



編者按:7月24日,在匹茲堡華人爲中華人民共和國駐紐約總領事孫國祥大使到訪舉辦的晚餐活動上,我認識了旅美科技協會的主席塗子沛先生,當日在餐會上塗先生也是唯一一位向孫國祥大使提出建議的華人:他認爲領事館工作效率和態度不盡人意,希望領事館能有改進。塗先生給我留下了深刻影響。後來我得知他撰寫的一部大作《大數據》,寫的是科技浪潮對國家社會政治、經濟、文化等方方面面的影響,非常精彩,國內已有四家媒體做了專版介紹。這裡本報將他的故事和他的書介紹給大家望共同分享!

塗子沛:

使用數據 是一個國家的 生存技能

塗子沛旅美華人,1996年畢業于華中科技大學計算機系。2006年辭去公職赴美留學,就讀于卡內基梅隆大學,獲信息技術碩士、公共管理碩士學位。現居匹茲堡,擔任KIT Solutions軟件公司數據中心主任、旅美科技協會匹茲堡分會主席。日前,其以探討科技背景下人與社會關係的專著《大數據》由廣西師範大學出版社出版。

先鋒語錄

★美國擊斃本·拉丹,其實就是數據在背後起作用。
★我們需要的是學習,怕的就是明明知道這條路可走,卻偏要繼續摸着石頭過河。
★真正的信息社會是一個信息自由流動而不受操縱的社會,這個社會的前提是人的自由和平等。

“文明是將一個人從一群人當中解放出來的過程。”電話那端,年近不惑的塗子沛用艾茵·蘭德的這句話,強調文明譜系中隱私權的坐標點。

33歲以前,塗子沛還是一名體制內的有爲青年。從重點中學,到名牌大學,再到廣東邊防大院,在這條中國式的精英道路上,他不允許自己有絲毫差池。

“十年中,我開發了中國的第一個反偷渡管理軟件,多次在全國性的新聞發佈會上做展示,擔任緝私艇指揮官,率隊在海上航行,撰寫多篇論文並彙編成冊……在廣東邊防工作的8年間被授予三等功兩次、嘉獎一次,2000年單位甚至爲我申報了二等功。”塗子沛曾在博客中這樣寫道。

但2006年8月的一天,揣着卡內基梅隆大學的碩士錄取通知書,他毅然逃離了這一眼望得見頭的“教科書式”生活。

經過6年打拼,塗子沛已是美國某軟件公司數據中心的主任,並爲國內多家媒體撰寫“美國觀察”——《微博能不能預測總統大選》、《微學位將掀起教育領域的革命》、《政府能否對公民進行網上監控》……他的專欄大多圍繞老本行“信息管理”展開。但正如他在新書《大數據》的扉頁上所題:“一個真正的信息社會,首先是一個公民社會。”

新書中,他系統梳理了美國半個多世紀信息公開的歷程以及信息技術創新的歷史,通過用豐富的圖表、注腳和事例試圖闡釋數據創新給公民、政府和企業所帶來的挑戰和變革。“儘管美國是該書的主體,卻又處處反觀中國當下的現實。”塗特意引用胡適和黃仁宇的話,前者批評中國習慣于“差不多先生”,而後者則認爲中國不懂用數據管理治國。如書中舉例,中國機動車輛相當于美國的三分之一,據官方數據,死于交通事故的人數卻是美國的兩倍。塗以此帶出美國循“數”管理的先進經驗。

談及寫作過程,他告訴記者,寫作這本書的八個多月里,每一章每一節,他都在想“這是不是中國社會所需要的?我有沒有傳達新的信息?”甚至連每章節開頭的相關名人名言,“找的也是華語世界沒有的。”

去年底,《大數據》書稿被國內多家出版社相中,看完樣稿後,廣西師大出版社的總編輯劉瑞琳給塗子沛復去短短幾個字:“我認爲這是一本中國需要的書。”最終,塗子沛被這句話所打動,決定將書稿交予廣西師大出版社出版。

除了上帝,任何人都必須用數據來說話。

現代管理學之父德魯克有言,預測未來最好的方法,就是去創造未來。如今,人類文明正在用大數據技術踐行這一預言:誰率先具備從各種各樣類型的數據中快速獲得有價值信息的能力,誰就是贏家。

