un poco más allá, el negocio.

El 25 de Mayo de 2018 se comenzará a aplicar la General Data Protection Regulation (GDPR) de la Unión Europea. Reemplazará a la Directiva actual y se aplicará directamente en los Estados Miembros sin necesidad de implementar una legislación nacional. Esta nueva regulación [implicará muchos cambios](#) en lo relativo a la protección de datos. El planteamiento europeo contrasta fuertemente con las [posturas de EEUU](#).

En el ejemplo de una planta de una vivienda en un portal de internet debemos pensar que para una persona que posea las bases de datos adecuadas **reconstruir de quién es esa vivienda** a través de otros datos (incluso públicos) puede llegar a ser muy fácil. Es lo que hace el buscador de Google. Seguro que muchos solamente utilizando el buscador serían capaces de encontrar dónde está esa vivienda.

Algunos podrán pensar que la solución es la **anonimización** de los datos. O encriptar algunos campos clave. Esto es, **sencillamente, inútil**. El conocimiento sobre [reidentificación de usuarios anonimizados en bases de datos](#) lleva a esa conclusión. La solución es transformar las bases de datos en otras más generalistas, pero entonces son las bases de datos las que se vuelven inútiles.

En este sentido es célebre la anécdota del Gobernador de Massachussetts [que recibió su propio historial médico por correo postal](#) después de que unos años antes el Gobierno de dicho Estado tomase la decisión de publicar las bases de datos

anonimizadas de historiales médicos por considerar que esto era beneficioso para la sociedad con fines de investigación.

¿Cómo pudo ocurrir esto? Una estudiante del MIT desanonimizó la base de datos de historiales médicos **combinándola con otras bases de datos**. Y aquí es donde entra la frase de la Política de Privacidad de Architizer. Recuerda: “**para complementar su información sobre ti**”.

BIM + big data: un riesgo importante que debe gestionarse adecuadamente

Imaginemos que no solamente tenemos las plantas en formato de imagen, sino muchos más parámetros de la arquitectura del edificio en formato digital. Aquí es donde entra con fuerza BIM.

El conocido como **BIM** (*Building Information Model*) añade **nuevos grados de riesgo**. Un edificio diseñado en un programa tipo Revit o Allplan convierte al edificio en un modelo paramétrico. El diseño es un conjunto de datos digitales dimensionales, cuantitativos y cualitativos del edificio. A partir del modelo BIM se pueden extraer fácilmente datos de superficie, materiales, tipo de estructura... y mucho más.

Es momento de revisar BIM como una fuente de datos personales. Pero, ¿hay **datos personales** de los futuros inquilinos de un edificio en un modelo BIM?

Imaginemos el caso más evidente de una vivienda para una persona discapacitada que ha implicado adoptar determinadas acciones de diseño. Esa discapacidad podría ser posible "leerla" del modelo BIM de su vivienda. ¿Debería poder circular por Internet ese modelo BIM?

Vayamos más allá: ¿podrían extraerse de ese archivo parámetros como la **orientación religiosa** de una persona? ¿O la **vida sexual**? ¿O sus inclinaciones **políticas**?

Cualquier arquitecto sabe que determinados modos de vida tienen plasmaciones espaciales construidas concretas. Una persona en silla de ruedas, el ejemplo más evidente quizá, requiere de unos espacios para radios de giro de la silla.

