IDP REVISTA D'INTERNET, DRET I POLÍTICA
REVISTA DE INTERNET, DERECHO Y POLÍTICA

www.uoc.edu/idp

ARTÍCULO

Esclavos de los macrodatos. ¿O no?

Mireille Hildebrandt
Cátedra de Entornos Inteligentes, Protección de Datos y Estado de Derecho
Institute for Computing and Information Sciences (ICIS)
Universidad Radboud de Nimega

Fecha de recepción: octubre de 2013
Fecha de aceptación: octubre de 2013
Fecha de publicación: octubre de 2013

Resumen
En este trabajo se debate la noción de macrodatos con relación a la monetización de los datos personales. Se revisa lo que afirman algunos de sus defensores y adversarios, según los cuales los macrodatos implican que «n = todos», en el sentido de que ya no es necesario utilizar muestras, puesto que disponemos de todos los datos, y se llega a la conclusión de que tal argumento es al mismo tiempo demasiado optimista e innecesariamente pesimista. Se presenta una serie de aspectos epistemológicos y éticos relacionados con las repercusiones de los macrodatos en nuestra percepción, cognición, imparcialidad y privacidad así como en los debidos procesos legales. A continuación, el artículo examina la idea de la gestión de datos personales centrada en el usuario, para averiguar hasta qué punto este tipo de gestión aporta soluciones a algunos de los problemas planteados por el enigma de los macrodatos. Se presta una especial atención al principio básico de la legislación sobre protección de datos, concretamente el principio de finalidad vinculante. Para terminar, este trabajo pretende indagar en la influencia que tiene la política de los macrodatos en la persona, la mente y la sociedad, y preguntarnos cómo podemos evitar el convertirnos en esclavos de los macrodatos.

Palabras clave
macrodatos, inteligencia artificial, monetización de los datos personales, gestión de datos personales centrada en el usuario, doble contingencia

Tema
macrodatos

Mireille Hildebrandt
Título original: *Slaves to Big Data. Or Are we?*

Slaves to Big Data. Or Are We?

Abstract

In this contribution, the notion of Big Data is discussed in relation to the monetisation of personal data. The claim of some proponents, as well as adversaries, that Big Data implies that 'n = all', meaning that we no longer need to rely on samples because we have all the data, is scrutinised and found to be both overly optimistic and unnecessarily pessimistic. A set of epistemological and ethical issues is presented, focusing on the implications of Big Data for our perception, cognition, fairness, privacy and due process. The article then looks into the idea of user-centric personal data management to investigate to what extent it provides solutions for some of the problems triggered by the Big Data conundrum. Special attention is paid to the core principle of data protection legislation, namely purpose binding. Finally, this contribution seeks to inquire into the influence of Big Data politics on self, mind and society, and asks how we can prevent ourselves from becoming slaves to Big Data.

Keywords

Big Data, artificial intelligence, monetisation of personal data, user-centric personal data management, double contingency

Topic

Big Data

Introducción

El problema que plantean los macrodatos es que n = todos.[1] Mejor dicho, el problema es la afirmación de algunos de sus defensores (y adversarios) para los que n = todos. «N = todos» resume perfectamente en qué consisten los macrodatos, cómo se definen y qué dificultades presentan. *Si esa afirmación fuera cierta*, los macrodatos podrían romper cualquier membrana protectora de nuestra vida interior y alterar el lugar más sagrado de nuestra privacidad y autonomía, ya que sus dueños llegarían a conocernos mejor —y a conocer mejor cualquier cosa— que nosotros mismos. *Si esa afirmación no fuera cierta*, los macrodatos también podrían eliminar nuestro sentido del yo y nuestra inferfaz con el mundo, hasta el punto de que no podríamos discutir sus resultados, tendríamos problemas para no rendirnos ante el conocimiento aparentemente claro y objetivo que proporcionan y no dispondríamos de las herramientas para averiguar qué perfil están creando de nosotros. *Si no fuera cierta*, los macrodatos generarían una discriminación errónea, pero *aun siendo cierta*, los macrodatos podrían generar una discriminación parcial o injustificable.

Por supuesto el problema es que hablar en términos de verdadero o falso con relación a los macrodatos es absurdo, porque los macrodatos tratan de elaborar modelos de datos. Sean modelos de arriba abajo o de abajo arriba, automatizados o incluso autónomos, la mejor pregunta es si los modelos funcionan, qué efectos tienen y cómo se distribuyen tales efectos. Finalmente —y lo más importante de todo— la pregunta es en qué clase de seres humanos nos convertiremos cuando interactuemos con los modelos que generan los macrodatos para analizarnos. He terminado hablando de los macrodatos como de «algo» que nos analiza, como si los macrodatos tuvieran mente propia. Sería muy fácil negarlo y atribuir todas sus predicciones a los diseñadores de tecnologías de macrodatos o a sus usuarios, las redes publicitarias, los *data brokers*, las autoridades judiciales, los científicos, las empresas de redes inteligentes y cualquier otro proveedor de servicios que se base

1. En la investigación empírica cuantitativa «n» equivale a la muestra, mientras que «todos» se refiere a la totalidad de la población. Si «n» fuera «todos», ya no nos podríamos referir a una muestra porque estaríamos investigando todos los casos de nuestro objeto de investigación, fuera cual fuera.

en los macrodatos para tomar decisiones, como conceder un crédito, dar un trabajo o hacer un seguro. O cualquier otro proveedor de servicios que *externalice* sus decisiones a los sistemas operativos autónomos de alta velocidad en tiempo real que cada vez determinan más nuestro entorno externo. Pensemos por ejemplo en los preparativos para la red eléctrica inteligente que procesará nuestros datos sobre consumo eléctrico en tiempo real y los combinará con unos precios flexibles, lo cual nos permitirá cargar energía en la red y venderla a nuestros vecinos más cercanos.[2] IBM ha acuñado el término de *informática autónoma*, que sugiere que estos sistemas informáticos adaptarán nuestro entorno externo exactamente igual que el sistema nervioso autónomo «dirige» nuestro entorno interno: de una forma a la que no tenemos acceso consciente y sobre la que no tenemos control directo.[3] Sin embargo, en la medida que los macrodatos sean lo bastante inteligentes para operar autónomamente, también tendrán que ser más listos que sus diseñadores y usuarios. Los macrodatos son inteligentes porque generan soluciones que nosotros no habríamos podido desarrollar, porque como seres humanos que somos no tenemos suficiente capacidad informática. Así, el uso de macrodatos genera una impredecibilidad parecida a la de los animales: por muy bien domados que estén, no controlamos completamente su comportamiento. Peor aún, pueden generar la volatilidad de las erupciones volcánicas, de los «actos de Dios» como decíamos antiguamente. Podría ser que, en la medida en que adoremos a los macrodatos, creamos en ellos y dependamos de sus oráculos, se conviertan en un nuevo panteón, repleto de nuevos dioses. Dioses creados por nosotros mismos, pero no necesariamente bajo nuestro control, lo cual indica que, en realidad, no es ninguna insensatez hablar de los macrodatos como si tuvieran mente propia, sin querer decir con ello que su mente es como la nuestra y sin olvidar que su mente ha sido desarrollada en principio por organismos científicos, gubernamentales o empresariales para obtener beneficios, para construir nuevo conocimiento y para mejorar la eficiencia y la efectividad de la Administración pública.

En este trabajo quiero plantear la cuestión de la doble contingencia —la interdependencia mutua— entre macrodatos, mentes de los individuos y sociedad humana. Para ello, en primer lugar (I) investigaré las diferentes definiciones de macrodatos, incluidas las provocaciones que han generado por parte de los que se mueven en el terreno de la ciencia de los datos. A continuación, indagaré en las soluciones ofrecidas (II) (por ejemplo por el Foro Económico Mundial) para restablecer cierto equilibrio entre las personas y las corporaciones que prácticamente poseen sus datos. También analizaré uno de los principios del gobierno constitucional, la finalidad vinculante (III) —sólidamente enraizado en el principio de legalidad (que no debe confundirse con legalismo). En el terreno del procesamiento de datos personales, este principio no solo está relacionado con los macrodatos que poseen los gobiernos basados en datos, sino también con los modelos de negocio de empresas privadas que monetizan los macrodatos. ¿Hasta qué punto la compartición, la venta y el posterior procesamiento de datos personales más allá del contexto de su recogida son legales y/o éticos? ¿Es compatible el análisis de los macrodatos con la especificación de una finalidad previa? ¿O la desviación de uso es el santo grial de los macrodatos y en la ciencia, en la empresa y en la Administración la extracción de datos transcontextual es lo que aporta valor añadido? ¿Debemos repensar la especificación de la finalidad en tanto que está totalmente en conflicto con la lógica interna de los macrodatos? Terminaré (IV) volviendo a la cuestión de la mutua interdependencia de los macrodatos, las personas individuales y la sociedad humana.

1. La definición de macrodatos: «N = Todos»

Si queremos mejorar el funcionamiento de nuestra web, podemos hacer una investigación de tipo AB,[4] lo cual significa hacer un pequeño cambio en el diseño del sitio A y dirigir a la mitad de los visitantes al sitio A y la otra mitad al sitio B, que es el mismo sitio pero con los pequeños cambios introducidos. Luego registramos todo lo que hacen los visitantes y calculamos cómo encajan las dos versiones del sitio con el comportamiento preferido (digamos el comportamiento de consumo). En lugar de tomar una muestra de los visitantes de nuestra web y llamarlos o mandarles un correo electrónico, solo tenemos que medir todos sus comportamientos y actuar ante nuestros hallazgos. Ya no dependeremos del subgrupo que conteste y no se producirán desviaciones

2. Hildebrandt (2013a).
3. Kephart *et al.* (2003, págs. 41-50).
4. Kohavi et al. (2007); Chopra (2010). Véase también: <http://elem.com/~btilly/effective-ab-testing/>.

debidas a los que dan respuestas políticamente correctas. No tendremos que conformarnos con lo que la gente dice que hizo o que hará, solo tendremos que calcular lo que hizo, lo que hace y lo que probablemente hará. Si nosotros somos los visitantes, haremos de conejillos de indias, pues nadie nos pidió nuestro consentimiento para participar en el experimento, nadie nos pagó por nuestra contribución a la mejora del funcionamiento de la web y, para ser sinceros, todo esto nos pasó desapercibido.