毫無疑問,我們正在一個真正的大數據時代:數據作爲新一輪信息戰的主角,將創造無限商機,既便利又危及着每個人的生活,Google、百度之類搜索服務,不再有立足之地,而數據技術變革,更是能推動政府信息公開、透明和社會公正……這就是爲什麼,塗子沛的《大數據》一書的出版恰逢其時。但如果認爲這只是一本屬於IT界技術人士關起門來討論的專業書籍,那你就大錯特錯了——公佈官員財產美國是怎麼做的,美國能讓少部分人腐敗起來嗎,美國式上訪是怎麼回事,憑什麼美國礦難那麼少,全民醫改美國做得到嗎,美國總統大選有什麼利器才能贏,下一輪全球洗牌我們的“世界工廠”會被淘汰嗎……在《大數據》中,看不到晦澀難懂的公式,也不會被密密麻麻的數字或生僻理論給嚇住,正相反,作爲一本由數據管理專家寫就的書,《大數據》太不像一本講數據的書,它充滿着一個人對社會趨勢的觀察和思考,同時是回到歷史的野心,甚至有着“我以我血薦軒轅”的家國情懷。

美國是全書主體,但又處處反觀中國當下的現實。回望中國,胡適批評“差不多先生”,黃仁宇求索“數目字管理”,作者從太平洋對面看到中美兩國的差距,深知中國缺少什麼、需要什么,故將十多年觀察、思索所得,淘洗成這一本書。“這本書以信息化爲背景,講述作爲一波浪潮的大數據,但我想說的不僅僅是技術。”塗子沛說。

史學大家、匹茲堡大學歷史系榮譽講座教授許倬雲,專門爲《大數據》作序:“我們要對塗子沛先生致敬與致謝,因爲他爲華文世界提出一個重要的話題。”

美國擊斃本·拉丹背後的大數據

記者(Q):有人說,《大數據》是一本很及時的書,你是怎麼想到去寫它的?

塗子沛(A):儘管是在國外寫作,但我寫的東西無一例外地注重中國社會應該借鑒的東西,中國需要什么我就寫什麼。所以他們評論這本書:這里有中國的問題,這里有中國的財富,這里有中國的鄉情。看到這句話的時候我眼淚都流出來了。這就是我的心里話,你問我爲什麼要寫這本書?其實投入跟產出是不成比例的,我們這種在美國工作的數據工作者,一個小時至少50美元,一篇專欄都要用一兩天來醞釀,然後才能寫出來。但我還是願意去做這個東西,就像這個社會總需要一個客觀理性的觀察者。

Q:怎樣定義大數據時代的現實意義?

A:大數據時代所涉及的問題主要是數據收集、開放和使用的問題,在美國,數據開放不僅僅在政府領域實施,還推進到商業和私人領域。現在越來越多的新的應用出現,比如你要購買一件商品,通過掃描條形碼就能夠獲得多家商場同一商品的價格信息,然後進行對比。大數據時代,這樣的應用會越來越多,在美國數據開放已經不只是滿足知情權的問題,而是推動經濟和社會創新的一個有利因素。

Q:爲什麼說中國社會需要大數據?

A:在世界已經起步的情況下,我必然會關心中國怎麼辦,首先我們不可能迴避這個時代,但中國所面臨的問題和挑戰也很大。這個時代是一個知識便捷地不斷擴大的時代。因爲有了數據,你就可以做分析,比如說前些天歐洲杯期間,每天晚上十一點到十二點,網上商店的交易量就一路飆升;也就是通過數據發現男人一看球女人就購物。通過這組數據,我們可以知道,購物是什麼時候?她們

買了什麼?是珠寶還是首飾?這都是商機。這在小數據時代是不可能發現的,因爲個人的行爲在小數據時代是無序的、零散的。但大數據時代,人的行爲被統計之後在數據上會呈現一種穩定的規律性,你可以去研究。

再例如,美國擊斃本·拉丹,其實就是數據在背後起作用,說白了,反恐工作就是海量的數據收集,關注那些沒有固定住所但是賬面上又有很多現金的、老打國際長途的、老購買一些炸藥的、常旅行的人,然後做數據挖掘和提煉,再通過監控然後一個一個排除。

信息社會首先是一個公民社會

Q:在走向大數據時代的國家中,包括中國嗎?

A:我認爲中國社會現在是被推進了一個信息時代,但是我們還不是一個真正完善的信息社會。

真正的信息社會是一個信息自由流動而不受操縱的社會,這個社會的前提是人的自由和平等。所以我覺得落腳點就在這裡:一個信息社會首先是一個公民社會。

Q:被推着進入一個大數據時代的中國,是否有自己的主動性?

