

Turning Their Data into Your Money (and vice versa)

Robert M. (r0ml) Lefkowitz
r0ml@1010data.com

1



Sophisticated Wall Street analysts working with 3-D goggles and parallel processors.

The answer: 10 day moving averages in pairs trading.

Today, we're talking about sentiment analysis and social graphs.

In that vein, we're going to look at the simple stuff.

2



Transactions

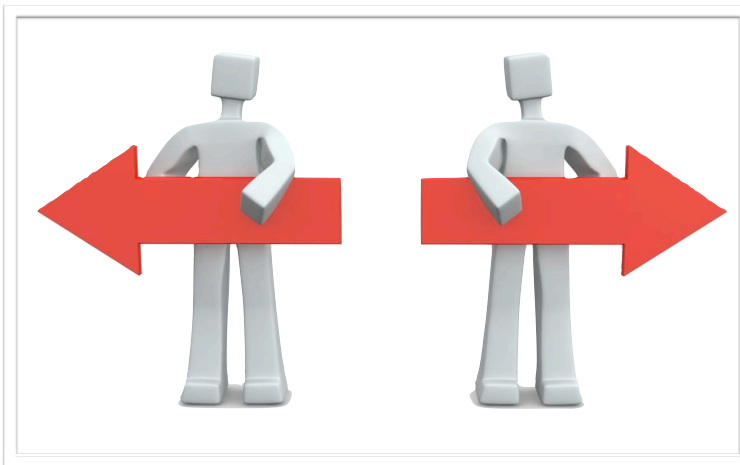
3

Buyers and sellers,
Senders and receivers,
Givers and takers.

Who, what, where, and how much.

Simple relational database stuff.

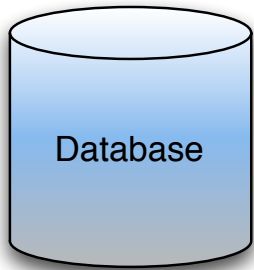
We're not going to talk about



Expenses – Revenues

4

The stuff you sell (revenue)
The stuff you buy (expenses)



5



You collect data about your transaction,
you analyze your data, you try to increase
sales or decrease expenses.

Recommendation engines.

6

Middleman

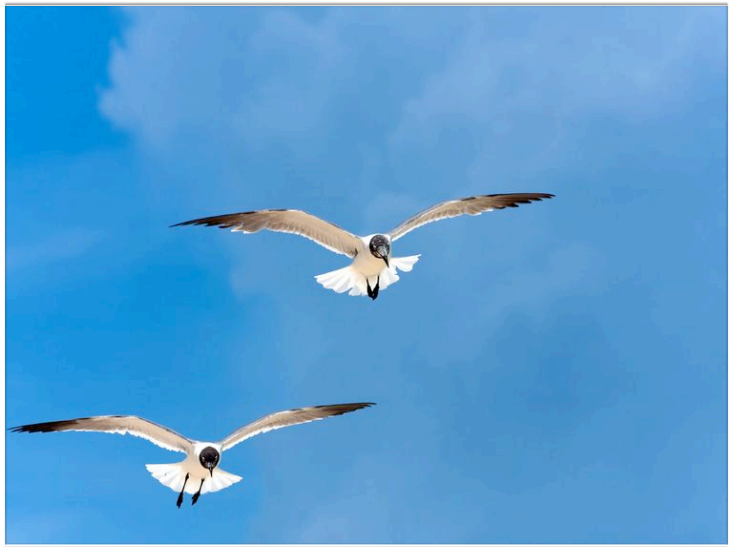


The internet is the greatest boon to the middleman ever created.

It's not about disintermediation. It's about intermediation.

OpenTable, as a f'r'instance, is a middleman between what

7



Middleman

Because, where there used to be one merchant, engaged in the transaction, when the middleman enters, there are now two.

And there are two databases.

8



Of course the middleman is not a new idea. The retail industry is an industry of middlemen. But the internet and computers and data collection are changing what it means to be a middleman.

The idea of the middleman was to collect a percentage of each transaction.

9



Say somebody buys a soda at RiteAid or Dollar General.

RiteAid collects the information about the transaction. The transaction happened at RiteAid, so it's RiteAid's data. Even though somebody (not RiteAid) bought Coke (not a RiteAid product).

10

Curiosity



11

And presumably, RiteAid can sell the data to Pepsi -- about who's buying Coca-Cola.

