

Data Selection and Procurement (and a Little Big Data)

Carl F. Mela¹

¹Department of Marketing
Fuqua School of Business, Duke University

2013 Workshop on Quantitative Marketing And
Structural Economics



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Hal Varian on Data (2009)

“I keep saying the sexy job in the next ten years will be statisticians. The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that’s going to be a hugely important skill in the next decades. *Because now we really do have essentially free and ubiquitous data.* So the complimentary scarce factor is the ability to understand that data and extract value from it.”



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Surely Google Can Help Me ...



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

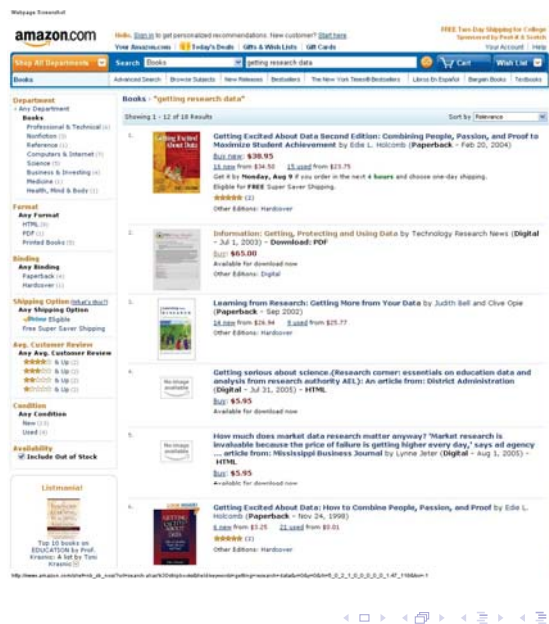
Project Management

Sharing Data

Big Data

Summary

Maybe There's a Book on This ...



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Determine the Data You Need

Start With the Research Question

Why use a structural model?

- ▶ Infer missing primitives (Nevo 2001)
- ▶ Test competing theories (Draganska, Klapper and Vilas-Boas 2010)
- ▶ Policy simulation (Duflo, Hanna and Ryan 2012)

Better data and/or better model?

- ▶ Chintagunta, Dube and Singh (2003) observe wholesale costs - instrument in demand estimation
- ▶ Fong, Simester and Anderson (2011) use price experiments to deal with price endogeneity
- ▶ Albuquerque and Bronnenberg (2009) micromoments to infer preference



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Determine the Data You Need

Typology of Data

There are several sources of data:

- ▶ Proprietary firm data (free, flexible, slow)
 - ▶ e.g., I married for it (Mela, Gupta and Lehmann 1997)
 - ▶ My emphasis today
- ▶ Public data and records (free, less flexible, medium)
 - ▶ e.g., Study of Adolescent Health (Mele 2013)
 - ▶ scrape the Internet (Roos, Mela and Shachar 2013)
- ▶ Private data Nielsen; IMS and JD Power (expensive, rigid, fast)
 - ▶ e.g., Impact RX physician prescriptions (Narayanan and Manchanda 2009)
 - ▶ Grants or library can help with cost (TNS Media data)



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Determine the Data You Need

Proprietary Data Example

How does marketing affect social networking? What information do you need?

Which interactions available and of interest (DV)?

- ▶ Frequency, intensity, and direction of interactions, syntax of interactions
- ▶ What drives the social interactions (IV)?
 - ▶ Marketing, past behaviors, etc.
- ▶ Plausible instruments to identify effect of interest?
 - ▶ Wind instruments for blogging on friending (Shriver, Nair and Hofstetter 2013)

How much data can you handle (agents, periods, states, & variables)?

- ▶ Consider simulating ideal data set and model prior to searching for the data!



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Determine the Data You Need

Proprietary Data Example

How does marketing affect social networking? What information do you need?

Which interactions available and of interest (DV)?

- ▶ Frequency, intensity, and direction of interactions, syntax of interactions
- ▶ What drives the social interactions (IV)?
 - ▶ Marketing, past behaviors, etc.
- ▶ Plausible instruments to identify effect of interest?
 - ▶ Wind instruments for blogging on friending (Shriver, Nair and Hofstetter 2013)

How much data can you handle (agents, periods, states, & variables)?