A:當然有,我們看到很多好的東西,比如前段時間食品安全出現問題了,就有人做了一個網站,把所有有問題的食品都收集到一起然後再集體發佈,這就是民間的努力。它的信息也不完整,也有很多很多問題,但它是一個很好的嘗試跟努力,這也說明我們有後來優勢。很多人說現在中國崛起,當然中國崛起是誰都能看得到的,但另一方面也必須要看到,在面對大數據時代的準備上,我們的旗幟才剛剛離地。

西方有太多東西值得我們學了。我們不是一味說西方好,而是首先站在一個理性的角度、客觀的角度,把它的好東西提煉出來。你看美國收集數據史上的里程碑:1973年引進了最早的數據體,中國是2006年,相差三十多年;1940年羅斯福總統已經進行民意調查,2002年我們有了自己的民意調查中心;1962年美國有了海浪監製計劃,開始收集環境數據,我們2010年開始有這方面的零星研究。所以說我們需要的是學習,怕的就是明明知道這條路可走,卻偏要繼續摸着石頭過河。

Q:所以你覺得中國的差距在于……

A:我們本來是一個數據大國,但我們的新聞每一年新增的產業數據是250拍,美國是3500拍,日本則是400拍,所以我們收集數據做得很差。這意味着,第一,我們的很多商務系統信息化不夠;第二,我們沒有民意的數據,不注重老百姓的觀念,不去做民意調查。但美國聯邦政府最重要的

一塊就是民意數據,它的一個民意調查可以做40年,跟踪一個人從小時候到青年時代,一直不會停止問卷調查,通過這個來統計社會觀點的變遷,無論對社會還是科學來說,都是一個巨大的積累。那對數據使用呢?我們就更差了。我們不尊重數據,理性精神不足,這是中國人的傳統。但在未來社會,使用數據不是一種文化也不是一種習慣,而是一個國家必須掌握的生存技能。大數據時代最關鍵是開放的時代,只有你開放了,數據才會有價值,也只有你開放了自己的,你才能共享別人的,目前全世界有三十多個國家開放數據,中國不在此列。但凡美國政府、英國政府的網站,都可以看到它的財政開支細目,甚至中國人在美國也可以提起信息公開的請求,例如你要求奧巴馬某次出行帶了多少隨從花了多少錢,他會把詳細的數據信息公開給你,供你使用,包括自由發佈。信息社會最怕的就是這種比較,所謂數據治國,就是讓一切回歸可量化的、透明的、易流通的狀態。

美國人爲什麼沒有身份證

Q:人時刻處在數據的收集、監控過程之中,隱私社會不會變成一個更加棘手的問題?

A:直到現在,“隱私”仍然是美國社會很容易引發討論的社會問題,比如,Facebook甚至設立了首席隱私官專門應對隱私問題。爲什麼?就因爲它的隱私政策一旦出了問題,就會引起整個運營上的震動。有幾個經典的例子:有人去找工作,雇主需要應聘者提供Facebook賬號,對其個人行爲進行考察。前幾個月,北卡有個很著名的案例,有人因爲雇主看了Facebook信息而沒有獲得這個工作,於是對雇主進行起訴,這個官司後來打贏了,北卡於是立法

——雇主不得對個人隱私進行網上監控——因爲他們認爲,公民隱私是自由的保障,這種保障,意味着個人可以和其他人的干擾對抗,也意味着個人隱私有不受新聞媒體、政府權力侵犯的保障。公民通過隱私來確定公共生活與自我的界限。

Q:喬治·奧威爾的《1984》中刻畫的老大哥的世界,是否也算是一種公權力通過數據對隱私的控制?

A:對,其實大數據在政府範圍內討論,的確存在老大哥的問題。這就需要公民社會給政府一個制約。例如美國社會沒有身份證,每個人只有一個ID號作爲辨識依據,但幾乎每年,關於辦不辦身份證、是否建立中央數據銀行,政府和老百姓之間都會吵得不可開交,每隔幾年都會打官司。身份證的確很有用,不僅是政府管控,經濟領域也很需要,一旦建立整個數據的流動就更有效了,對通用信息的發展也是有利的,但是老百姓就是不同意,國會也不同意。他們反對的依據就是1974年通過的《隱私法》。所以儘管每年大選都有人提身份證的事兒,但到現在都沒有通過。技術從來不具備自我道德的淨化和對危險的自我警惕,所以我爲什麼一再強調,一個信息社會,首先是公民社會,沒有這種完善的公民精神和制約權力,大數據時代就是一個老大哥的世界。