Dollar General provides (e.g.) POS data about competitors products, but not inventory data on competitors products.s



Clickstreams

At Esther Dyson's last PC Forum -- one of the sessions had a gentleman from a company that tracked web browsing behavior. (Maybe Nielsen?)

Given a set of URLs that somebody had visited, they could predict
a) who somebody's cell phone carrier was, and
b) if they were thinking of switching, and
c) who they were thinking of switching to.

12

You could sell this data to



Competition

13



We call it “competition” -- but it’s really about “peeking”. About finding out stuff that other people don’t want you to know.

14



Anonymity

What we learned from the AOL data release in 2006, is that anonymity is a very fragile thing.

By analyzing data, and cross referencing it against other data, one can learn more than you might think. Or, in other words, as you bring more data together, it becomes more valuable.

15

A screenshot of a Bloomberg news article. The Bloomberg logo is at the top left in orange. To its right are links for 'Anywhere | Professional | Solutions | About'. Below the logo is a navigation bar with 'QUICK', 'NEWS', 'VIEW', 'MARKETS', 'PERSONAL FINANCE', 'TV', 'RADIO', and 'MORE'. Underneath is a 'Related News' section with links for 'Finance', 'U.S.', 'Bonds', and 'Real Estate'. The main headline reads 'Mortgage Investors Are Duped by False Owner-Occupancy Claims, Study Finds'. Below the headline is the author 'By Jody Shenn - Jan 21, 2011 11:58 AM ET'. At the bottom left are social media icons for 'Tweet', 'More', 'Print', and 'Email'. To the right of these icons is a short excerpt: 'Borrowers either never moved into or later vacated homes for at least 25 percent of U.S. mortgages that were described as being for "owner occupied" properties when bundled into securities, according to 1010data.'

By combining data from CoreLogic (mortgage origination data) and Equifax (credit reporting) we were able to synthesize information which was not available in either separately -- i.e. identify which mortgages were fraudulently obtained.

In order to do so, one needed to look at every mortgage, and then see if one could figure out who's mortgage it was.

16

~~That means comparing every mortgage to~~



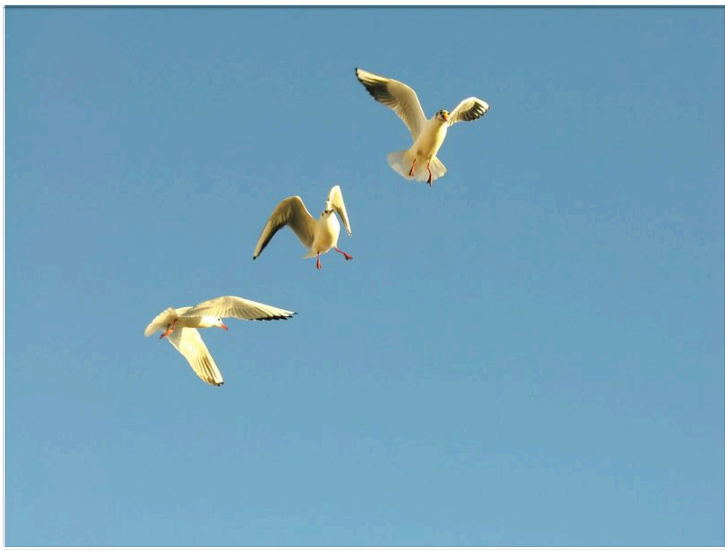
Cash

17



Another middleman

18



Another middleman

Now we have the retailer collecting all this information about the transaction, and the credit card company collecting all this information about the transaction.

And the credit card company collects data about transactions that the retailer doesn't (because it's at another retailer). And the retailer collects data about transactions that

19

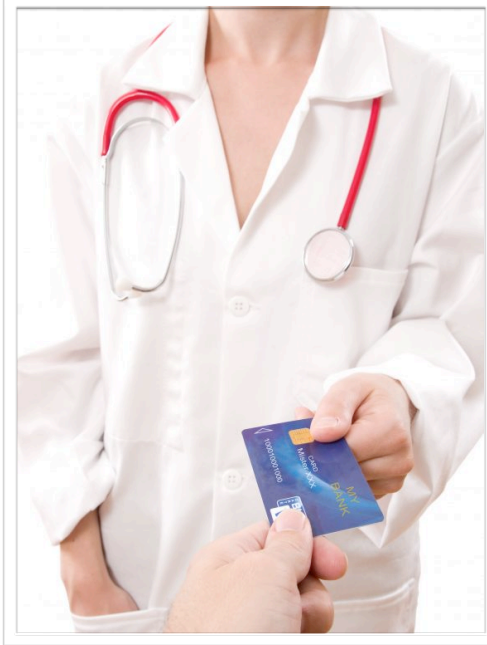
The value of the sum is greater than the sum of the value.