- ▶ Consider simulating ideal data set and model prior to searching for the data!



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Determine the Data You Need

Public Data Example

Effort and Success of Studies Using Personal Data Collection versus Company-Supplied Data Bases

Gerard J. Tellis

Paper	Topic	Time	Effort	Sources	Published In	Award	Citations Google
Golder and Tellis 1993	Pioneering advantage	2 years	550 man hours	Numerous magazines and 100+ books	JMR	William Odell	> 460
Chandy & Tellis 2000	Incumbent's Curse	4 years	10 researchers, numerous hours	Over 250 books, over 500 articles in periodicals	JM	Harold Maynard	> 300
Golder and Tellis 1997	New product takeoff	6 months	400 man hours	Numerous periodicals and books	MktScience	Frank M. Bass	> 160
Johnson and Tellis 2008	Success of Entry in China and India	Intermittent over 9 years	10 people, numerous hours	ABI/inform, Lexis Nexis, FACTIVA and others	JM	Paul Root/MSI	> 60

References

Chandy, Rajesh and Gerard J. Tellis (2000), "The Incumbent's Curse? Incumbency, Size and Radical Product Innovation," *Journal of Marketing*, 64, 3 (July), 1 to 17.

Golder Peter N and Gerard J. Tellis (2003), "Pioneering Advantage: Marketing Logic or Marketing Legend," *Journal of Marketing Research*.

Golder, Peter N. and Gerard J. Tellis (1997), "Will It Ever Fly? Modeling The Takeoff of New Consumer Durables," *Marketing Science*, 16, 3, 256-270.

Johnson, Joseph and Gerard J. Tellis (2008), "Drivers of Success for Market Entry Into China and India," *Journal of Marketing*, 72 (May) 1-13.



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Determine the Data You Need

Marketed Data Examples

Goldfarb and Xiao (2011)

- ▶ Do managers think strategically about entry?
 - ▶ telephone entry data \$5000 (DV)
 - ▶ census.gov and data.gov and dataferret for local market conditions (IV)
 - ▶ Internet search (phone directories, who's who, linked-in, etc.) for managerial experience (IV)



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Find the Right Contact

Sources

- ▶ Corporate
- ▶ Personal
- ▶ Collegial
- ▶ Public



- Data Procurement
- Carl F. Mela
- Determine The Data You Need
- Find the Right Contact**
- Make the Pitch
- Negotiate the NDA
- Data Delivery
- Data Checking
- Project Management
- Sharing Data
- Big Data
- Summary

Find the Right Contact

Corporate - 1

Past colleagues from work (Yao and Mela 2009)

Consulting

- ▶ free consulting for free data
- ▶ old data from old consulting

Cold calls (they work for Catherine!)



- Data Procurement
- Carl F. Mela
- Determine The Data You Need
- Find the Right Contact**
- Make the Pitch
- Negotiate the NDA
- Data Delivery
- Data Checking
- Project Management
- Sharing Data
- Big Data
- Summary

Find the Right Contact

Corporate - 2

Relevant work leads to relevant data - get known!

- ▶ Advisory boards (Bronnenberg Kruger and Mela 2008)
- ▶ Industry Conferences and Seminars
 - ▶ ARF, ART, IIR, DMEF, WOMMA, etc.
 - ▶ Wilbur and DVR/set top data
 - ▶ Kamakura, Fader ART Forum Tutorials
- ▶ Local lunches
- ▶ Publish in HBR & other managerial outlets
- ▶ Publish books and software
- ▶ PR your work



Data
Procurement

Carl F. Mela

Determine The
Data You Need

**Find the Right
Contact**

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Find the Right Contact

Personal

Family (Yao and Mela 2010; Ansari and Mela 2003)

Friends (College buddies)



Data
Procurement

Carl F. Mela

Determine The
Data You Need

**Find the Right
Contact**

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Make the Pitch

What does the firm gain (incentive compatible)?