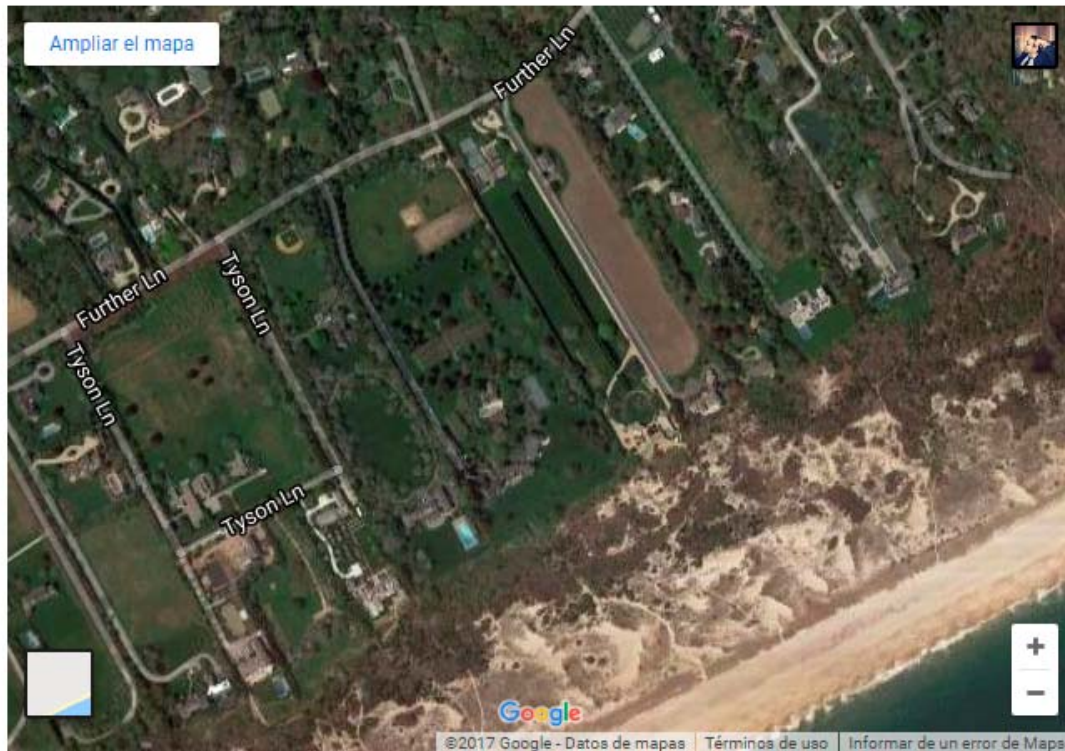


Viviendas para discapacitados. [Fuente](#)

Muchos dirán: bueno sí, pero *solamente* con un modelo BIM de su edificio es muy difícil estar seguro de que una persona es de una determinada manera. Y aquí es donde viene la clave donde enlazamos con el ejemplo del gobernador de Massachussetts o la Política de Privacidad de Architizer citadas. Quizá solamente con un modelo BIM no podamos extraer ciertos datos. Pero **combinando** la base de datos que nos da un modelo BIM con otras bases de datos podemos construir armas muy potentes contra la privacidad.

Hazte las siguientes preguntas: ¿qué podríamos llegar a saber si combinamos la información de un modelo BIM con los datos procedentes de **tarjetas de crédito** de sus inquilinos? ¿Y la información de sus **redes sociales** (incluido el procesamiento de los textos de los mensajes)? ¿Y de las redes sociales de sus vecinos? ¿Y la de los **móviles**? ¿Y de dispositivos como **Echo**?

Es conocido que a partir de la manera en que se usa el móvil [se puede llegar a ser capaz de saber la edad y el género de una persona](#) con una precisión bastante elevada. Y es conocida la inmensa cantidad de información sobre personas concretas que se puede obtener por ejemplo solamente con el big data procedente de los [desplazamientos en taxi de una ciudad](#). En la imagen, extraída de [este artículo](#), se muestra cómo es posible identificar, en bases de datos públicas, el punto desde el que alguien cogió un taxi para un trayecto en el que se gastó 400\$, dejando más de 100 de propina. ¿Vivía en la casa de al lado?



But I do know the *exact* Brooklyn Heights location and time from which someone (not necessarily the owner) hailed a cab, rode 106.6 miles, and paid a \$400 fare with a credit card, including a \$110.50 tip. **If the TLC truly wanted to remove potentially personal information, they would have to remove latitude and longitude coordinates from the dataset entirely.** There's a tension that public data is supposed to let people know how well the taxi system serves different parts of the city, so maybe the TLC should provide census tracts instead of coordinates, or perhaps only coordinates within busy parts of Manhattan, but providing coordinates that uniquely identify a rider's home feels excessive.

Ahora imaginemos que tuviéramos no la planta de una sola vivienda, sino las plantas de miles o incluso millones de edificios, con lo que la afectación potencial de la aplicación de determinadas tecnologías llegase a millones de personas. Es lo que está ocurriendo.

Es momento de revisar BIM como una fuente de datos personales.

El salto cualitativo del big data en arquitectura

No entraré en profundidad en detalles técnicos sobre este punto, pero adelanto solamente que el disponer de datos masivos aporta tres novedades sustanciales:

- la **posibilidad de predecir**
- la posibilidad de **generar inteligencia artificial** que realice tareas que sustituyan a las personas

- la posibilidad de **conocer cuáles son los parámetros más relevantes que condicionan un comportamiento** en las personas.