El enigma de los macrodatos: sus implicaciones como agentes de un cambio radical

En su impresionante *Big Data. A Revolution that Will Transform Hw We Live, Work and Think*, Mayer-Schönberger *et al.* describen así los macrodatos:

> «Cosas que se pueden hacer a gran escala que no se pueden hacer a una escala más pequeña».[5]

Es un importante punto de partida, porque evidentemente los macrodatos no solo son una gran bolsa de datos. La dimensión complementaria, que forma parte integrante de la noción de macrodatos, está constituida por las técnicas para extraer modelos relevantes a partir de datos almacenados o incluso en tiempo real. Dichas técnicas se denominan descubrimiento de conocimiento en las bases de datos y ahora la mayoría de ellas están vinculadas al aprendizaje de máquinas. Así se ha definido el descubrimiento de conocimiento en las bases de datos:

> «El proceso no trivial de identificar modelos de datos válidos, nuevos, potencialmente útiles y en última instancia comprensibles».[6]

El aprendizaje de máquinas se ha definido así:

> «Una máquina aprende con relación a una tarea concreta T, un indicador de desempeño I, un tipo de experiencia E, si el sistema realiza con fiabilidad su desempeño I en la tarea T de acuerdo con la experiencia E».[7]

Ambas nociones constituyen el núcleo de *Inteligencia artificial: Un enfoque moderno* (AIMO por sus siglas en inglés),[8] que no debe confundirse con GOFAI (*good old fashioned artificial intelligence* o buena y anticuada inteligencia artificial). Esta, fundamentada en modelos deductivos basados en normas o casos, supone que la inteligencia puede ser modelada e imitada sobre la base de un modelo formal de inteligencia humana. El enfoque moderno, concretamente el aprendizaje de máquinas, se basa en la noción de agencialidad, definida como la capacidad de ser interactiva, autónoma y adaptativa.[9] Sé que muchos autores todavía tienen dudas sobre las máquinas que aprenden, pero a mí me parece más productivo admitir que efectivamente las máquinas aprenden, a gran velocidad y de un modo distinto y a la vez parecido a cómo aprendemos nosotros, lo cual no implica que las máquinas piensen como pensamos nosotros, o sientan como sentimos nosotros. Lo que esto plantea es si nuestros propios procesos de aprendizaje están empezando a cambiar, como consecuencia de tener que interactuar con máquinas que aprenden. ¿Qué repercusión tiene la función de autocompletar en nuestra forma de escribir, cómo influye en la fluidez de nuestro lenguaje, a qué productivos malentendidos da lugar entre amigos que se comunican? Recordemos que Zizek dijo que la comunicación es un malentendido con éxito,[10] refiriéndose con ello a que nunca podemos ver realmente en la mente de los demás, a que nunca podemos estar seguros de si queremos decir lo mismo con las mismas palabras. Zizek nos recuerda que esto no es ningún problema que haya que solucionar, sino una fuente de creatividad. La función de autocompletar, que anuncia perfectamente otros tipos de entornos inteligentes, como la inteligencia ambiental y la internet de las cosas,[11] ¿es una fuente de creatividad, o lo que pretende es la perfección y al final dominará la producción de significado, aunque lo haga desde la perspectiva de una máquina? ¿Internalizaremos la tendencia a la desambiguación inherente al lenguaje de las máquinas, llegaremos a creer que la desambiguación equivale a la comunicación perfecta? En otras palabras, la pregunta no es si las máquinas son capaces de aprender «realmente», sino si nosotros llegaremos a parecernos más a las máquinas porque así nos será más fácil anticiparnos a

5. Mayer-Schönberger, *et al.* (2013, pág. 6).
6. Usama M. Fayyad *et al.* (1996, pág. 41).
7. Mitchell (2006).
8. Russell *et al.* (2010).
9. Floridi *et al.* (2004, págs. 349-379).
10. Zizek (1991, pág. 30).
11. Van Den Berg (2010); Aarts *et al.* (2003).

cómo ellas se anticipan a nosotros. Y si fuera así, ¿hay algo importante en el aprendizaje humano, en el pensamiento humano y en el sentimiento humano que queramos conservar?

Pero sigamos primero los pasos de Mayer-Schönberger y Cukier en su afán por explicar los macrodatos. Su análisis empieza con la noción de «n = todos». En la investigación cuantitativa tradicional, los científicos que querían desvelar regularidades en una población no tenían más remedio que investigar una muestra y confiar en la estadística para hacer una extrapolación de la muestra a la población. Examinar la totalidad de la población era imposible o demasiado caro. Una población puede ser un conjunto de personas, pero también un conjunto de animales, de plantas, de piedras, de paisajes, de células, de moléculas o de cualquier otra cosa, de acontecimientos o de procesos. Así pues, la investigación empieza con una hipótesis que se comprueba sobre la muestra. La muestra consiste en «n» casos de la población relevante, con lo cual, en la medida en que se supone que la muestra representa correctamente a la población, los hallazgos sobre esa muestra serán válidos para la población. Para realizar esta investigación tradicional se debe desarrollar la hipótesis, preparar una muestra representativa, llevar a cabo la investigación y calcular las conclusiones, lo cual requiere tiempo, experiencia en la materia de la que se trate y, según el tipo de pruebas que se deban hacer, quizás también instrumentos caros. Sean cuales sean las conclusiones, seguirán siendo inciertas puesto que no es posible recoger todos los casos relevantes de la población.

«N = todos» significa que la muestra es igual a la población. Implica que, en el caso de los macrodatos, no existe la incertidumbre que genera el salto entre muestra y población. O, para formularlo sin tanta contundencia, significa que el aumento exponencial de «n» reduce substancialmente la incertidumbre. Esto está relacionado con la idea de que la disponibilidad de casi todos los casos de una población dada compensa las posibles inexactitudes. En efecto, Mayer-Schönberger y Cukier afirman que en algunos casos la falta de precisión se corregirá gracias a posteriores registros de más datos. El aumento de conocimiento que es posible gracias a que «n = todos» invita a más «datificación» y promete infinitas oportunidades de extracción de datos a la búsqueda de nuevos modelos pertinentes. Es así porque dichos patrones pueden posibilitar nuevos modelos de negocio o, en el caso de la Administración pública, nuevas argumentaciones para una gobernanza más eficaz y efectiva. Podemos decir que la «datificación» es el proceso de traducir el flujo de vida en bits y bytes discretos, legibles por máquinas, medibles y manipulables[12] La datificación refuerza la ilusión de que «n = todos» porque hace posible una discretización aparentemente ilimitada debido a la reducción de costes y al aumento exponencial de la capacidad informática.

En realidad la actual explosión de datos hace dos cosas. Primero convierte los datos en ruido: la gran cantidad de bits y bytes los hace ilegibles al ojo humano. Segundo, para convertir ese ruido en información o incluso en conocimiento, se han desarrollado y aplicado técnicas informáticas de recuperación de la información. Y, como hemos dicho anteriormente, no son simplemente consultas que recuperan el *input* original, sino que cada vez más se trata de extraer operaciones que recuperan modelos ignorados previamente. Modelos invisibles derivados de las inferencias estadísticas. Estas inferencias pueden denominarse «derivados de datos», tal como sugiere acertadamente Louise Amoore.[13] Derivados de datos que prevén *futuros presentes*, o, en otras palabras, anticipaciones del *presente futuro*. Y, como ha planteado Elena Esposito,[14] esos *futuros presentes modelarán el presente futuro*. Cuanto mejores sean las predicciones (el futuro presente) más personas podrán actuar sobre él y así cambiar la causa y el curso del presente futuro.

Mientras, tal como dicen Mayer-Schönberger y Cukier, la velocidad a la que los nuevos datos están disponibles y con la que pueden extraerse y comprobarse las correlaciones dentro de los conjuntos de datos, parece que succione la vida a partir de la búsqueda de la causalidad, que se está convirtiendo rápidamente en una búsqueda pasada de moda, una búsqueda del «por qué» en una era que funciona mejor con «cómo, cuándo, dependiendo de qué», sin tiempo para determinar las causas que se esconden «detrás» de las

12. Aquí se usa manipulación en el sentido neutro de alterar, editar o mover texto o datos en un ordenador de una forma hábil. La competencia para manipular bits y bytes puede provocar la capacidad de manipular a una persona en el sentido peyorativo de controlar o influir en esa persona o situación con inteligencia y deslealtad y sin escrúpulos.
13. Amoore (2011, págs. 24-43).
14. Esposito (2011).

correlaciones. Porque resulta que en el tiempo que hemos tardado en empezar nuestra investigación, las correlaciones pueden haber sido falsificadas, demostrar que son espurias o simplemente ir seguidas de nuevas correlaciones que «funcionan» mejor. Esta observación se ha hecho muchas veces, especialmente Chris Anderson en su provocador artículo «The End of Theory»[15] publicado en la revista *Wired Magazine*. En efecto, el cambio de la causación a la correlación se basa en una comprensión consecuencialista del significado: para explicar el significado de una correlación no retrocedemos a la causación, sino que miramos hacia adelante, a eso que esta causación puede afectar. Desde la perspectiva del pragmatismo filosófico es fascinante, ya que recuerda una de las llamadas máximas pragmáticas sobre el significado de los conceptos. Esta máxima parece ser particularmente «acertada» para la era de los macrodatos, si sustituimos el concepto por la correlación o modelo:

> «Consideremos qué efectos, que puedan tener concebiblemente repercusiones prácticas, concebimos que tiene el objeto de nuestra concepción. Entonces, nuestra concepción de dichos efectos será la totalidad de nuestra concepción del objeto».[16]

Otra cuestión importante que debaten a continuación Mayer-Schönberger y Cukier es el paso de la experiencia al análisis de datos. Al parecer, no hay ningún campo en el que el análisis de datos no emerja como un agente de cambio radical, que reorienta los procesos de trabajo, las metodologías, los modelos de negocio y las argumentaciones. Un buen ejemplo de ello es la revelación de las prácticas de vigilancia secreta de la Agencia Nacional de Seguridad por un administrador de sistema.[17] Para sobrevivir, tanto la industria como el gobierno deben adaptar sus sistemas de decisiones a las operaciones de procesamiento de datos que progresivamente determinan lo que es posible, lo que permite y a la vez limita nuestra percepción del mundo y plantea la cuestión del libre albedrío. Mayer-Schönberger y Cukier sugieren que estamos al borde de la dictadura de los datos, queriendo decir con ello que seremos incapaces de percibir la realidad si no es a través de las técnicas y tecnologías de los macrodatos. Lo que los autores sugieren es que los datos, la extracción de datos, la comunicación máquina a máquina y los sistemas de decisión informática pronto tomarán el relevo.