- ▶ What decisions will they make better?
- ▶ What is impact on profits or revenue?
- ▶ How will your contact get promoted for helping you?

What does it cost the firm?

- ▶ Dollar and time costs
- ▶ Do your own data cleaning if it's easiest for firm
- ▶ "Executives rank providing data to academics right there with root canals and colonoscopies. "Have cool tool" is not effective"
 - ▶ W. Kamakura



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Make the Pitch

Peter Rossi's advice: "Go to the absolute top people in the org. It is almost impossible to get data from mid level people. Then get down on your knees and beg."



Your pitch should be concise

- ▶ What will the firm gain, what will it cost, what is the time line, and why do they need you?
- ▶ Most do not care at all about publications

Role play: persuade Google exec to provide you data on key word search ...



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Negotiate the Non Disclosure Agreement (NDA) | Firm

The firm's goal is to gain actionable insights not shared by other firms nor interpreted by analysts

The firm also seeks to keep its clients confidential

Your goal is to publish

Disguising data and/or aggregating results achieves these aims

- ▶ Be careful - in IRI data set some markets had only a few mass merch – even if disguised one could infer
- ▶ Netflix prize cancellation

Trust is critical



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

**Negotiate the
NDA**

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Negotiate the Non Disclosure Agreement (NDA) | Firm

The firm's goal is to gain actionable insights not shared by other firms nor interpreted by analysts

The firm also seeks to keep its clients confidential

Your goal is to publish

Disguising data and/or aggregating results achieves these aims

- ▶ Be careful - in IRI data set some markets had only a few mass merch – even if disguised one could infer
- ▶ Netflix prize cancellation

Trust is critical



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

**Negotiate the
NDA**

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Negotiate the Non Disclosure Agreement (NDA) I

Firm

The firm's goal is to gain actionable insights not shared by other firms nor interpreted by analysts

The firm also seeks to keep its clients confidential

Your goal is to publish

Disguising data and/or aggregating results achieves these aims

- ▶ Be careful - in IRI data set some markets had only a few mass merch – even if disguised one could infer
- ▶ Netflix prize cancellation

Trust is critical



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

**Negotiate the
NDA**

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Negotiating the NDA II

University

The University's goal is not to be sued



Your goal is not to be sued. Use university lawyers!

Email your faculty; several examples they'll redact

Check with human subjects as well (Institutional Review Board) for schools and governments regulations



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

**Negotiate the
NDA**

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Negotiating the NDA II

University

The University's goal is not to be sued



Your goal is not to be sued. Use university lawyers!

Email your faculty; several examples they'll redact

Check with human subjects as well (Institutional Review Board) for schools and governments regulations



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

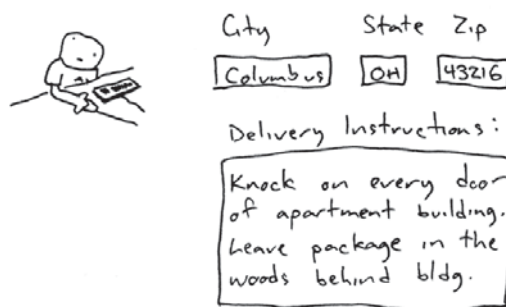
Summary

Data Delivery

Large USB drives work great.

"Get physically in the building with an external hard drive having made an appointment with the IT guy and stand over him while he downloads the data." - Catherine Tucker

Use metadata and dictionaries prior to entire data set!



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Data Checking

Check variable names and columns

- ▶ WTF is “user_action_id skc_url”?

Are the variable values feasible?

- ▶ Frequencies, means, regression, correlations
- ▶ Which variable let you address research question?

Are the fields well populated?

Pivot variables (relational databases e.g., Sprint)

Do dictionaries make complete sense?