Sabiendo cuáles son los parámetros más relevantes [recientes estudios](#) muestran cómo utilizando técnicas de big data es posible **vender 13 veces más** que con las técnicas tradicionales basadas en la experiencia de los especialistas en marketing.

¿Se comprende ahora el **potencial disruptivo** y de creación de asimetrías y distorsión de los entornos socioeconómicos de estas tecnologías?

El big data y sus aplicaciones en Inteligencia Artificial pueden servir para construir **sistemas que generen diseños** basados en el **aprendizaje realizado sobre millones de proyectos de arquitectura**. Y es posible que estos sistemas sean tan avanzados que **ni los arquitectos ni los propios programadores entenderán** los procesos que el sistema desarrolla para tomar decisiones, [como ya está ocurriendo](#) en otros campos donde se están implementando sistemas de Inteligencia Artificial.

Esta es una amenaza para los puestos de trabajo en arquitectura tan real como lo es el coche autoconducido o Uber para los taxistas.

El arquitecto será fundamental en estos procesos, ya que los algoritmos no son perfectos, y generan problemas conocidos que deben ser controlados, pero quizá muchos dejen de ser necesarios. De hecho, uno de los mayores riesgos a los que se enfrenta la humanidad son los sistemas pilotados por algoritmos que pueden generar sesgos y desviaciones sistémicas. Algo que ha llegado a la [Oficina Ejecutiva del Presidente de los EEUU](#) y que se conoce como "**data discrimination**". En este asunto, nuevamente aparecen [noticias](#) que dan cuenta de estos riesgos.