N no son todos y todos no son N

Parece como si Mayer-Schönberger y Cukier empezasen con una especie de injustificable tecnooptimismo y terminasen con un tecnopesimismo igualmente injustificable. Me referiré brevemente a las seis provocaciones desarrolladas por Boyd y Crawford,[18] en contra del enigma de los macrodatos. En primer lugar, Boyd y Crawford coinciden en que la investigación automatizada cambia nuestra definición de conocimiento. Los macrodatos no son únicamente una añadidura a la generación de conocimiento o a la gestión del conocimiento. Son agentes de un cambio radical. Implican otra forma de entender lo que se considera conocimiento y crean diferentes fundamentos para los sistemas de decisiones humanos, máquina a máquina e híbridos. Sin embargo, al contrario que Mayer-Schönberger y Cukier, Boyd y Crawford no creen en que «N = todos». Las afirmaciones de objetividad y precisión son engañosas. En mi opinión esto tiene que ver con el hecho de que la cuantificación siempre implica una calificación previa. Para traducir el flujo de la vida en bits y bytes discretos y legibles por máquinas debemos calificar lo que se considera como el mismo tipo de datos, qué realidades encajan con los objetos y atributos en los modelos de datos utilizados para elaborar los macrodatos, lo cual implica interpretación. Como ya se ha dicho,[19] no existe nada como «datos puros», los datos se realizan, igual que los hechos. Como dicen los franceses: *les faits sont faits*. En este sentido, N nunca son Todos, porque el flujo de la vida se puede traducir en datos legibles por máquinas de diferentes formas y escojamos la que escojamos tendrá un enorme impacto en el resultado de las operaciones de extracción de datos.

En consonancia con lo anterior, Boyd y Crawford afirman que los macrodatos no siempre son los mejores datos ni necesariamente los peores; a veces los mejores datos son los microdatos. Aquí tenemos, pues, una afirmación verda-

15. Anderson (2008). Una anterior filósofa de la ciencia, Isabelle Stenger, trazó la forma en que operan las correlaciones en la era de la extracción de datos; de hecho produciendo significado en lugar de desvelando causas o razones existentes previamente; véase Stengers (1997, págs. 62-63).
16. Peirce (1958).
17. Davidson (2013).
18. Boyd *et al.* (2012).
19. Gitelman (2013).

deramente revolucionaria en lo que respecta al enigma de los macrodatos. Creo que para los científicos de datos esto no es nada nuevo ni sorprendente. El tiempo, los conocimientos humanos y el poder de la informática son escasos, al contrario de lo que les gusta proclamar a algunos partidarios de los macrodatos. Para traducir entidades, acontecimientos y procesos en datos discretos es necesario interpretar tales entidades, acontecimientos y procesos, así como una anticipación reiterativa del efecto que producirá un modelo de datos específico, y cómo va a posibilitar y a restringir los resultados de las operaciones de extracción de datos.

Esto está relacionado con la distinción que hacen Boyd y Crawford de las *redes sociales* entre humanos de carne y huesos, por un lado, y las *redes de comportamiento* o grafos sociales observados por el software de los proveedores de servicios, por el otro. Confundir los comportamientos legibles por máquinas con la acción parece formar parte del argumento que justifica el análisis predictivo. Aunque pueda ser una «confusión» muy productiva, en realidad lo que puede desencadenar es una situación en que los comportamientos pongan fin a la acción; ¿a quién le importa que tengamos razones o intenciones, si nuestros comportamientos se corresponden con nuestra disposición genética, si encajan con nuestros grafos sociales en línea o con los datos combinados recogidos por los organismos gubernamentales a lo largo de nuestra vida? Boyd y Crawford sientan así las bases para dos cuestiones éticas:

1. El hecho de que los datos personales estén disponibles públicamente ¿hace que su explotación y monetización sea ética o no?

2. ¿Deberíamos aceptar las nuevas desigualdades creadas por las asimetrías de conocimiento entre sujetos de datos y controladores de datos o no?

   Podemos añadir cuatro aspectos epistemológicos que incorporan una serie de cuestiones éticas quizá no tan obvias pero mucho más generalizadas.

3. Las ciencias naturales y sociales tradicionales empiezan a partir de una reflexión teórica que se comprueba derivando una hipótesis que, a su vez, se puede comprobar con una muestra (lo cual se llama falsacionismo y se supone que da lugar a conocimientos sólidos). ¿Qué significa saltarse la teoría y limitarnos a generar y comprobar hipótesis con «una» población?

4. La ciencia de los datos facilita una serie de técnicas alternativas para detectar modelos en los conjuntos de datos, que a menudo tienen resultados alternativos. Si los proveedores de servicios privados y públicos utilizan mayoritariamente solo un subconjunto de esas técnicas, ¿qué implicará para la robustez de esos resultados y para la medida en que informan la arquitectura de los sistemas informáticos autónomos?

5. En el pasado decíamos: «si las personas definen una situación como real, esa situación es real en sus consecuencias» (teorema de Thomas);[20] ahora hemos de aceptar que «si las máquinas definen una situación como real, esta situación es real en sus consecuencias». ¿Qué repercusiones tendrá ello en cómo se orientan y a qué dan importancia los sistemas de decisiones que dependen del análisis de macrodatos?

6. Si las entrañas de estos sistemas de decisión son opacos para las personas afectadas por sus operaciones, e incluso para las que los gestionan, ¿a dónde irá a parar la democracia y el estado de derecho? ¿Acaso los debidos procesos legales y una administración justa, la elección como consumidores informados y el derecho a no depender de una toma de decisiones invisible están a punto de convertirse en conceptos ilusorios?

2. La gestión de los datos personales en la era de los macrodatos

Datos voluntarios, datos observados y datos inferidos

Volvamos al experimento AB. Los datos se extraen para mejorar la experiencia del usuario, o el funcionamiento de una web, o la rentabilidad de un modelo de negocio, pero el visitante no los facilita. No se rellena ningún formulario ni se hace ninguna pregunta. Son datos observados y suelen consistir en datos relativos al comportamiento. En su proyecto sobre *Rethinking Personal Data*, el Foro Económico Mundial hace poco publicó un informe titulado «Unlocking the Value of Personal Data: From Collection to Usage» (Liberar el valor

20. Cf. Merton (1948, págs. 193-210).

de los datos personales: del almacenamiento al uso).[21] Uno de los aspectos claves de su supuesto nuevo enfoque son las «nuevas formas de comprometer a las personas, de ayudarlas a entender y de darles las herramientas para que puedan hacer una verdadera elección basada en un claro intercambio de valor». Algo muy interesante. Durante mucho tiempo, el valor de los datos personales se ha considerado que estaba relacionado con la personalidad de los individuos. La doctrina legal alemana, por ejemplo, entiende que los derechos fundamentales a la privacidad y a la protección de datos son derechos de la personalidad, refiriéndose con ello a que están relacionados con la dignidad y la autonomía de las personas y que deben considerarse como constitutivos del yo,[22] no como objetos de comercio. En el terreno de las relaciones consumidores-empresas, así como en el del gobierno electrónico, el marco legal de la protección de datos en la Unión Europea se centra en la minimización de datos. Los individuos, como consumidores o como ciudadanos, solo deberían aportar los datos que son necesarios para una finalidad concreta y su uso solo es legal en la medida en que esa finalidad (o una finalidad compatible) se mantiene. Esto también es válido cuando los datos se han facilitado con consentimiento.[23] Sin embargo, cuando ya empezamos a pensar en términos de «un claro intercambio de valor» y hablamos de los datos personales como de «una nueva clase de activos», el valor monetario de los datos personales efectivamente se libera. De hecho, en un informe anterior el Foro Económico Mundial destaca este aspecto y hace hincapié en el oculto potencial de los datos personales como «oportunidades no explotadas para el crecimiento socioeconómico»[24] e insta a un reanudar el debate en torno a la recogida y uso de esos datos teniendo en cuenta la actual monetización de los datos personales. Este renovado debate comienza a partir de una tipología de datos alternativa, que distingue claramente entre datos voluntarios, datos observados y datos inferidos, en lugar de hacerlo entre datos personales y datos no personales. Al parecer, la legislación tradicional sobre protección de datos todavía está centrada en los datos voluntarios, incluso en el caso del acceso de terceras partes a los datos personales o en el de obligación legal de facilitar datos. Los datos voluntarios se definen como «creados y compartidos explícitamente por individuos, por ejemplo, los perfiles de las redes sociales». Quiero añadir que todos los formularios que rellenamos y las tarjetas de crédito que facilitamos conscientemente son datos voluntarios. Fijémonos en que la diferencia entre datos voluntarios y datos observados no consiste en si la persona ha dado su consentimiento ni siquiera tiene que ver con si esos datos se deben considerar datos personales. El procesamiento de ambos tipos de datos puede implicar o incluso requerir consentimiento y ambos pueden ser calificados tanto de datos personales como de datos no personales, por ejemplo, dependiendo del uso de las técnicas de anonimización. El argumento de la investigación AB, la gestión del tráfico, los programas de espionaje de la Agencia Nacional de Seguridad (NSA), las fuerzas del orden público o la detección del fraude raramente se limitan al procesamiento de datos voluntarios. Se basan con mucha más frecuencia en datos observados, normalmente datos comportamentales, medidos por maquinaria de software que extrae, comparte y vende dichos datos observados para conseguir el santo grial de los macrodatos, que son los datos inferidos. Así pues, tenemos datos voluntarios, datos observados y datos inferidos y me atrevo a decir que estos tres diferentes tipos de datos personales y no personales requieren una protección legal diferente.[25] Una cosa es dar el consentimiento para que se compartan los datos de una tarjeta de crédito con la finalidad de comprar algo en línea, o que se comparta una fotografía colgada en Facebok, y otra muy distinta es dar consentimiento para la compartición máquina a máquina de nuestro comportamiento en línea, o de nuestro comportamiento en el transporte público, o de nuestro comportamiento biométrico. Además, los datos inferidos –por ejemplo, los perfiles derivados de la extracción de datos de datos agregados anonimizados– pueden tener un enorme impacto sobre la persona. Si resulta que tres o cuatro puntos de datos relativos a una persona concreta encajan con determinados datos inferidos (con un perfil) –que no tienen por qué ser datos personales y que, por tanto, estarán fuera del alcance de la legislación sobre protección de datos–, es posible que esa persona no consiga el trabajo que desea, que le aumenten la prima de su seguro, que la