People have known this for years.

Mailing lists have often been shared because when companies combine lists, the combined list is more valuable to each party than the individual lists are.

20

Privacy



But there are laws about who can tell/sell what to whom.

So, if the transaction is about something medical, it's different than if it isn't medical.

I.e., I can sell data about who's buying Coke and Pepsi -- but not about who's buying Thorazine or

21



Here's an example -- the telephone company can't sell the data about who you called. Everybody knows that (especially after the HP pretexting scandal in 2006). 2006 was a great year for data privacy issues.

22



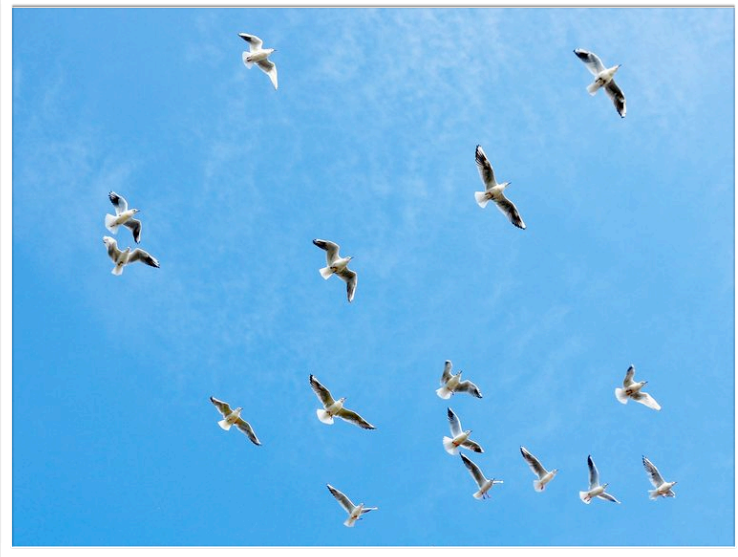
Who called?

So, what about the data about who called you?

What about aggregate data about how many people called you by zip code?

Could you buy the aggregate transaction data about how many calls cardiologists receive by zip code?

23



Lots of middlemen

And they all collect data, and that data is valuable. And they could sell it individually (as RiteAid does).

But the data, when combined, is more valuable than when it is not combined. So there is an opportunity to combine all this

24

Summary

- Middlemen have visibility into “other people’s data”.
- The value of combining datasets exceeds the combined value of datasets.
- Detail transactions are more valuable than aggregate statistics.

25

Our retail customers sell data about transactions to their suppliers.



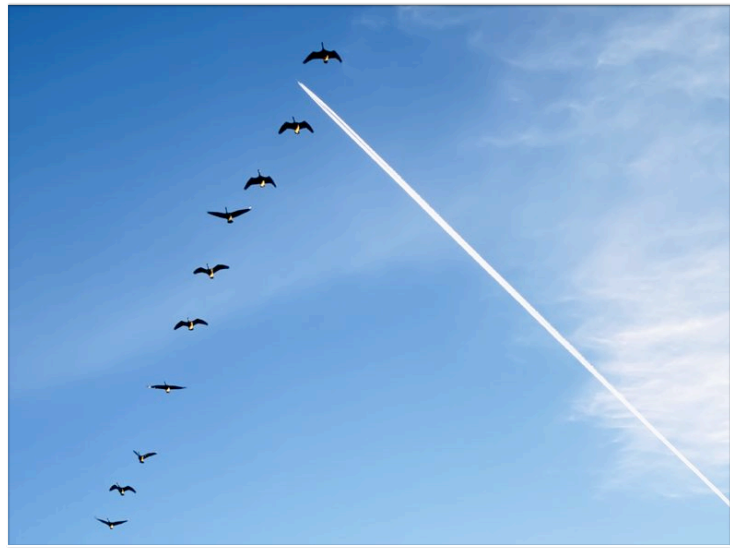
Let me suggest that there is an opportunity for Data middlemen.