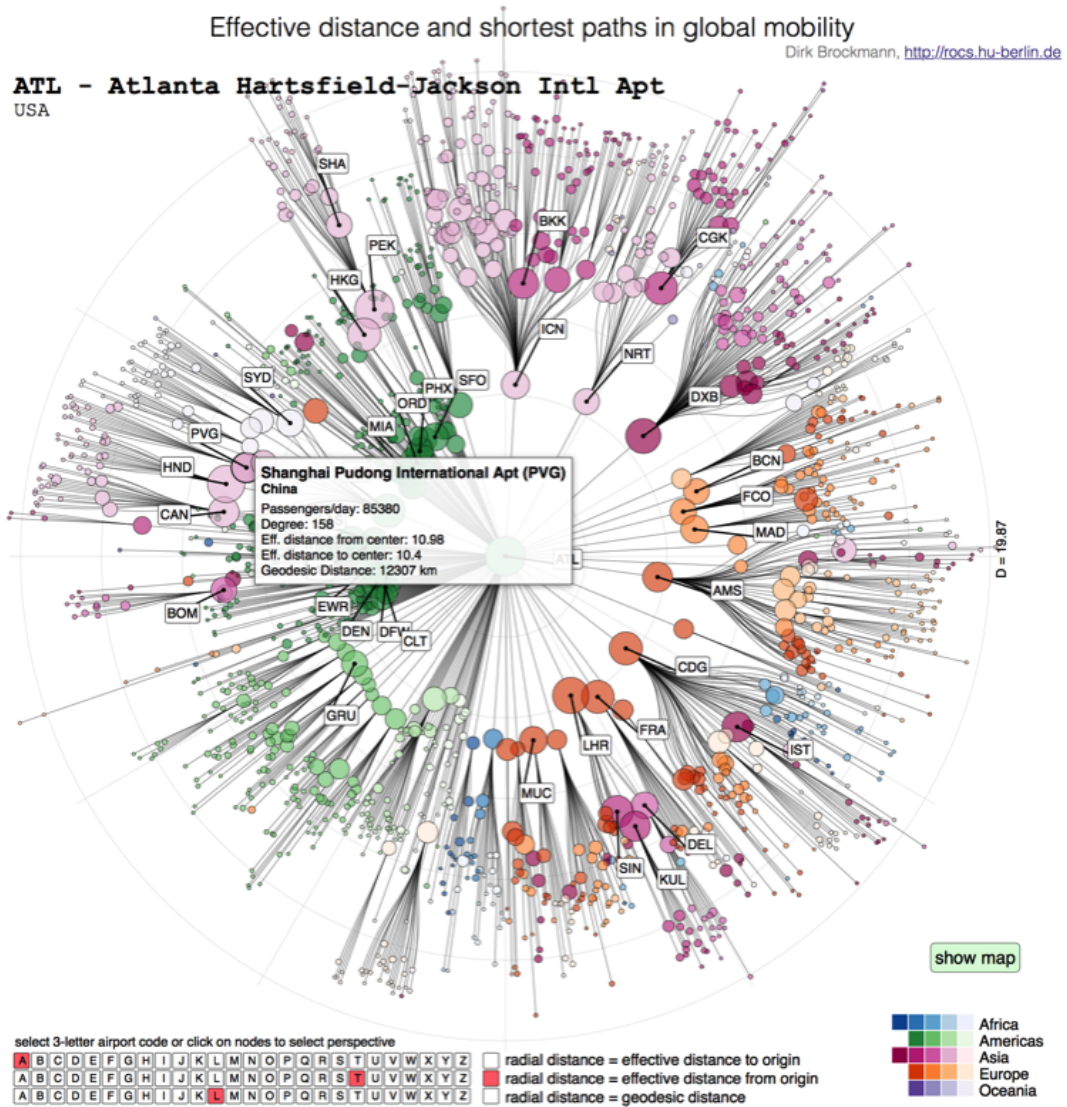
Nuevas maneras de hacer arquitectura

El potencial del uso de los datos en arquitectura va mucho más allá de lo imaginable actualmente, y los arquitectos deben familiarizarse con este ámbito de conocimiento. El big data es un asunto [técnicamente complejo](#), pero no inaccesible.

Algunos arquitectos que decidan entrar en este campo poco a poco irán habituándose a términos como *spectral clustering*, o *eigenvector*, que tan ajenos pueden sonar ahora.

Y esta entrada **generará nuevas corrientes de conocimiento en la disciplina**. Conceptos como la salubridad de un edificio pueden adquirir nuevos matices y convertirse en cualidades dinámicas medidas en tiempo real. Gracias al big data, añadido a la capa relativa a la estructura social y espacial de un edificio podremos evaluar la salubridad de un edificio en relación a su situación respecto de una [epidemia](#) de gripe, por ejemplo.

Conceptos del campo de conocimiento del big data, como "[distancia efectiva](#)", una suerte de distancia virtual entre lugares, en la que la distancia se mide en parámetros no evidentes, como "tiempo de llegada de una enfermedad a una ciudad desde la fuente de contagio inicial" pueden traer nuevas visiones y nuevos modos de desarrollar proyectos.



[Fuente](#)

Y sobre todo el big data puede ayudarnos a **responder a algunas de las cuestiones más trascendentes** para la arquitectura desde hace siglos.

Los edificios definidos como entidades digitales a través de modelos BIM, y en conexión con sistemas digitales como los teléfonos móviles y aparatos incorporando Internet de las Cosas, pueden transformar los edificios en Living Labs.

Los Living Labs son algo relativamente conocido en el sector retail. Múltiples empresas, algunas españolas como [Mercadona](#), los utilizan y han transformado la manera de entender estos espacios. Pero el potencial de estas técnicas para el diseño de edificios de uso residencial, hospitalario, educativo, etc. está por explorar, así como sus implicaciones en términos de seguridad o privacidad.

El potencial del uso de los datos en arquitectura va mucho más allá de lo imaginable actualmente. Muchos estudios de arquitectura están explorando qué puede significar esta nueva tecnología para la disciplina. Del mismo modo, universidades punteras en

todo el mundo, como el MIT están invirtiendo importantes esfuerzos en estar en primera línea de este revolucionario campo de investigación.

La necesidad de una infraestructura de gestión de los datos de arquitectura y de los arquitectos. El caso de España

Mientras esta revolución tiene lugar es necesario gestionar adecuadamente todos los riesgos que para la profesión de arquitecto y para la propia arquitectura se están generando. Antes de que sea demasiado tarde.

Subir determinada información a Internet es peligroso, todos lo sabemos. Pero alguna información aparentemente inofensiva [puede llegar a ser peligrosa](#) cuando se procesa con herramientas de big data.

En el caso de la **información de la arquitectura**, a los riesgos de privacidad se suman otros que tienen que ver con la seguridad, la economía, el medio ambiente, e incluso **militares**.

La gestión de grandes bases de datos de arquitectura y de los arquitectos debe estar gestionada y supervisada adecuadamente. Insistimos, [una base de datos anonimizada no es suficiente](#), y con ella una persona con los conocimientos adecuados puede extraer [datos muy relevantes](#) sobre la vida de personas concretas. Si además, las bases de datos se combinan el riesgo aumenta exponencialmente.