21. Foro Económico Mundial (2103).
22. Rouvroy *et al*. (2009); Hornung *et al*. (2009, págs. 84-88).
23. De Hert *et al*. (2006).
24. Foro Económico Mundial (2011).
25. Los datos impersonales no son ni datos personales ni datos no personales en el sentido de que la distinción no es relevante. Los datos anonimizados e inferidos pueden ser datos no personales en lo que se refiere al artículo 2(X) de la Directiva 95/96 CE, pero cuando se aplican a un individuo su impacto puede ser más considerable que el uso de datos voluntarios.

policía decida que su correo electrónico sea revisado o que no pueda acceder a los estudios que ha elegido.

Sin embargo, en el marco de la protección de datos de la Unión Europea, existe el derecho a no estar sujeto a este análisis de perfil por el hecho de estar completamente automatizado.[26] Hay tres importantes excepciones a ese derecho: consentimiento, contrato y obligación legal. Si una de esas excepciones es de aplicación, existe el derecho de transparencia: nos deben comunicar que el análisis de perfil ha determinado la decisión y nos deben informar sobre qué factores se han ponderado. Para que sea un derecho efectivo, quienes utilizan el análisis de datos que tiene un impacto significativo sobre la persona deben ofrecer transparencia del procesado dorsal [*backend system*] (lo que realmente hace el software) de forma clara y comprensible. Y puesto que los usuarios finales no deberían ser completamente dependientes de quienes «poseen» los servidores de datos y las tecnologías de aprendizaje de máquinas, necesitarán sus propias herramientas de transparencia, por ejemplo en el procesado frontal [*frontend*] del sistema. Estas herramientas de transparencia pueden crearse como plataformas, manejables por los consumidores o por terceras partes de confianza, que permitan compartir y extraer datos de consumidores para predecir cómo se van a monetizar probablemente dichos datos o cómo podrán ser usados por los cuerpos policiales. Estas plataformas utilizarían máquinas de inferencia para replicar los perfiles a los que los realizan, o, dicho en otras palabras, para adivinar cómo se nos van a anticipar, para leer cómo vamos probablemente a ser leídos, para anticiparnos a cómo se van a anticipar a nuestras intenciones.[27]

Gestión de datos personales centrada en el usuario

Llegamos ahora a las soluciones propuestas hasta el momento. Tras el fracaso del uso a gran escala de las herramientas de mejora de la privacidad (PET, *privacy enhancing technologies*), seguidas de la noción de privacidad por diseño (PbD, *privacy by design*), considerada como una obligación legal en la propuesta de Regulación General de Protección de Datos (GDPR, *General Data Protection Regulation*) bajo el título Protección de Datos mediante Diseño (DPbD, *data protection by design*), el recién llegado se llama gestión de datos personales (GDP).[28] Esta noción se entiende mejor como un intento de construir arquitecturas y marcos de confianza que hagan posible la minimización de datos y el consentimiento informado, con la esperanza de volver a llevar a los usuarios finales a la ecuación de los ecosistemas de los datos personales. No entraré en tecnicismos, sino que me referiré a la definición que dan Bus *et al*. Nguyen (2013) de GDP sensible al contexto como una aplicación de TIC que:

> «Permite que el individuo controle el acceso a sus datos personales y su uso de tal forma que le da suficiente autonomía para determinar, mantener y desarrollar su identidad como individuo, lo cual incluye presentar aspectos (atributos) de su identidad según el contexto de las transacciones (comunicación, compartición de datos, etc.), así como considerar limitaciones relativas a sus preferencias personales y normas culturales, sociales y legales».[29]

La definición contiene muchos términos vagos, pero a veces ser más específico significa disminuir la protección. Es necesario definir nociones como «identidad», «suficiente», «contexto» y «protección», pero cuanto más definamos, más reduciremos la aplicabilidad del concepto —por lo tanto, si queremos seguir con más definiciones es preferible limitarlas al ámbito operacional—.[30] Es importante recalcar que el término identidad se usa aquí de dos formas distintas. Primero, como una referencia al yo y segundo, en el sentido técnico del conjunto completo de atributos que definen a una persona o en el sentido técnico de uno o más puntos de datos que únicamente identifican a una persona. La idea que hay detrás de la GDP es que el uso de identidades técnicas tiene un impacto sobre el desarrollo de la identidad en el sentido de individualidad, lo cual quiere decir que el uso de identidades en el yo técnico debería restringirse a proteger la construcción de la identidad en el sentido

26. Art. 15 de 12 D 95/46 CE y art. 20 de la GDPR propuesta. En lo relativo a las diferencias, véase Hildebrandt (2012, págs. 41-56).
27. Sobre la antelación a nuestras intenciones en función de la publicidad comportamental, véase McStay (2011, pág. 3).
28. Evidentemente muchas entidades comerciales y organismos gubernamentales están literalmente manejando los datos personales de otras personas. Cuando me refiero a GDP, quiero decir una GDP centrada en el usuario, que facilite una medida de control a la persona propietaria de los datos. Esto puede incluir o no transparencia sobre qué perfiles encajan con los puntos de datos de una persona, incluso si dichos puntos de datos no se consideran como datos personales (porque son simples atributos, no fácilmente vinculables a un identificador único).
29. Bus *et al.* (2013).
30. Lo cual permitirá a las personas discutir una interpretación restrictiva.

de la individualidad.[31] Una forma de conseguir protección contra las «limitaciones irrazonables a la construcción de la identidad propia»[32] sería utilizar la GDP como instrumento para una divulgación mínima, por ejemplo permitiendo la autenticación mediante credenciales basadas en atributos en lugar de la identificación completa. Al revelar solo el atributo necesario para la provisión de un servicio (superar o no una edad determinada, ser hombre o mujer, poseer o no cierto título, tener suficiente crédito) se previene la diseminación innecesaria de datos personales. Una de las desventajas de esta forma de actuar es que entonces el perfil de las personas se construye basándose en sus atributos; aun siendo anónimas pueden convertirse en público objetivo, ya que los macrodatos pueden realizar inferencias a partir del uso de sus atributos. Otra desventaja es que los proveedores de servicios necesitarán pruebas de que el atributo en cuestión es efectivamente un «verdadero» atributo de esa persona, lo cual requerirá un vínculo con una identidad raíz que sea la identidad real o verdadera. Sobre todo en situaciones en las que no se requiere este vínculo puede que aumenten las limitaciones sobre la construcción de identidad, basándose en un análisis de perfiles personalizado.

La monetización de nuestros propios datos personales

Algunas formas de GDP tratan esforzadamente de hacer posible que la propia persona interesada monetice sus datos personales. Ello es interesante porque simplemente se reconoce que los datos personales y otros datos actualmente están monetizados y se acepta que ello tiene consecuencias para las personas con las que están relacionados los datos o para aquellas a las que se aplican los derivativos de datos.[33] En lugar de luchar contra la monetización, la idea es que esta puede crear valor añadido, pero también implica que las personas cuyos datos se utilizan como recurso obtienen una parte de los beneficios (Novotny *et al.*, 2013). Aquí podemos aplicar el principio de maximin de Rawl:[34] quien sea capaz de crear valor añadido tiene derecho a una porción más grande del pastel, siempre y cuando los más desaventajados no vean reducida la suya. Es una forma de lograr justicia distributiva. Se basa en la idea de que si todos compartimos el mismo pastel, la distribución, en principio, debería ser equitativa, mientras que quien agrande el pastel tiene derecho a una porción un poco mayor, con el objetivo de fomentar ese agrandamiento. La restricción moral del principio maximin es que nunca quedarán en desventaja los que tienen las porciones más pequeñas, porque ellos también tienen que mejorar (o como mínimo conservar su porción original). Así pues, la GDP no debería crear asimetrías que, como ciudadanos, como consumidores, empeorasen nuestra situación en lo relativo a nuestra parte del valor monetario respecto al momento anterior a la llegada del análisis de macrodatos, lo cual implicaría nuestra participación en la monetización y nos permitiría obtener una parte de los beneficios. También mantendría el estímulo para inventar aplicaciones para el análisis de macrodatos, ya que quien crea valor añadido consigue una buena parte de los beneficios.