That is, if the internet is the great enable of middlemen -- and people want to buy and sell data, then there is room for a middleman between those buyers and sellers.

We're one of those middlemen.

There's the usual advantage of middlemen

26



Speed

One, we could do it better.

So, there are all these databases from all these middlemen -- but we might provide a product that was faster and better. And there's value in that -- because detail is more valuable than aggregate -- so you need to process all that detail.

27



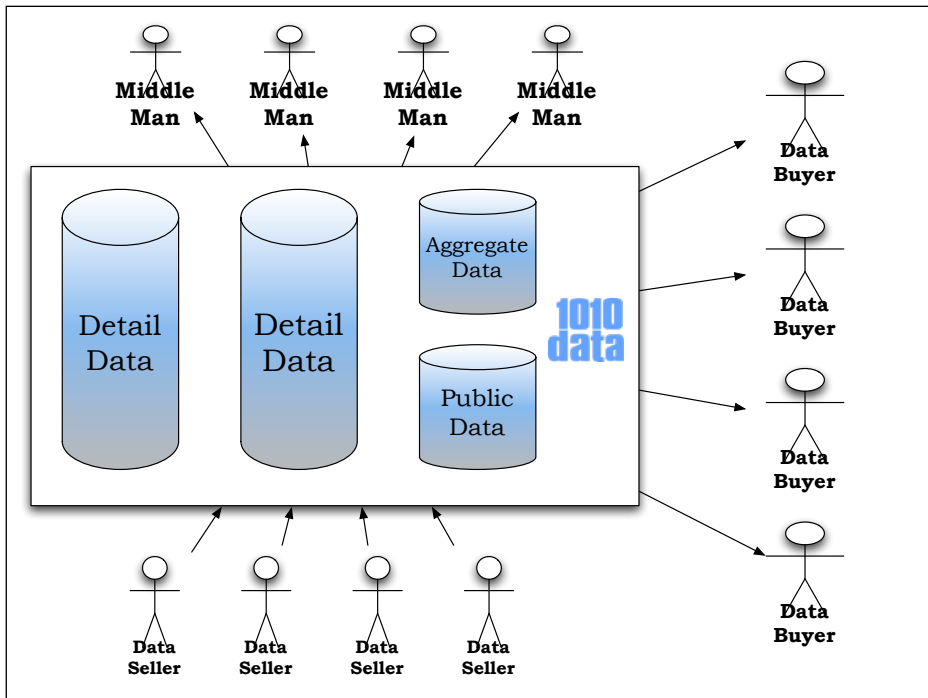
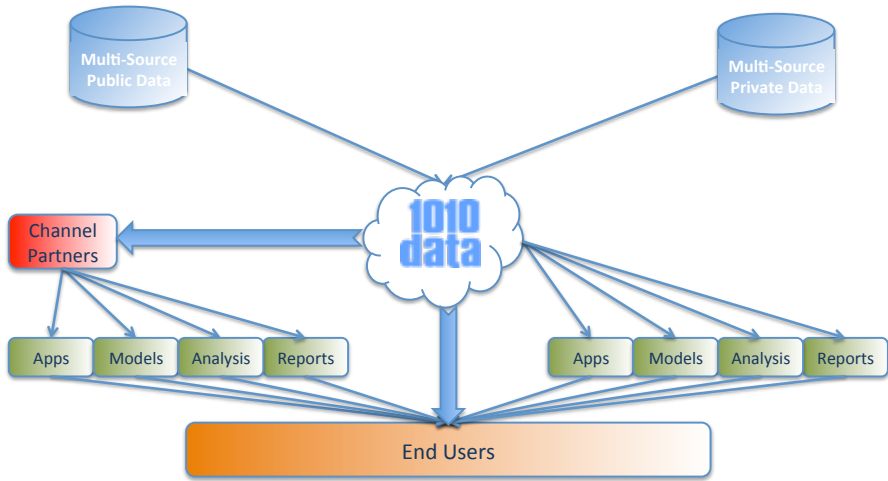
But it isn't just the speed, it's the network effects. Middlemen build network effects.

You need to be able to combine your data with the data from other middlemen to make it more valuable. And, presumably, every company could undertake to establish a relationship with every data vendor and deal with acquiring and managing every database.

Or, there could be a middleman.

28

Vision: Analytical Ecosystem



Turning Their Data into Your Money (and vice versa)

Robert M. (r0ml) Lefkowitz
r0ml@1010data.com