Es preciso evaluar la necesidad de crear una **entidad que gestione los datos procedentes del sector de la arquitectura** o, si se quiere más ampliamente, de la construcción, y que implemente los protocolos y tecnologías que se están desarrollando a nivel científico para garantizar la protección más adecuada y más ética sobre esta información.

Posiblemente lo ideal es que se cree una estructura institucional, del mayor rango posible, para tratar esta información. Dado que la protección de datos es [diferente en las distintas regiones del mundo](#), una estructura unificada mundial parece poco viable, aunque deseable.

En España **los Colegios de Arquitectos** acumulan cantidades ingentes de información sobre arquitectura, principalmente a través del Visado y los Archivos Documentales y podrían cumplir esta función. Como vocal de cultura de un Colegio de Arquitectos que he sido durante 6 años, y responsable de la creación y desarrollo de la única red social de arquitectura entidad colaboradora oficial del CSCAE: Plazatio, conozco esta realidad institucional.

Actualmente los Colegios en su mayoría disponen de lo que se conoce como **"dark data"**, esto es, datos que se guardan pero que no se analizan suficientemente ni se explotan para que aporten valor. O en algunos casos no están estructurados.

Si los Colegios de Arquitectos estructurasen, almacenasen, analizasen y destilasen adecuadamente la información que poseen sobre el sector, ¿cuánto valor se podría trasladar a la sociedad y al entorno construido?

¿Seríamos capaces de anticipar crisis del sector? ¿Seríamos capaces de reducir impactos medioambientales?

En este asunto hay que reflexionar sobre cómo están los Colegios de Arquitectos, y en especial el CSCAE, abordando la revolución de Internet y las nuevas tecnologías. Cuando se habla de Google o Facebook se dice: “Recuerda, si no eres el cliente eres el producto”. **¿Están siendo los arquitectos el producto de plataformas como Architizer, Archilovers, etc?**

Es habitual ver campañas en las que los arquitectos animan a sus contactos a “votar” su obra en este tipo de plataformas, para conseguir premios honoríficos. Cuando los arquitectos hacen esto están dando, sin saberlo, a esa web su **red de tus contactos** vinculados a esa obra. Esta red es uno de los factores más importantes cuando se trata, por citar un ejemplo, de vender apps de móvil. Y sin embargo, el arquitecto no gana un solo euro de ese trabajo de crear esa red, ni de las ventas generadas. Esto no es conocido por el colectivo de arquitectos.

Cuando esa base de datos, con la red de ese arquitecto incluida, sea enorme y se activen **sistemas predictivos de machine learning** que hagan innecesarios muchos arquitectos, todo desarrollado en base a datos de esos mismos arquitectos, ¿exigirán esos arquitectos su parte del negocio? ¿Exigirán que le dejen usar los datos para poder hacer tus proyectos urbanos, o pagarán por ellos?

No olvidemos algo crucial. Una vez se es el primero en tener un sistema predictivo potente, los competidores probablemente quedarán atrás rápidamente ya que el sistema se retroalimenta. El buscador de Google es un buen ejemplo.

Por ello es probable que sí sea necesario crear una entidad que gestione los datos procedentes del sector de la arquitectura

Desde el Colegio Territorial de Arquitectos de Alicante se ha creado bajo mi dirección la red socioprofesional Plazatio, para que los arquitectos dispongan de una alternativa que busca garantizar **un uso ético de los datos**. Garantizado en la medida que lo que en ella se almacena tiene un control que depende de órganos de gobierno elegidos y gestionados democráticamente.

Hay que tener en cuenta que **ningún Colegio de forma individual dispone de suficientes datos para generar una infraestructura de gestión de big data**.

El tener datos de comportamiento online de colegiados puede llegar a tener otras implicaciones importantes a las que el CSCAE debe prestar atención. Un mal uso de estos datos **podría afectar incluso a la estructura democrática**. También para la estructura democrática de los Colegios de Arquitectos y sus procesos electorales.

Si se consiguieran reconstruir datos de la red de relaciones entre colegiados se podría llegar a analizar quiénes son los colegiados con más centralidad en la red, teniendo por tanto mayor potencial de influencia. Es conocido que este tipo de técnicas las empleó Obama en las segundas elecciones a las que optó a la Presidencia de los EEUU.