Desde luego, existen otra clase de razones para involucrar a los ciudadanos y consumidores en la creación de valor añadido: mejoraría nuestra autonomía, nos permitiría descubrir cómo podemos influir en los sistemas autónomos de toma de decisiones, compensaría las asimetrías de conocimiento que de otro modo persistirán. En pocas palabras, reinstalaría el sistema de control y equilibrio en el que se basa el estado de derecho, y así reinventaría de algún modo el estado de derecho en la era de los macrodatos. Y no solamente con relación al gobierno, sino también con otros destacados participantes que pueden ser más poderosos que un gobierno. Sin embargo, existen también grandes preocupaciones respecto a los sistemas de GDP que permiten que las personas interesadas moneticen sus datos personales. La principal pregunta es hasta qué punto la GDP puede simplemente ser cooptada por la industria y los organismos gubernamentales para monetizar más nuestros datos, precisamente mediante la implicación de nuestra iniciativa.[35] Por ejemplo, ¿qué sucedería si pudiéramos prever qué comportamientos aumentarán el valor monetario de nuestros datos de comportamiento observados? ¿Qué consecuencias tiene el ser conscientes de que podemos ganar dinero casando esos perfiles inferidos que nos convierten en entidades rentables? ¿Cuándo empezaremos a leer ciertos contenidos solo porque hacen posible la monetización, en

31. Hildebrandt (2008).
32. Agre *et al.* (2001, pág. 7).
33. Hildebrandt *et al*. (2013).
34. Rawls (2005).
35. Tiene que ver con un tipo de autocensura sobre nuestro propio comportamiento, cf. Hutton *et a*l. (1988).

lugar de leer contenidos que no tienen ningún interés para los *brokers* de datos, las redes publicitarias y los especialistas en marketing viral de la era de los macrodatos? ¿Nos influirán los micropagos automatizados que acompañarán nuestros comportamientos legibles por máquinas? ¿Nos va a convertir esto –a nosotros, grupos observados de puntos de datos– en esclavos de los macrodatos? ¿Sí o no?

3. Finalidad vinculante en la era de los macrodatos

Antes de contestar a la pregunta sobre si somos esclavos de los macrodatos, investigaré uno de los principios fundamentales de la legislación sobre protección de datos, sometido ahora a una extrema presión para que ceda ante un modo más flexible de abordar el uso de datos. Esto tiene que ver con el principio de finalidad vinculante, que implica dos reglas interconectadas: (1) el procesamiento de datos personales solo está permitido si hay finalidades explícitas y concretas, y (2) el posterior procesamiento no está permitido si la finalidad ya no se mantiene, salvo si hay otra finalidad que no es incompatible con la original.[36] La finalidad vinculante está íntimamente relacionada con la noción de divulgación mínima o minimización de datos. Cuando la finalidad ya no se mantiene, el procesamiento de datos es ilegal *incluso si hubo consentimiento*. Bajo la actual legislación de la Unión Europea, una persona no puede renunciar a su derecho de conformidad con la limitación de finalidad, porque sea cual sea el motivo de aplicación (artículo 7 de la Directiva de Protección de Datos) serán de aplicación todas las condiciones del procesamiento lícito de datos (artículo 6 de la Directiva de Protección de Datos). En el caso de consentimiento (uno de los motivos del artículo 7), es de aplicación la limitación de la finalidad. Mis preguntas en este trabajo son cómo está relacionado este principio con «n = todos» de los macrodatos, y cómo está relacionado con el tipo de soluciones de la GDP. Más concretamente, cómo está relacionada la finalidad vinculante con la integridad contextual, la destacada propuesta de Helen Nissembaum para repensar la privacidad y la protección de datos, en una era en que la oposición entre la esfera privada y la pública es demasiado burda para que funcione.[37]

Finalidad vinculante y «N = todos»

En la medida en que los macrodatos permiten «hacer a gran escala lo que no se puede hacer a una escala más pequeña», la finalidad vinculante parece que está en conflicto con el argumento de los macrodatos. Los macrodatos quieren *n*, ni más ni menos. La minimización de datos, que es posible, por ejemplo, mediante tecnologías de credenciales basadas en atributos, significa que se reducen los puntos de datos proporcionados (datos voluntarios u observados) a solo los que son necesarios para la finalidad del procesamiento de datos. Por ejemplo, en lugar de dar nuestro DNI para demostrar la edad que tenemos al comprar bebidas alcohólicas, solo tenemos que afirmar que tenemos más de dieciocho años y aportar la prueba necesaria sin revelar otros puntos de datos (es decir, la edad concreta, si somos hombre o mujer, etc.). La minimización de datos también significa que en lugar de permitir que terceras partes rastreen nuestros patrones de navegación por diferentes webs, solo damos permiso a las webs que visitamos para observar los comportamientos necesarios para las operaciones técnicas y funcionales de la web en cuestión. De este modo, el objetivo del procesamiento de datos observados se limita a lo necesario para una experiencia de usuario fluida dentro del campo de nuestra atención. Pero, ¿qué pasa si los datos observados se extraen para servir al interés legítimo de la red publicitaria que nos rastrea y localiza? ¿Qué ocurre si el objetivo explícito y específico de Google Adwords es crear valor añadido tanto para el publicista como para la web que ofrece espacio a la subasta de anuncios basada en publicidad comportamental? ¿Y si los departamentos gubernamentales deciden volver a utilizar datos con el fin de cumplir una nueva obligación legal, ahorrando así a los ciudadanos la aburrida tarea de aportar los mismos datos una vez tras otra? ¿Es posible que los datos recogidos para facturar el consumo de electricidad se utilicen para detectar un fraude, si el fraude está relacionado con la seguridad social? ¿Puede ser que una obligación legal para las empresas de redes inteligentes, que consista en proporcionar datos para la policía, invalide el principio de la finalidad vinculante? El «n = todos» de la detección de fraudes implica un argumento para tener siempre más puntos de datos sobre los ciudadanos, estableciendo correlaciones entre, por ejemplo, consumo eléctrico, movilidad, localización, datos de tráfico

36. Art. 6b, c, d, e de la D 95/46/CE y art. 5b, c, d, e y 6(4) de la propuesta GDPR.
37. Nissenbaum (2010).

de telecomunicaciones y comportamientos fraudulentos. ¿Cuál era la lógica del artículo 6 (4) de la versión primera de la propuesta Regulación General de Protección de Datos?

> «Cuando la finalidad de un posterior procesamiento no es compatible con la finalidad para la que se han recogido los datos personales, el procesamiento debe tener una base legal como mínimo en uno de los motivos a los que se refieren los puntos (a) a (e) del párrafo 1. Ello será de particular aplicación a cualquier cambio en los términos y condiciones generales de un contrato».

¿Querría ello decir que, contrariamente a la legislación actual, uno podría dar su consentimiento para que se vuelvan a utilizar datos para una finalidad incompatible? O ¿sería posible que los gobiernos crearan nuevas «obligaciones legales» para volver a utilizar datos personales, o decir que la reutilización es necesaria y proporcional «para llevar a cabo un trabajo de interés público»? ¿Podría haber sido esta una buena forma de hacer que la protección de datos sea compatible con las posibilidades de crear valor añadido en la era de los macrodatos? ¿Sí? El grupo de trabajo del artículo 29 sobre protección de datos opinó que esto erosiona tanto la minimización de datos como la finalidad vinculante, porque el consentimiento significa muy poco en estos tiempos de sobrecarga cognitiva y las nuevas obligaciones legales no deberían justificar automáticamente la reutilización de datos.[38] Por otro lado, si los macrodatos son interesantes porque generan modelos que de otro modo no hubiéramos podido prever y así hacen posible un uso impredecible, entonces la finalidad vinculante es presuntuosa y parte de una premisa equivocada. No sabemos por adelantado qué uso va a ser posible, y para descubrirlo primero tenemos que extraer los datos, y para saber qué datos son relevantes tenemos que extraer cuantos más datos mejor («n = todos»). El valor de los macrodatos solo será un valor libre si admitimos la novedad que supone el conocimiento inferido y repensamos la finalidad vinculante de acuerdo con el potencial innovador de sus resultados.[38]

Si la GDP es la solución, ¿cuál era el problema?

Investiguémoslo comprobando si la GDP es combinable con la finalidad vinculante. La GDP significa que una persona tiene acceso a dónde, cuándo y cómo se procesan sus datos personales, así como control sobre todo ello. Y, quizás también, sobre la finalidad con la que se procesan sus datos, aunque esto ya no es tan evidente. Si la GDP solo permite garantizar el acceso a los datos y tener información sobre qué entidad los utiliza, no representa ninguna ayuda para hacer posible la limitación de finalidad. Para conseguir dicha limitación de finalidad, es necesario revisar la política sobre privacidad, las condiciones del servicio o el acuerdo de permiso del usuario con antelación y confiar en que quien tiene control sobre los datos se atendrá a ello. Alternativamente, sería necesario que los datos viajaran con una mochila de metadatos que determinasen su uso legítimo, incluidas las condiciones de su eliminación (o sea, aplicabilidad inmediata cuando se den determinadas circunstancias). Lo preferible sería que esta mochila permitiera alguna forma de información máquina a máquina sobre lo que sucede, mientras que la GDP, por ejemplo, podría tener un panel o tablero de control que permitiría la destrucción selectiva de flujos de datos individuales si se descubriera que se hacía de ellos un uso ilegal o, en el caso de procesamiento de datos con consentimiento, si el uso ya no fuera deseable (si se retirase el consentimiento).

