



Linking Financial Data Sets: Possible Problems and Solutions

Sept. 30, 2014

Kathleen Dreyer, Columbia University

Todd Hines, Princeton University

Using financial data

Very common occurrence

- Financial data comes from multiple sources

Another very common occurrence

- The different sources use different identifiers

In order to link/merge the data together, common identifiers are required.



Identifiers

Identifiers used to match data sets:
(ranked from less precise to more precise)

1. Name or ticker
2. CUSIP
 - a. SEDOL or ISIN for non-US companies
3. Unique id, e.g. GVKEY, PERMNO

Examples of financial data

Compustat data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	GVKEY	datadate	fyyear	indfmt	consol	popsrc	datafmt	TIC	CUSIP	Company Name	curcd	Total Assets	Total Liabilities
2	3226	20111231	2011	INDL	C	D	STD	CMCSA	20030N101	COMCAST CORP	USD	8573	13241
3	3226	20121231	2012	INDL	C	D	STD	CMCSA	20030N101	COMCAST CORP	USD	19991	16714
4	3226	20131231	2013	INDL	C	D	STD	CMCSA	20030N101	COMCAST CORP	USD	14075	18912
5	176404	20111231	2011	INDL	C	D	STD	TWC	88732J207	TIME WARNER CABLE INC	USD	6398	5370
6	176404	20121231	2012	INDL	C	D	STD	TWC	88732J207	TIME WARNER CABLE INC	USD	4877	5325
7	176404	20131231	2013	INDL	C	D	STD	TWC	88732J207	TIME WARNER CABLE INC	USD	2144	5226

ThomsonReuters data

Date Announced	Target Name	Target CUSIP	Acquiror Name	Acquiror CUSIP	% of Shares Acq.	Enterprise Value at Announcement (\$mil)	Master Deal Type
02/13/2014	Radiant Services LLC	9A8261	Accenture LLP	9A7666		- np	MA
02/13/2014	Continental Carbonic Products	8A9236	Matheson Tri-Gas Inc	59179J	100.00	234	MA
02/13/2014	Easton-Bell Sports Inc-Easton	8A9198	Bauer Performance Sports Ltd	07162P	100.00	np	MA
02/13/2014	Advantage Business Media LLC	06110R	Owner Resource Group LLC	69123M	100.00	np	MA
02/13/2014	Time Warner Cable Inc	88732J	Comcast Corp	20030N	-	70,677	MA
02/13/2014	LCA-Vision Inc		501803 PhotoMedex Inc		719358	-	77 MA
02/13/2014	Third Coast Search	9A1486	HireBetter LLC	9A1470	100.00	np	MA
02/13/2014	Phillips 66 Co-Gold Line Syste	8A9580	Phillips 66 Partners LP		718549	100.00	np MA
02/13/2014	Limelight Networks Inc	53261M	Limelight Networks Inc	53261M		- np	MA
02/13/2014	Daniels Uselton & Clay PC	8A9527	Collinsworth Bright & Co PC	8A9523	100.00	np	MA
02/13/2014	Fortress Investment Group LLC	34958B	Fortress Investment Group LLC	34958B		25.28	1,437 MA



Linking by company name

NEWS

Hillshire Brands gains in debut

From wireports

357 words

30 June 2012

Chicago Daily Herald

DHLD

1

English

© 2012 Chicago Daily Herald. Provided by ProQuest Information and Learning. All Rights Reserved.

NEW YORK -- **Hillshire** Brands Inc. gained on its debut on the New York Stock Exchange Friday after emerging from Sara Lee Corp.

The new company, which trades under the ticker "HSH," focuses on meat brands such as Jimmy Dean, Ball Park and its namesake **Hillshire** Farm, which Sara Lee acquired in 1971. It also sells Sara Lee frozen desserts and bakery products.

The other company created after Sara Lee's split in two is an international coffee and tea seller called D.E. Master Blenders 1753, a reference to one of the company's coffee brands. It is based in the Netherlands and will trade on an exchange there.

Hillshire says the spinoff will let it focus on developing meat products that appeal to changing tastes.