Mientras todo esto llega hay una realidad urgente. **De momento, todos los Colegios y estudios de arquitectura deben adecuarse a la nueva regulación europea.** Recordamos, antes del 25 de Mayo de 2018 deberá estar listo, y el CSCAE tendrá que tomar nota e impulsar las medidas que corresponda.

Si los datos son el petróleo del s.XXI es preciso preguntarnos si queremos una sociedad de jefes de los datos. O si queremos democracia. Y actuar en consecuencia, ya que el tiempo se agota.

Objetivos

Nuestro proyecto se propone evaluar:

- Los riesgos de privacidad anexos a las bases de datos de proyectos de arquitectura digitalizados en distintos formatos, incluido BIM, y en especial en el caso de su difusión en abierto.
- El potencial de extracción de datos personales de la arquitectura en formato digital de forma aislada y en combinación con distintos tipos de bases de datos, y en especial los riesgos potenciales de que empresas como Facebook o Google adquieran bases de datos de empresas como Architizer, Archdaily, Houzz y similares. Y en especial si ésta incluye información de interacciones y redes de arquitectos / clientes / interesados.

Para ser capaces de estimar el grado de necesidad de:

- organismos de control y gestión de las bases de datos de arquitectura
- iniciativas legales
- limitar o restringir el tipo de datos que pueden vender o combinar.
- Medidas relativas a las normas de uso y políticas de privacidad de plataformas como Architizer.
- necesidades formativas a incluir en la titulación de arquitectura
- inclusión de parámetros de revisión a incluir en el Manual de Calidad del visado del CSCAE, atendiendo a nuevos factores de riesgo para la *seguridad de las personas*.

Y permitir el desarrollo de **innovaciones** vinculadas a:

- desarrollo de nuevas tipologías arquitectónicas
- desarrollo de funcionalidades específicas en software BIM como Revit y Allplan
- desarrollo de formatos de representación de proyectos en BIM optimizados para la interacción digital con productos que incorporen Internet de las Cosas (*IoT*).
- desarrollo de la estructura de seguridad de redes sociales de arquitectura como Plazatio, Architizer, Archilovers, Archdaily, etc.
- generación automática de proyectos mediante la aplicación de algoritmos de Machine Learning

Metodología a seguir

Se aplicará el método científico. Dado que se dispone de 6 meses para desarrollar el proyecto, se plantearía un proyecto de pequeño tamaño, siguiendo el principio "to fail small and quickly", para ser capaz de corregir el rumbo sin grandes implicaciones. En la fase inicial del proyecto se ajustaría su profundidad en función de los medios con los que se pudiera contar. Lo que a continuación se recoge es una estimación en un entorno ideal.

Personas e instituciones a implicar

Precisaríamos enlazar a diferentes grupos de interés para que nos aportasen su conocimiento, bases de datos y tecnología:

- Colegio Territorial de Arquitectos de Alicante - Plazatio.com (ya enrolado si fuese necesario).
- Architizer, portal de referencia mundial en arquitectura, con sede en Nueva York.
- Estudio de Arquitectura que desarrolle proyectos en BIM y que conozca BIM a un nivel profundo. Podría ser un estudio español (Estudio de Arquitectura de Antonio Galiano, con quien tengo contacto, y que es además profesor de BIM y fundador del BIM Lab de la Universidad de Alicante, centro autorizado de Autodesk, la empresa que produce Revit) o un estudio de Nueva York contactado a través de la Universidad.
- Especialista en GIS (Ángel Castaño, ingeniero español, ya enrolado si fuese necesario).
- Un equipo académico especialista en Big Data, de la Universidad de Columbia, o del MIT (Connection Science Department, David Shrier y Yves-Alexandre de Montjoye), quienes han sido profesores míos en el curso de Big Data & Social Analytics.
- Un equipo académico especialista en privacidad, del mismo departamento del MIT, o del equivalente en la Universidad de Columbia.
- Opcionalmente: compañías de telefonía o tecnológicas que trabajen con big data (Facebook, Google, Movistar, Orange, Vodafone). Tengo contacto con la directora científica de Vodafone, Nuria Oliver.
- Departamento de Arquitectura de la Universidad.

Proceso

En la investigación se presentarían varios retos. En función del acceso a la información de que pudiéramos disfrutar, el siguiente esquema sería aplicado sobre bases de datos 100% reales (edificios reales, con datos reales de personas reales, con comportamientos reales), en parte sobre bases de datos de estudios científicos en abierto, en especial en lo relativo a las bases de datos de comportamientos (por ejemplo podríamos usar la base de datos *Friends and*

Family). En caso de no poder contar con acceso a bases de datos reales, para lo que sería importante el apoyo de la Universidad, se desarrollaría un estudio más teórico.