Pero si nos fijamos en la estructura de incentivos del análisis de macrodatos, volvemos a enfrentarnos a la noción de «n = todos». La GDP se puede utilizar para facilitar la extracción de cuantos más datos mejor. En particular, si la GDP se diseña de tal forma que nos permite compartir la monetización de nuestros puntos de datos, nuestro tablero de mandos de GDP puede convertirse en una *play station* que genere verdaderos beneficios. Podemos aprender a cambiar los flujos de datos para generar los máximos beneficios, podemos manejar nuestros datos de tal forma que tengamos libre acceso a la mayoría de los servicios. ¿O no? Quizá podemos manipular los flujos de datos –de nuestros datos observados de comportamiento de navegación– de tal modo que encajen con perfiles rentables. Podemos aprender a aprovecharnos del sistema. Y, desde luego, va a haber quien intente piratear nuestros flujos de datos para conseguir acceder ilegalmente a nuestros beneficios. Pero manipular flujos de datos o cambiar patrones de datos comportamentales implica probablemente cambiar nuestro comportamiento. ¿Queremos sinceramente sintonizar nuestros comportamientos legibles por máquinas con motores de explotación comercial (y seremos realmente nosotros los

38. Grupo de trabajo del art. 29, Opinión 03/2013, WP203 sobre el principio de finalidad vinculante, en 38.
39. Massiello *et al.* (2010, págs. 119-124).

que los explotemos cuando los sistemas de GDP se pongan en marcha)? ¿O quienes nos ofrecen dinero nos empujarán sutilmente hacia una actitud sumisa? Sumisa ante cualquier plan que los que están dispuestos a pagar por nuestros datos nos tengan destinado: ¿todo lo que hagamos va a medirse, almacenarse y extraerse para hacernos más influenciables, más sumisos y más previsibles?

Entonces, ¿qué problema nos resuelve la GDP? ¿Deja que nos hagamos la ilusión de controlar nuestros propios puntos de datos y nos invita así a colaborar en nuestra propia subversión? ¿Estamos en nuestras últimas fases antes de convertirnos en un compuesto de puntos de datos mercantilizados y mercantilizables, en un recurso cognitivo para los sistemas informáticos inteligentes que controlan nuestros entornos externos, las infraestructuras decisivas, los sistemas de redistribución de ingresos (llamados impuestos), y cerca de las futuras infraestructuras de impresión en 3D distribuida? ¿La asimetría de conocimientos entre los usuarios finales, por un lado, y las empresas, los ingenieros y los diseñadores de sistemas distribuidos interconectados de procesamiento de datos personales, por el otro, no es ya tan acusada que la mera idea de recuperar el control pone de manifiesto un malentendido fundamental de hasta qué punto ya estamos bajo control (su control)? De momento, propongo que no nos dejemos arrastrar por el tecnopesimismo, aunque volveré a este punto en la última parte de mi trabajo.

Finalidad vinculante e integridad contextual

Antes de eso, investigaré rápidamente el vínculo entre finalidad vinculante e integridad contextual. Ambas nociones tienen que ver con la idea de que las legítimas expectativas en torno a la compartición de datos dependen en parte del papel que desempeña quien controla los datos (¿nuestro médico, la Agencia de Seguridad Nacional o Facebook?), de la finalidad del procesamiento (¿mejorar nuestra salud, prevenir ataques terroristas, aumentar el valor de las acciones?) y del contexto (¿estamos hablando de consejos médicos, de seguridad y protección del comercio?). Otra diferencia es que la finalidad vinculante requiere una previa articulación de finalidades específicas, explícitas y legítimas que vinculen y así restrinjan el posterior procesamiento de los datos personales. La integridad contextual parece, por un lado, *más vaga*, y no necesita una previa determinación de objetivos explícitos y específicos; y, por otro lado, *más precisa* ya que hace que el procesamiento legítimo dependa de las legítimas expectativas ligadas a un contexto concreto, independientemente de articulaciones explícitas de finalidades concretas. Otra diferencia es que el principio de limitación de finalidad es aplicable a los datos personales individuales, mientras que la integridad contextual atañe a los flujos de datos. Lo primero es una obligación legal relativa al procesamiento de datos dentro de las jurisdicciones de la UE, lo segundo es un principio ético, desarrollado en los Estados Unidos la especialista en ética Helen Nissenbaum[40] e incorporado a la Ley de los Derechos de Privacidad del Consumidor (*Consumer Privacy Bill of Rights*) anunciada por la administración de Obama (pero sin poder vinculante y evidentemente no pensada para ser aplicable a la Agencia de Seguridad Nacional).[41] La integridad contextual es especialmente importante en los Estados Unidos debido a la llamada doctrina de la tercera parte, según la cual una vez que los datos se han facilitado a otra parte, se consideran públicos y pueden ser compartidos por esa otra parte, por ejemplo, por los cuerpos policiales, salvo que se haya establecido de otra forma en el contrato pertinente, en la política de privacidad o en el estatuto. Quizá la finalidad vinculante dependa demasiado de la vieja idea de que es posible y razonable tomar una decisión sobre la finalidad del procesamiento de datos con antelación, cuando precisamente el valor añadido de los macrodatos reside en parte en la posibilidad de descubrir nuevos objetivos que pueden crear una situación en la que todos salimos ganando. Sin embargo, la integridad contextual podría restringir la fusión de la base de datos al prohibir el uso de datos en un contexto que no se ajusta al contexto en el que se recogieron. Exactamente igual que en el caso de la finalidad vinculante, esto plantea la cuestión del valor añadido que representan los hallazgos inesperados en la extracción transcontextual de datos: ¿hallazgos que pueden dar lugar a nuevos tratamientos médicos, a una mejor predicción de las intenciones terroristas, a más posibilidades de elección para el consumidor?

La finalidad vinculante, así como la integridad contextual, desafían la argumentación de los macrodatos; ¿podemos tener nuestro pastel y comérnoslo? Para averiguarlo, tendremos que analizar los peligros que entraña la reuti-

40. Nissenbaum (2010).
41. Véase: <http://www.whitehouse.gov/the-press-office/2012/02/23/we-can-t-wait-obama-administration-unveils-blueprint-privacy-bill-rights>.

lización de los mismos datos para una finalidad distinta. Es decir, después del consentimiento o basada en una competencia legal, y también los peligros derivados de la extracción de datos transcontextuales. Estos peligros pueden resumirse como los de una sociedad de la vigilancia, en la que esta vigilancia puede ser ejercida por el gobierno, pero también por las empresas comerciales y en la que las mayores amenazas llegan cuando ambas entidades se sientan para intercambiar los datos que poseen respectivamente. Cuando la Agencia Nacional de Seguridad se introduce en los metadatos de las grandes compañías de internet, y la ciencia y el comercio se nutren de datos abiertos, crece el enigma de «n = todos» y se posibilita el uso de refinados derivativos de datos para lograr estremecedoras ilusiones de omnisciencia. En el siguiente apartado volveré a lo que es realmente el problema de este tipo de ilusión omnisciente de «n = todos». Aquí afirmaría que no, que no siempre podemos tener nuestro pastel y comérnoslo. A veces podemos y a veces no. Depende. Pero si no podemos, tenemos que inventar maneras para que no nos atraigan a compartir nuestros puntos de datos a cambio de satisfacciones momentáneas o recompensas inmediatas. Aunque la economía comportamental es bastante superficial en su análisis de la interacción, confirma el viejo mito de Odiseo y las sirenas. Teniendo en cuenta que según parece preferimos decididamente la satisfacción inmediata a recompensas a más largo plazo, es necesario que levantemos una protección en nuestro entorno que nos ayude a mantener el rumbo. Exigir una determinación previa de la finalidad del procesamiento de datos puede ser una manera de evitar una recogida de datos demasiado entusiasta. Y como disposición legal puede tener más fuerza que un principio ético para atenerse al contexto.

A continuación, volvamos a la cuestión de en quién nos podemos convertir, como individuos y como sociedad, si cada vez dependemos más de las infraestructuras de los macrodatos. Esta reflexión tendría que ayudarnos a valorar si una especificación previa de la finalidad y la limitación del uso son condiciones para que evolucionemos hacia el tipo de personas y el tipo de comunidad que deseamos.

4. Final: ¿Quiénes somos en la era de los macrodatos?

La función de autocompletar: ¿se nos empuja sutilmente hacia la sumisión?

Esto nos lleva otra vez al tema del impacto que tiene el análisis de los macrodatos sobre nuestra mente, nuestro yo y nuestra sociedad. Morozov ha resumido perfectamente los inconvenientes de la imaginada omnisciencia («n = todos»), combinados con la mentalidad solucionista de los frikis de Silicon Valley. En su último libro nos recuerda que:

> «[i]mperfección, ambigüedad, opacidad, desorden y la oportunidad de equivocarse, de pecar, de cometer errores: todo esto forma parte de la libertad humana y cualquier intento encaminado a desenraizarlo también desenraizará esa libertad».[42]

El filósofo del derecho Roger Brownsword también planteó un argumento parecido al referirse al solucionismo tecnolegal:[43] si las tecnologías hacen posible que impongamos la sumisión, es que ya no estamos en el reino de la ley. Para poderse calificar de ley, necesitamos el derecho a desobedecerla, a desafiar su validez ante normas legales superiores y a combatir su aplicación en determinados casos. Los controles y equilibrios del estado de derecho y la división de tareas entre legislador, administración y tribunales implican que la ley apela a la razón y no es inamovible. Cuando las tecnologías imponen la sumisión, estamos en el reino de la administración o la disciplina. Si los macrodatos permiten una anticipación persistente y subliminal a nuestras intenciones, si autocompletan nuestros entornos basándose en preferencias inferidas y posibilitan el tipo de calculada influencia que nos convierte a todos en individuos que respetan la ley, simpáticos, sanos y productivos, deberíamos dar un paso atrás y reconsiderarlo.

Por todo ello no deberíamos dejarnos engañar por los tecnooptimistas que deben de tener sus razones para empujarnos hacia la autocompleción. Y tampoco nos deberíamos dejar impresionar excesivamente por los tecnopesimistas que anuncian el fin del mundo. Aunque es acertada la observación de que por ahora somos presa fácil para los entornos inteligentes que nos tientan para subvertirnos, al tiempo que nos volvemos adictos a la última aplicación, también

42. Morozov (2013, pág. xiv).
43. Brownsword (2005, págs. 1-22).

es acertada la observación de que estamos aprendiendo. Aprendemos, igual que máquinas que ejecutan el código de la última forma de inteligencia artificial. La pregunta es qué aprendemos. Como hemos indicado más arriba, lo esencial es que tenemos que inventar formas de anticiparnos a cómo los entornos inteligentes se anticipan a nosotros, a adivinar cómo nos leen, a averiguar qué *futuros actuales* se infieren y cómo influyen en nuestra *actualidad futura*. Porque aunque las máquinas desarrollen muchos *futuros presentes* (predicciones de nuestros comportamientos futuros) solo habrá un presente futuro. Para reconciliarnos con los entornos inteligentes basados en los datos, tenemos que aprender a ser más inteligentes que ellos, a ir un paso por delante de ellos mientras ellos intentan ir un paso por delante nuestro. Quizá nos parezca un tarea agotadora, pero no tiene por qué serlo. Al contrario, es de dónde venimos y lo que valoramos de la sociedad humana: la reiteración de una anticipación mutua y doble.