Source: Factiva.com

Linking on ticker symbol

Source: Bloomberg

GRAB		Ticker Symbol Change	
<Menu> To Return			
Company	Sprint Communications Inc		
Ticker	0848680DUS		
Announced Date	08/09/05		
Effective Date	08/15/05		
Old Ticker	FON US		
New Ticker	S US		
Reason	Name Change		
		Action ID	18954835

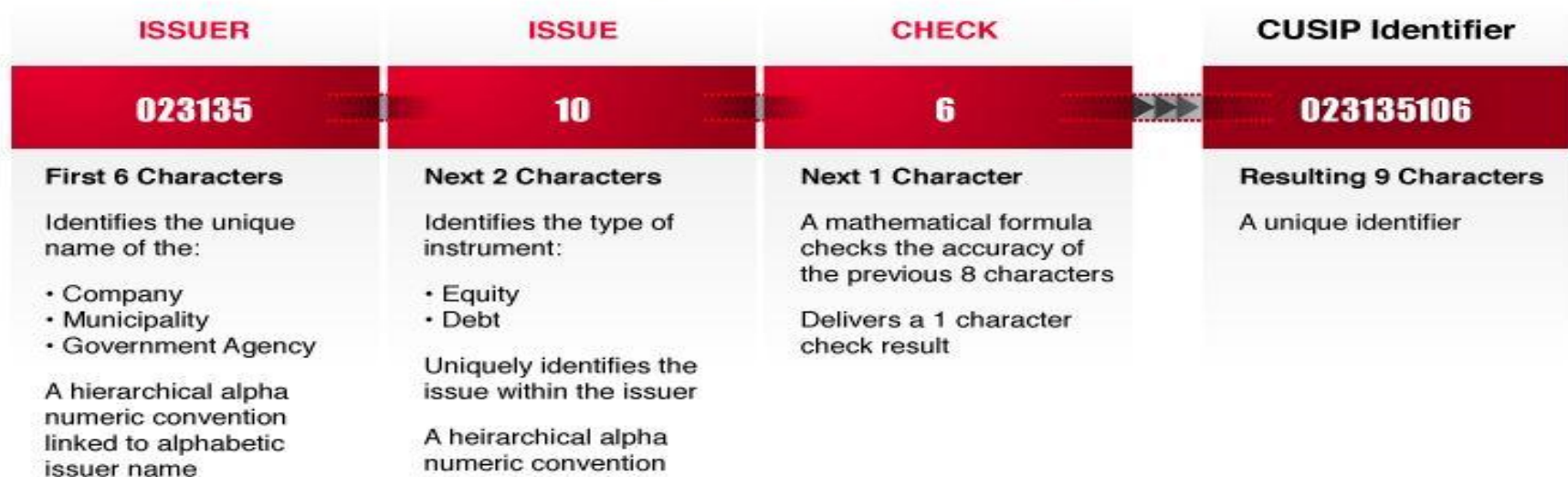
Australia 61 2 9777 8600 Brazil 5511 3048 4500 Europe 44 20 7330 7500 Germany 49 69 9204 1210 Hong Kong 852 2977 6000
Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000 Copyright 2014 Bloomberg Finance L.P.
SN 817397 G687-3793-2 23-Sep-14 11:03:01 EDT GMT-4:00

Linking on CUSIP

Structure of a CUSIP

CGS provides a 9-character standard CUSIP identifier for issuers and their financial instruments offered in the U.S. and Canada

Example: Amazon.com Inc - Common Stock



Source: CUSIP Global Services (<http://www.cusip.com/cusip/about-cgs-identifiers.htm>)



Difficulties with CUSIP

- Dell changed their name in July 2003 from Dell Computer Corp to Dell Inc.
- The ticker remained DELL but the CUSIP changed
- Before name change CUSIP was 24702510
- After name change (from 7/22/03) it was 24702R10
- CUSIPs have only existed since the late 1960s.
 - Can't use to match for earlier data