Desde una perspectiva amplia, desarrollaríamos 3 fases: análisis descriptivo, análisis predictivo y análisis prescriptivo.

Fase 1. Análisis descriptivo (3 MESES)

a) Revisión sistemática de literatura:

- revisión de la legislación de privacidad española para establecer los tipos de datos personales y su clasificación por sensibilidades. Se podría analizar el caso de otras legislaciones nacionales.
- revisión de proyectos de arquitectura del siglo XX y XXI tratando de extraer conexiones entre gestos de diseño que se plasman en la documentación de proyectos arquitectónicos y datos personales sensibles de los usuarios que estuvieran afectados por aspectos de privacidad definidos por la legislación.
- artículos científicos en Elsevier buscando ejemplos de análisis de bases de datos de arquitectura y big data, de análisis de la conexión entre BIM y big data, de análisis de la conexión entre big data, privacidad y arquitectura.
- También en el ámbito del análisis automatizado de planos para extracción de parámetros en relación a aspectos de privacidad, desde la geomática, geoestadística y visión artificial. Analizaríamos qué algoritmos de análisis hay disponibles. Haríamos una clasificación de los tipos de datos que se pueden captar directamente, y aplicando estos algoritmos.
- análisis de la estructura de datos de un archivo BIM en formato Revit.

b) Adquisición de datos y almacenaje:

- Analizaríamos algunas plataformas tipo Architizer, Archilovers, Archdaily, Plazatio, plataforma de visado del Colegio Territorial de Arquitectos de Alicante, programa de memorias del Colegio de Arquitectos de la Comunidad Valenciana, bases de datos abiertas de GIS. Podría conseguir datos del Colegio Territorial de Arquitectos de Alicante fácilmente, con una autorización especial. Sería más complicado conseguir datos de Architizer o empresas terceras como Facebook, pero bajo el paraguas de una Universidad, podría ser factible. En cualquier caso, no es un requisito indispensable poder acceder a esta última información.
- Analizaríamos los tipos de datos de comportamiento humano que es posible obtener de bases de datos públicas o de pago, para

evaluar qué datos pueden ser potencialmente combinados.
Haríamos un primer listado de potenciales proveedores de dichas bases de datos: Vodafone, Facebook, Google...

- Daríamos un primer listado de expertos clave internacionales en este campo.

c) Data Mining and Limpieza:

- analizaríamos el nivel de estructuración de los datos. Afrontaríamos dificultades en la modelización y análisis, dado que nos encontraríamos con datos incompletos, falta de datos, falta de datos estructurados o con diferentes estructuras, sesgos y ruido. Utilizaríamos técnicas de tratamiento de bases de datos para su limpieza y preparación.

d) Agregación e Integración de datos:

Usaríamos un enfoque multimodal, tratando de obtener distintos grados y tipos de información, tanto cualitativa como cuantitativa, de edificios, entornos urbanos, personas y comportamientos.

- clasificaríamos los proyectos de arquitectura en función del tipo y grado de estructuración de los datos de los mismos: planos en formato de mapa de bits, webs de arquitectura, proyectos en formato BIM...
- Clasificaríamos los tipos de datos.

e) Análisis y Modelización de datos:

- Analizaríamos coincidencias y divergencias entre los datos captados por las distintas plataformas.
- Listaríamos los gestos de diseño que tienen una correlación o relación causal con aspectos de privacidad sensibles de los usuarios.
- Estableceríamos nuevos indicadores clave, como combinación de los datos disponibles.
- Desarrollaríamos gráficos estadísticos y tablas.

f) Interpretación de los datos.

- Evaluaríamos distintas combinaciones de bases de datos y estudiaríamos cómo afectan a parámetros clave de big data como la unicidad, o la uniqueness de los usuarios.
- Listaríamos los potenciales riesgos de privacidad encontrados, de bases de datos de forma aislada y en cada combinación.

- Estableceríamos oportunidades para los distintos formatos en su potencial combinación con otras bases de datos.