¿Una doble contingencia de interacción persona-máquina?

Los sociólogos lo han denominado la «doble contingencia» de la interacción humana.[44] Es lo que constituye tanto la sociedad humana como el yo individual. Es lo que crea incertidumbre en torno a si queremos decir lo mismo cuando utilizamos la misma palabra y nos presiona para buscar técnicas y tecnologías que estabilicen el significado a pesar de su inherente inestabilidad. La doble contingencia significa que yo necesito anticiparme a cómo tú me vas a entender, para tener un significado lógico. Mead lo llamó «adoptar el papel del otro», imaginándonos cómo los demás nos ven y así nacer como persona, desarrollar un sentido del yo.[45] Si le digo a una niña: «tú Sally» señalándola con el dedo, y «yo Mireille» señalándome a mí misma, la niña repetirá: «tú Sally» señalándose a ella, y «yo Mireille», señalándome a mí. La corregiré, pero la niña se sorprenderá y repetirá de nuevo el gesto y el nombre. En el momento en que Sally entiende que para mí ella es «tú», mientras que para sí misma ella es «yo», vuelve a nacer, capaz de adoptar la perspectiva de otra persona. Ahora la niña puede reinventarse, predecirse a sí misma, reflejarse en sí misma, provocar expectativas de sí misma y, finalmente, ser provocativa al violar dichas expectativas. Aquí es donde emergen nuestro sentido del humor y nuestro sentido de la libertad humana. Esta libertad no se puede reducir a la elección del consumidor o a la libertad de las restricciones externas, va ligada a la capacidad de revisión de nosotros mismos y de cambiar el rumbo, basadas en cómo prevemos que serán interpretadas nuestras acciones. Y todo ello es posible adoptando la perspectiva del otro. Esta es la razón por la cual Ricoeu hablaba de *Oneself as Another*.[46] Y también es la razón por la que Zizek destacó que la «comunicación es un malentendido con éxito»;[47] no cualquier malentendido, sino uno que tenga éxito. ¿En qué? En alimentar la ambigüedad productiva del lenguaje humano, que nos permite actuar de común acuerdo a pesar de los recurrentes cambios de significado. En generar nuevos significados a partir de los intersticios de los malentendidos no intencionados, lo cual abre las puertas a nuevas formas de ver las mismas cosas, que de este modo se transforman en otras cosas y nos permiten reflexionar sobre las implicaciones de nuestras acciones, revisándolas con una mirada nueva la de los otros.

¿Qué pasa (1) si resulta que ahora son las máquinas las que se anticipan a nosotros, y qué pasa (2) si nosotros nos empezamos a anticipar a cómo esas máquinas de análisis de perfiles se anticipan a nosotros? Parece que para «entrar en nuestro propio yo» en un entorno inteligente tengamos que adoptar la perspectiva de las máquinas de inferencia. Mientras las máquinas tratan de averiguar quiénes somos, nosotros trataremos de burlar el sistema y decidir por nosotros mismos si efectivamente somos el tipo de persona que las máquinas han calculado. Si la GDP hace posible que adivinemos el valor de nuestros datos personales, la doble contingencia puede reinstalarse. Entonces podremos desarrollar, mediante la tecnología, la capacidad de adivinar cómo nos predicen y aprender cómo anticiparnos a la anticipación de nuestras intenciones. Suena bien. Sin embargo, hay tres excepciones.

La primera tiene que ver con la pregunta de qué pasa si las máquinas se nos anticipan. Tenemos que reconocer que las máquinas solo pueden tener en cuenta datos legibles por máquinas y sus inferencias son contingentes sobre una población que consiste en datos legibles por máquinas. Por lo tanto, N no puede ser Todos, porque no todos pueden

44. Vanderstraeten (1995).
45. Mead *et al.* (1962).
46. Ricoeur (1992).
47. Zizek (1991), nota 10.

ser discretizados. La datificación es al mismo tiempo una multiplicación de la realidad, una virtualización en el sentido de Deleuze[48] y una reducción de la realidad, porque necesariamente traduce el flujo de la vida en puntos de datos discretos. Por cierto, lo mismo vale para el lenguaje escrito. Pero el lenguaje escrito es visible para aquellos que han aprendido a leer y escribir, mientras que el lenguaje informático es el conocimiento secreto de los expertos.

La segunda concierne a la pregunta de qué nos pasaría si empezáramos a anticiparnos a esas máquinas. ¿Acaso al averiguar cómo «piensan» esas máquinas nos iríamos pareciendo a ellas y perderíamos parte de la ambigüedad inherente al uso del lenguaje hablado y escrito? ¿Tendrá razón Brian Christian cuando dice que en lugar de ser las máquinas las que se van pareciendo a nosotros, somos nosotros los que nos estamos volviendo como máquinas?[49] ¿Estará en lo cierto Maryanne Wolf al decir que la morfología y el comportamiento de nuestro cerebro cambiarán y que nos tendremos que preguntar qué hay que conservar, destacando que no podemos dar por sentado que nuestro cerebro se adapte sin perder lo que ha desarrollado a lo largo de nuestra evolución como animales lectores?[50]

La tercera se refiere a la transparencia que reinstala la doble contingencia. El modelo de la GDP, descrito anteriormente, puede posibilitar la transparencia intuitiva a través de la monetización. La introducción de un *tertium comparationis* en forma de dinero, de precio, podría darnos la capacidad de prever cómo casan nuestros propios puntos de datos con los modelos comportamentales inferidos. Pero, como hemos visto antes, esto podría crear perversos incentivos. Sin embargo la pregunta es si tenemos alternativas.

Hoy, la transparencia que se ofrece siempre que se requiere consentimiento previo informado crea un «desbordamiento del búfer»: la cantidad de información que implica inunda nuestra limitada racionalidad, lo cual permite la manipulación, por lo que escapa de nuestra atención. A pesar de que hay quien aplaude la iluminación de las *idees claires et distinctes* de Descartes, otros pueden opinar que generan sobreexposición, lo cual sugiere erróneamente la posibilidad de luz sin sombras. La metáfora del desbordamiento del búfer sugiere en realidad que puede que necesitemos una iluminación selectiva y que tengamos una acuciante necesidad de sombras. Por lo tanto, la pregunta más interesante será qué debería quedar en un primer plano y dónde necesitamos oscuridad. En la pintura renacentista las técnicas del claroscuro, el *chiaroscuro*, el *Helldunkel*, se inventaron y aplicaron para sugerir profundidad y para iluminar lo que se quería destacar. Jugando con la luz y la sombra, la pintura atrae la atención del espectador y crea la peculiar experiencia de vernos arrastrados dentro de la obra —como si estuviéramos de pie en la oscuridad, atraídos por la luz—. El análisis de los macrodatos nos invita a reinventar algo parecido al claroscuro, una medida de transparencia que nos permite prever lo que nos espera. Esto nos permitiría enfrentarnos a cómo se nos está agrupando, correlacionando, enmarcando y leyendo, lo cual facilitaría los prerrequisitos para el debido proceso. Finalmente, con ello podríamos especular con nuestras sombras digitales y adquirir el grado de fluidez que hemos aprendido a adquirir con la lectura y la escritura.[51]

Termino mi trabajo haciendo referencia al enigma de la esfinge en la tapa de un libro que acabo de coeditar con Katja de Vries. En ella se ve a Edipo en el claro del claroscuro. Edipo destaca, fuerte y voluntarioso, y mira un poco impaciente. La esfinge está en la sombra de una cueva, potencialmente irritada porque un intruso ha conseguido por fin desvelar su enigma. Sin embargo, aunque Edipo haya resuelto el enigma, no puede evitar la fragilidad fundamental que predice la esfinge. A pesar de que la monetización de los puntos de datos personales pueda ayudar a reinventar una nueva versión de la doble contingencia que constituye nuestro mundo, no puede resolver la incertidumbre fundamental que sostiene. En realidad, necesitamos herramientas de transparencia que nos ayuden a reinstalar dicha incertidumbre, en lugar del exceso de determinación que de otro modo podría hacer posible la monetización.

48. Deleuze (1994); Lévy (1998).
49. Christian (2011).
50. Wolf (2008).
51. Este párrafo y el siguente se basan en Hildebrandt (2013b, págs. 238-9 y 241).

Bibliografía

AARTS, Emile; MARZANO; Stefano (2005). *The New Everyday. Views on Ambient Intelligence*. Rotterdam: 010, 2003. ITU. «The Internet of Things». Ginebra: International Telecommunications Union (ITU).

AGRE, Philip E.; ROTENBERG, Marc (2001). *Technology and Privacy: The New Landscape*. Cambridge, Massachusetts: MIT, pág. 7.

AMOORE, Louise (2011). «Data Derivatives on the Emergence of a Security Risk Calculus for Our Times». *Theory, Culture & Society,* vol. 28, n.º 6, págs. 24-43.

ANDERSON, Chris (2008). «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete». *Wired Magazine*, vol. 16, n.º 7.

BOYD, Danah; CRAWFORD, Kate (2012). «Critical Questions for Big Data». *Information, Communication & Society*, vol. 5, n.º 4, págs. 662-679.
http://dx.doi.org/10.1080/1369118X.2012.678878

BROWNSWORD, Roger (2005). «Code, control, and choice: why East is East and West is West». *Legal Studies*, vol. 25. n.º 1, págs. 1-22.
http://dx.doi.org/10.1111/j.1748-121X.2005.tb00268.x

BUS, Jacques; NGUYEN, Carolyn (2013). «Personal Data Management – A Structured Discussion». En: Mireille HILDEBRANDT, Kieron O'HARA, Michael WAIDNER (eds.). *The Value of Personal Data. Digital Enlightenment Yearbook 2013*. Ámsterdam: IOS Press.