Examples of unique identifiers

From CRSP

<input type="checkbox"/>	SIC	PERMNO	PERMCO	CUSIP_FULL	TICKER	COMPANY_NAME	FIRST_DATE	LAST_DATE
<input type="checkbox"/>	3577	39087	21833	85206110	FON	SPRINT CORP	1992-02-27	2005-08-14
<input type="checkbox"/>	4813	86414	21833	85206150	PCS	SPRINT CORP	1998-11-24	2004-04-22
<input type="checkbox"/>	3663	14040	54473	85207U10	S	SPRINT CORP NEW	2013-07-12	2013-12-31
<input type="checkbox"/>	4813	39087	21833	85206110	S	SPRINT NEXTEL CORP	2005-08-15	2013-07-10



From Compustat

<input type="checkbox"/>	SIC	FIRST_DATE	LAST_DATE	GVKEY	COMPANY_NAME	TICKER	CUSIP_FULL	CIK
<input type="checkbox"/>	4841	1995-01-01	2001-12-31	066704	COMCAST CABLE COMMUNICATIONS	3513B	20029P935	00010405
<input type="checkbox"/>	6799	1983-01-01	1992-12-31	012721	COMCAST CABLEVISION-PHILA	CCPI.1	200299105	00007114
<input type="checkbox"/>	4812	1995-01-01	1998-12-31	112453	COMCAST CELLULAR CORP	9659B	199904939	00010418
<input type="checkbox"/>	4841	1974-01-01	2013-12-31	003226	COMCAST CORP	CMCSA	20030N101	00011666
<input type="checkbox"/>	4841	2001-01-01	2001-12-31	151503	COMCAST CORP-PRO FORMA	CMCS.P	20099X933	
<input type="checkbox"/>	4841	1978-01-01	1999-12-31	006282	COMCAST JOIN HOLDINGS	CMCSK1	20099X008	00002756

Let's try an example!

A researcher would like to examine the effectiveness of a merger.

What data do we need?



Effectiveness of mergers

- To analyze, we'll need
 - Merger data (acquiror, target company, deal size, deal dates, etc.)
 - Stock price information
 - Did the deal have a positive effect on the acquirer's stock price?
 - Fundamentals (financial statement) information
 - Analyze the financials of the acquiror both and after the deal

Matching example

Unfortunately, all these variables come from different databases. Fun!

Some common databases for:

- M&A data -- SDC Platinum (a.k.a. Thomson ONE), Zephyr
- Stock price - CRSP (US stock prices)
- Fundamentals (financial statement data) - Compustat North America (US company fundamentals)

Sara Lee (Hillshire Brands) - company history

CRSP Names File					
PERMNO	DATE	NAMEENDT	COMNAM	NCUSIP	TICKER
22840	19460710	19540223	CONSOLIDATED GROCERS CORP		
22840	19540224	19620701	CONSOLIDATED FOODS CORP		
22840	19620702	19680101	CONSOLIDATED FOODS CORP		CFD
22840	19680102	19850401	CONSOLIDATED FOODS CORP	20921910	CFD
22840	19850402	20010823	SARA LEE CORP	80311110	SLE
22840	20010824	20020101	SARA LEE CORP	80311110	SLE
22840	20020102	20040609	SARA LEE CORP	80311110	SLE
22840	20040610	20110207	SARA LEE CORP	80311110	SLE
22840	20110208	20120628	SARA LEE CORP	80311110	SLE
22840	20120629	20131231	HILLSHIRE BRANDS CO	43258910	HSB



Sara Lee (Hillshire Brands) selected M&A transactions

Thomson (SDC) M&A Data

Date Announced	Acquiror Name	Acquiror CUSIP	Target Name	Target CUSIP	Value of Transaction (\$mil)
05/12/2014	Hillshire Brands Co	432589	Pinnacle Foods Inc	72348P	4,284.19
07/02/2001	Sara Lee Corp	803111	Earthgrains Co	270319	2,905.88
04/22/1999	Sara Lee Corp	803111	Chock Full O'Nuts Corp	170268	242.48
09/06/1978	Consolidated Foods Corp	209219	Hanes Corp	410342	320.51



CRSP Data

Thomson (SDC) M&A Data

PERMNO	COMNAM	6 digit NCUSIP	Acquiror CUSIP	Date Announced	Acquiror Name	Target Name	Target CUSIP	Value of Transaction (\$mil)
22840	CONSOLIDATED FOODS CORP	209219	209219	09/06/1978	Consolidated Foods Corp	Hanes Corp	410342	320.51
22840	SARA LEE CORP	803111	803111	07/02/2001	Sara Lee Corp	Earthgrains Co	270319	2,905.88
22840	SARA LEE CORP	803111	803111	04/22/1999	Sara Lee Corp	Chock Full O'Nu	170268	242.48
22840	HILLSHIRE BRANDS CO	432589	432589	05/12/2014	Hillshire Brands Co	Pinnacle Foods	72348P	4,284.19

Why is it important we did this?