Fase 2. Análisis predictivo (1 MES)

En base al análisis desarrollado, trataríamos de:

- establecer hipótesis para predecir cuándo la base de datos de proyectos podría suponer riesgos de privacidad: qué tamaño debería alcanzar, cuándo se estima que alcance ese tamaño, etc.
- establecer los datos clave a captar en plataformas como Architizer, o en softwares de generación de memorias de proyectos, para mejorar las capacidades predictivas y de aplicación de técnicas de machine learning a los proyectos de arquitectura.
- establecer hipótesis relativas a las características de los proyectos, tratando de establecer patrones y tendencias. Por ejemplo, podríamos realizar análisis que muestren correlaciones o relaciones causales entre por ejemplo tipologías de proyecto y su mayor riesgo en términos de privacidad en Internet. También, podríamos establecer relaciones entre la actividad en Internet de los arquitectos en las plataformas para las que tuviéramos datos y los proyectos diseñados por ellos.

Fase 3. Análisis prescriptivo (2 MESES)

En base a los estudios previos, definiríamos:

- Cómo debería ser la base de datos de proyectos de arquitectura para que puedan aplicarse procedimientos de machine learning sobre ella, si existe ese potencial y cuándo podría alcanzarse.
- la urgencia o necesidad de una entidad que almacene y gestione los datos de la arquitectura.
- Daríamos un primer listado de potenciales mejoras en Software para BIM para permitir / evitar la extracción de información de privacidad de los archivos.
- Procedimientos de diseño arquitectónico que contemplen la seguridad de los datos en Internet que puedan suponer riesgos para la privacidad.
- Mejoras en las políticas de privacidad de plataformas como Architizer
- Sugerencia de iniciativas legales
- necesidades formativas a incluir en la titulación de arquitectura

- sugerencia de inclusión de parámetros de revisión a incluir en el Manual de Calidad del visado del CSCAE, atendiendo a nuevos factores de riesgo para la *seguridad de las personas*.
- Sugerencia de desarrollo de nuevas tipologías arquitectónicas más “seguras” en términos de privacidad en Internet
- Sugerencia de desarrollo de funcionalidades específicas en software BIM como Revit y Allplan para mejorar las características de privacidad de los proyectos en caso de ser difundidos a través de Internet.
- Sugerencia de formatos de representación de proyectos en BIM optimizados para la interacción digital con productos que incorporen Internet de las Cosas (*IoT*).
- Sugerencias de mejora para la estructura de seguridad de redes sociales de arquitectura como Plazatio, Architizer, Archilovers, Archdaily, etc.

Retos tecnológicos y relativos a las bases de datos

Afrontaremos diferentes retos debidos a las propias características del big data y de los formatos de los proyectos. Estableceremos estos retos, en función de algunas de las características clave de las bases de datos en big data:

- **Volumen:** tendremos que poder almacenar terabytes de información, algunos con información sensible, en entornos seguros.
- **Variedad:** dado que planteamos un enfoque multimodal, habremos de tratar con múltiples formatos de datos, estructurados y no estructurados. El tratamiento de los datos será más complejo.
- **Veracidad:** debido a la creciente complejidad de la estructura de los datos, posibles inclusión de datos anónimos, imprecisos o inconsistentes, presentes en las bases de datos grandes, será necesario realizar un proceso de limpieza de datos, que realizaremos empleando el lenguaje de programación Python.
- **Visualización:** Visualizar parte de la información de manera que sea fácilmente legible puede ser un reto, especialmente si se trata de información de redes. Emplearemos las librerías de Python desarrolladas al respecto para esta labor.
- **Valor:** Extraer valor de las bases de datos es el gran reto al que nos enfrentaremos.

Retos de gestión

Existirán varios retos, a nivel de gestión, que habremos de resolver:

- a)** Retos de privacidad: Puede ser necesario estar en contacto con los organismos de protección de datos americanos o españoles (AEPD) para desarrollar los protocolos, acuerdos y estructura de protección de datos en función del acceso que obtengamos a la información de bases de datos.
- b)** Retos de seguridad: deberemos emplear una infraestructura tecnológica segura, de manera que los datos estén seguros. La infraestructura de la Universidad será posiblemente la mejor candidata.
- c)** Intercambio de información: habremos de establecer contacto previo al inicio de la beca con las entidades a las que vayamos a solicitar información de sus bases de datos, de manera que haya tiempo suficiente para gestionar los permisos.
- d)** Costes operacionales: se pretende trabajar con lenguajes de programación como Python, y sus librerías gratuitas asociadas. No se espera que se generen costes por el acceso a la información, aunque probablemente se solicite que se compartan los resultados de la investigación, algo que habría que asumir.