CHOPRA, Paras (2010). «The Ultimate Guide To A/B Testing». *Smashing Magazine*. <http://www.smashingmagazine.com/2010/06/24/the-ultimate-guide-to-a-b-testing/>

CHRISTIAN, Brian (2011). *The Most Human Human. What Talking with Computers Teaches Us About What It Means to Be Alive*. Nueva York: Doubleday.

DAVIDSON, Joe (2013). «NSA to cut 90 percent of systems administrators. Federal Eye». En: *Washington Post*. 13 de agosto de 2013. <http://www.washingtonpost.com/blogs/federal-eye/wp/2013/08/13/nsa-to-cut-90-percent-of-systems-administrators/>.

DE HERT, Paul; GUTWIRTH, Serge (2006). «Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power». En: Erik CLAES, Antony DUFF, Serge GUTWIRTH (eds.). *Privacy and the Criminal Law*. Amberes, Oxford: Intersentia.

DELEUZE, Gilles (1994). *Difference and repetition*. Nueva York: Columbia University Press.

ESPOSITO, Elena (2011). *The Future of Futures: The Time of Money in Financing and Society*. Edward Elgar.

GITELMAN, Lisa (ed.) (2013). *'Raw data' is an oxymoron*. Cambridge, Massachusetts – Londres, Inglaterra: MIT Press.

FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic [*et al.*] (1996). *Advances in Knowledge Discovery and Data Mining*. Meno Park, CA - Cambridge, MA – Londres, Inglaterra: AAAI Press / MIT Press, pág. 41.

FLORIDI, Luciano; SANDERS, J. W. (2004). «On the Morality of Artificial Agents». *Minds and Machines,* vol. 14, n.º 3, págs. 349-379.
http://dx.doi.org/10.1023/B:MIND.0000035461.63578.9d

FORO ECONÓMICO MUNDIAL (2011). *Rethinking Personal Data: Strengthening Trust*.
<http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf>

FORO ECONÓMICO MUNDIAL (2013). *Rethinking Personal Data: Unlocking the Value of Personal Data: From Collection to Usage*. <http://www.weforum.org/reports/unlocking-value-personal-data-collection-usage>

HILDEBRANDT, Mireille (2013a). *Legal Protection by Design in the Smart Grid*. Informe, encargado por Smart Energy Collective. Nijmegen/Groningen. <http://works.bepress.com/mireille_hildebrandt/42/>

HILDEBRANDT, Mireille (2013b). «Profile Transparency by Design: Re-enabling Double Contingency». En: M. HILDEBRANDT, E. DE VRIES (eds.). *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*. Abingdon: Routledge, págs. 238-9.

HILDEBRANDT, Mireille (2012). «The Dawn of a Critical Transparency Right for the Profiling Era». En: *Digital Enlightenment Yearbook 2012*. Ámsterdam: IOS Press, págs. 41–56.

HILDEBRANDT, M. (2008). «Profiling and the identity of the European citizen». En: M. HILDEBRANDT, S. GUTWIRTH (eds.). *Profiling the European citizen. Cross-discplinary perspectives*. Dordrecht: Springer. http://dx.doi.org/10.1007/978-1-4020-6914-7

HILDEBRANDT, M.; O'HARA, K.; WAIDNER, M. (2013). *The Value of Personal Data. Dig*ital Enlightenment Yearbook 2013. Amsterdam: IOS.

HORNUNG, G.; SCHNABEL, Ch. (2009). «Data protection in Germany I: The Population census decision and the right to informational self-determination». *Computer Law & Security Reports*, 1, págs. 84–88. http://dx.doi.org/10.1016/j.clsr.2008.11.002

HUTTON, Patrick H.; GUTMAN, Huck; MARTIN, Luther H. [*et al.*] (1988). *Technologies of the self: a seminar with Michel Foucault*. Amherst: University of Massachusetts Press.

KEPHART, Jeffrey O.; CHESS, David M. (2003). «The Vision of Autonomic Computing». *Computer*. Enero, págs. 41–50. http://dx.doi.org/10.1109/MC.2003.1160055

KOHAVI, Ron; HENNE, Randal M.; SOMMERFIELD, Dan (2007). «Practical guide to controlled experiments on the web: listen to your customers not to the hippo». In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. Nueva York, NY: ACM, págs. 959–967. http://dx.doi.org/10.1145/1281192.1281295

LÉVY, Pierre (1998). *Becoming Virtual. Reality in the Digital Age*. Nueva York y Londres: Plenum Trade.

LUHMANN, Niklas (1995). *Social Systems*. Stanford: Stanford University Press.

MASSIELLO, Betsy; WHITTEN, Alma (2010). «Engineering Privacy in an Age of Information Abundance». En: *Intelligent Information Privacy Management*. AAAI, págs. 119–124.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt, pág. 6.

MCSTAY, Andrew (2011). *The mood of information: a critique of online behavioural advertising*. Nueva York: Continuum, pág. 3.

MEAD, George Herbert; MORRIS, Charles William (1962). *Mind, self, and society from the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

MERTON, Robert K. (1948). «The Self-Fulfilling Prophecy». *The Antioch Review*, vol. 8, n.º 2, págs. 193–210. http://dx.doi.org/10.2307/4609267

MITCHELL, Tom M. (2006). «Introduction». *The Discipline of Machine Learning*. Carnegie Mellon University, School of Computer Science. http://www-cgi.cs.cmu.edu/~tom/pubs/MachineLearningTR.pdf.

MOROZOV, Evgeny (2013). *To save everything, click here: the folly of technological solutionism*. Nueva York: Public Affairs, pág. xiv.

NOVOTNY, Alexander; SPIEKERMANN, Sarah (2013) «Personal Information Markets and Privacy: A New Model to Solve the Controversy». *WI'2013*. Leipzig. <http://ssrn.com/abstract=2148885> http://dx.doi.org/10.2139/ssrn.2148885

NISSENBAUM, Helen (2010). *Privacy in context: technology, policy, and the integrity of social life*. Stanford, California: Stanford Law Books.

PEIRCE, Charles Saunders (1958). *Selected Writings, Edited with an Introduction and notes by Philip P. Wiener*. Nueva York: Dover.

RAWLS, John (2005). *A theory of justice*. Cambridge, Massachusetts: Belknap Press.

RICOEUR, Paul (1992). *Oneself as Another*. Chicago: The University of Chicago Press.

ROUVROY, A.; POULLET, Yves (2009). «The right to informational self-determination and the value of self-development. Reassessing the importance of privacy for democracy». En: S. GUTWIRTH, P. DE HERT, Y. POULLET (eds.) *Reinventing Data Protection*. Dordrecht: Springer.

RUSSELL, Stuart J.; NORVIG, Peter; DAVIS, Ernest (2010). *Artificial intelligence: a modern approach*. Upper Saddle River, New Jersey: Prentice Hall.

STENGERS, Isabelle (1997). *Sciences et pouvoirs*. París: La Découverte, págs. 62-63.

VAN DEN BERG, Bibi (2010). *The Situated Self: Identity in a world of Ambient Intelligence*. Nijmegen: Wolf Legal Publishers.

VANDERSTRAETEN, R. (2007). «Parsons, Luhmann and the Theorem of Double Contingency». *Journal of Classical Sociology*, vol. 2, n.º 1, págs. 77-92.

WOLF, Maryanne (2008). *Proust and the Squid: The Story and Science of the Reading Brain*. Icon Books Ltd.

ZIZEK, Slavoj (1991). *Looking awry: an introduction to Jacques Lacan through popular culture*. Cambridge, Massachusetts: MIT Press, pág. 30.

Cita recomendada

HILDEBRANDT, Mireille (2013). «Esclavos de los macrodatos. ¿O no?». *IDP. Revista de Internet, Derecho y Política.* Número 17, pág. 7-26. UOC. [Fecha de consulta: dd/mm/aa]
http://journals.uoc.edu/index.php/idp/article/view/n17-hildebrandt/n17-hildebrandt-es
http://dx.doi.org/10.7238/idp.v0i17.197

Sobre la autora

Mireille Hildebrandt
hildebrandt@law.eur.nl
Cátedra de Entornos Inteligentes, Protección de Datos y Estado de Derecho
Institute for Computing and Information Sciences (ICIS)
Universidad Radboud de Nimega

Mireille Hildebrandt empezó su andadura académica en la antropología cultural para sustituirla luego por el derecho. Obtuvo su grado en Derecho en la Universidad de Leyden, en los Países Bajos, y en su tesis doctoral, que leyó en la Erasmus University Rotterdam y cuyo tema fue la filosofía del derecho penal, integró la antropología jurídica y la historia del derecho para desarrollar una fenomenología hermenéutica del castigo.

Actualmente ocupa la cátedra de Entornos Inteligentes, Protección de Datos y Estado de Derecho en el Institute for Computing and Information Sciences (ICIS) en la Universidad Radboud de Nimega. Es profesora asociada de Jurisprudencia en la Erasmus School of Law y desde 2002 trabaja en comisión de servicios a tiempo parcial en el Centre for Law Science Technology and Society (LSTS) en la Vrije Universiteit Brussels. Su línea principal de investigación se centra en la relación entre la infraestructura sociotécnica emergente (internet, web 2.0, inteligencia ambiental) y la autonomía de la persona humana supuesta y a la vez producida por la democracia constitucional. Junto con Serge Gutwirth ha publicado *Profiling the European Citizen* (Springer 2008) y con Antoinette Rouvroy *Law, Human Agency and Autonomic Computing* (Routledge 2011).
Su web personal es: <http://works.bepress.com/mireille_hildebrandt/>

Faculty of Science
University of Nijmegen
Postbus 9010
6500GL Nijmegen
The Netherlands