- We would have thought the the SDC M&A deals were for three completely different companies:
 - Hillshire Farms
 - Sara Lee
 - Consolidated Food Corp
- All three have different names, different CUSIPs, different tickers

Will this solve all your problems?

- Most companies will “match” if you use permanent identifiers, but there still can be problems
 - There is a lot of financial engineering now - spinoffs, tracking stocks, complicated merger deals, etc. Sometimes the researcher will need to read up on the specifics of a deal in order to match it correctly.

Matching CRSP to Compustat

- A very common problem is to need to match CRSP stock data to Compustat fundamentals (financial statement) data
- Problem
 - Compustat has a “unique identifier” GVKEY
 - But the other company identifiers (name, CUSIP) are what is called “header data”
 - Only the most recent value (name, CUSIP) is available in Compustat

Matching CRSP to Compustat (cont.)

- For Hillshire Farms (f.n.a. Sara Lee) this means that only the most recent company name (**Hillshire Farms**) and most recent CUSIP (the one starting with **432589**) is in the Compustat record.
- There is a subscription database called CRSP/Compustat merged.
 - If you have a GVKEY, you can use the database to get the Permno. And vice versa.
- Most institutions don't have a subscription, so you have to do the matching yourself

Compustat Record

<u>GVKEY</u>	<u>FIRST_DATE</u>	<u>LAST_DATE</u>	<u>COMPANY_NAME</u>	<u>TICKER</u>	<u>CUSIP_FULL</u>
9411	1/1/1950	12/31/2013	HILLSHIRE BRANDS CO	HSH	432589109

Matching CRSP to Compustat

As when we were matching the SDC data, having access to the CRSP file, which details all company names, tickers and CUSIP numbers used, will allow us to match.

Sara Lee (Hillshire Brands) - company history

CRSP Names File					
PERMNO	DATE	NAMEENDT	COMNAM	NCUSIP	TICKER
22840	19460710	19540223	CONSOLIDATED GROCERS CORP		
22840	19540224	19620701	CONSOLIDATED FOODS CORP		
22840	19620702	19680101	CONSOLIDATED FOODS CORP		CFD
22840	19680102	19850401	CONSOLIDATED FOODS CORP	20921910	CFD
22840	19850402	20010823	SARA LEE CORP	80311110	SLE
22840	20010824	20020101	SARA LEE CORP	80311110	SLE
22840	20020102	20040609	SARA LEE CORP	80311110	SLE
22840	20040610	20110207	SARA LEE CORP	80311110	SLE
22840	20110208	20120628	SARA LEE CORP	80311110	SLE
22840	20120629	20131231	HILLSHIRE BRANDS CO	43258910	HSB



CRSP Names File

Compustat Record

PERMNO	COMNAM	NCUSIP	<u>CUSIP</u>	<u>GVKEY</u>	<u>FIRST_DATE</u>	<u>COMPANY_NAME</u>	<u>TICKER</u>
22840	HILLSHIRE BRANDS CO	432589	432589	9411	1/1/1950	HILLSHIRE BRANDS CO	HSB

How researchers do matching

- Researchers seldom look at one company at a time
- Usually are looking at hundreds or thousands of companies and deals
- This means they'll generally use statistical software such as Stata, SAS, R, etc. to work with the data
 - This software allows researchers to write programs that will follow the same matching rules that we just performed for one company, Hillshire Farms (Sara Lee), but for many companies at once.

Questions?

Contact information:

Kathleen Dreyer: kd2145@columbia.edu

Todd Hines: thines@princeton.edu