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

Project
Management

Sharing Data

Big Data

Summary

Project Management - I

Early Phase

- ▶ After data checking, follow up quickly
- ▶ Some data collected going forward after detailing needs
 - ▶ Yao and Mela (2010)

Mid Phase

- ▶ Bimonthly or monthly phone or email updates



Data
Procurement

Carl F. Mela

Determine The
Data You Need

Find the Right
Contact

Make the Pitch

Negotiate the
NDA

Data Delivery

Data Checking

**Project
Management**

Sharing Data

Big Data

Summary

Project Management - II

Completion Phase

- ▶ Phone or on site conference
- ▶ Contact ensures all relevant players are there
 - ▶ Forget the method (beyond concept)
 - ▶ Focus on what they should do differently and profit/sales outcomes

Implementation Phase (Misra and Nair 2011)

- ▶ Frequent meetings, institutional constraints, and measurement



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Sharing Data

Sharing makes us all better off

Most NDAs will not allow you to share

Some firms insist data returned when project done

Negotiate to share data after finite period

- ▶ Old data not so relevant to firm



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Big Data

What Is It?

What is Big Data (Wikipedia)?

- ▶ Moving target: Don Lehmann inverted matrices by pen
- ▶ “Capture, curation, storage, search, sharing, transfer, analysis, and visualization.”

What is the point of Big Data?

- ▶ Determine needs and collect data as noted above
- ▶ Why archive everything “just in case?”



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Big Data

Data Structure Issues (Too Much and Too Little)

Often a Relational Database Structure e.g. SQL

Complete but sparse

- ▶ Low response rates and rare events
 - ▶ Non-parametrics for rare events (Naik and Tsai 2001)
 - ▶ Data fusion for missing data (Gilula et al. 2006; 2011)

Large N and Large P

- ▶ 100s of covariates (P)
 - ▶ Space reduction methods (Naik and Tsai 2004); variable selection (Fan and Li 2001), etc.
- ▶ Many millions of observations (N)
 - ▶ Sampling with macromoments/machine learning?



Data Procurement

Carl F. Mela

Determine The Data You Need

Find the Right Contact

Make the Pitch

Negotiate the NDA

Data Delivery

Data Checking

Project Management

Sharing Data

Big Data

Summary

Big Data

Other Issues

Data Visualization (even for small n and p; Mela et al. 2013)

- ▶ Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. "A tour through the visualization zoo".
- ▶ Rafe M. J. Donahue. Fundamental Statistical Concepts in Presenting Data. Processing



Algorithms and Computation (Miranda and Fackler 2002; Judd 1998)

- ▶ Efficiency in computation
- ▶ Use efficiency for more data or more complexity?

- Data Procurement
- Carl F. Mela
- Determine The Data You Need
- Find the Right Contact
- Make the Pitch
- Negotiate the NDA
- Data Delivery
- Data Checking
- Project Management
- Sharing Data
- Big Data**
- Summary

Big Data

Quoted Advice from the Gurus

- ▶ Start with an existing capability for data management and analytics
- ▶ Add some unstructured, large-volume data
- ▶ Throw some product/service innovation into the mix
- ▶ Add a dash of Hadoop
- ▶ Cook up some data in a high-heat convection oven
- ▶ Embed this dish into a well-balanced meal of processes and systems
- ▶ Promote the chef to Chief Analytics Office

http://digitalcommunity.mit.edu/community/featured_content/big-data/blog/2013/07/16/tom-davenport-reports-on-big-data-in-big-companies

- Data Procurement
- Carl F. Mela
- Determine The Data You Need
- Find the Right Contact
- Make the Pitch
- Negotiate the NDA
- Data Delivery
- Data Checking
- Project Management
- Sharing Data
- Big Data**
- Summary

In Conclusion

Data

- ▶ Types of data (proprietary, public, and for sale) and how to get it (steps on right)
- ▶ Big data becomes data - so see above

Acknowledgments

- ▶ Thanks for your interest
- ▶ Thanks to Peter Fader, Avi Goldfarb, Mike Hanssens, John Hauser, Wagner Kamakura, Peter Rossi, Gerry Tellis, Catherine Tucker for ideas and information
- ▶ Thanks to Brett Gordon, Rick Staelin, and Raph Thomadsen for the conference
- ▶ Thanks to all the firms who coughed up data for me