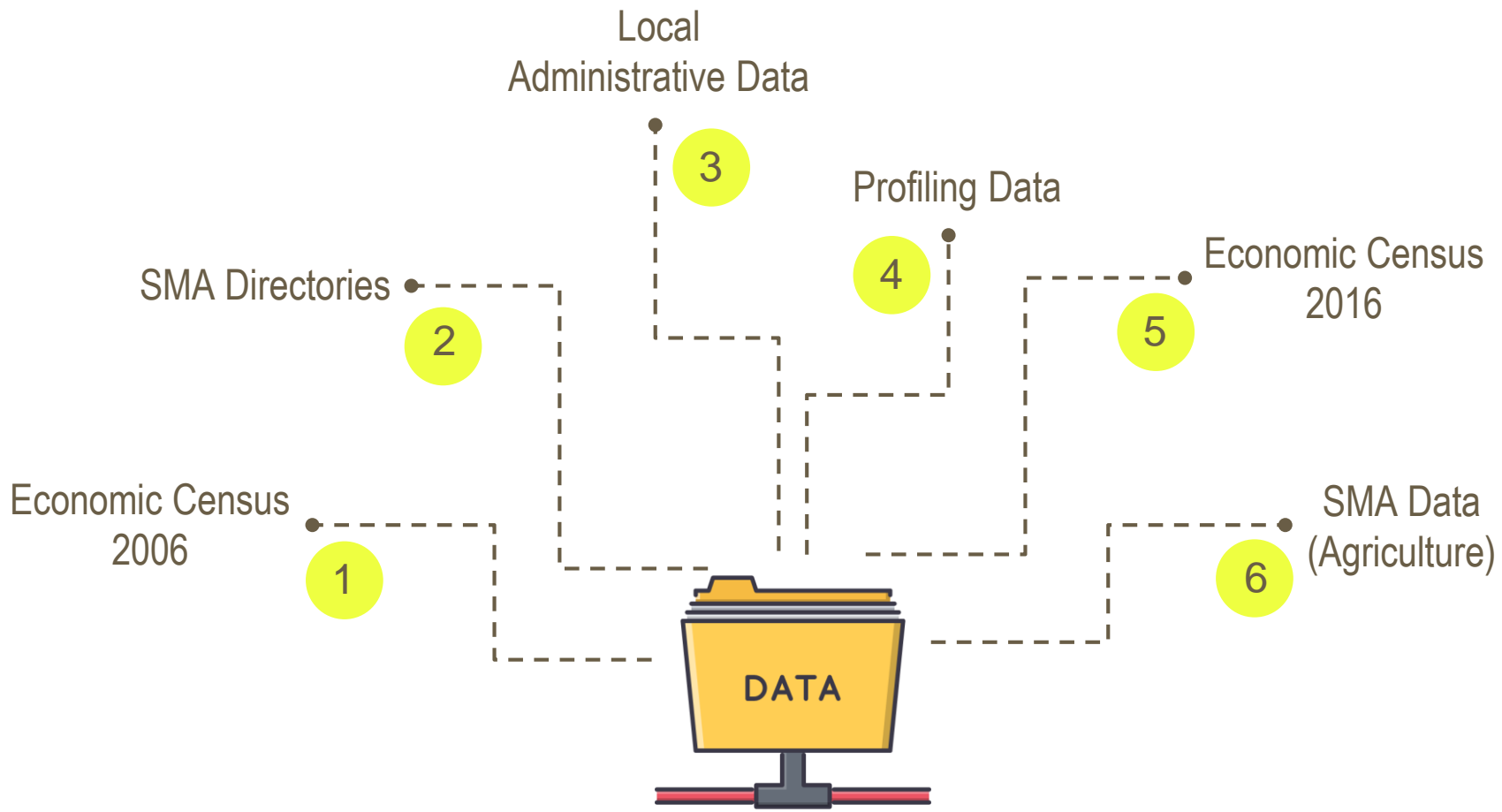
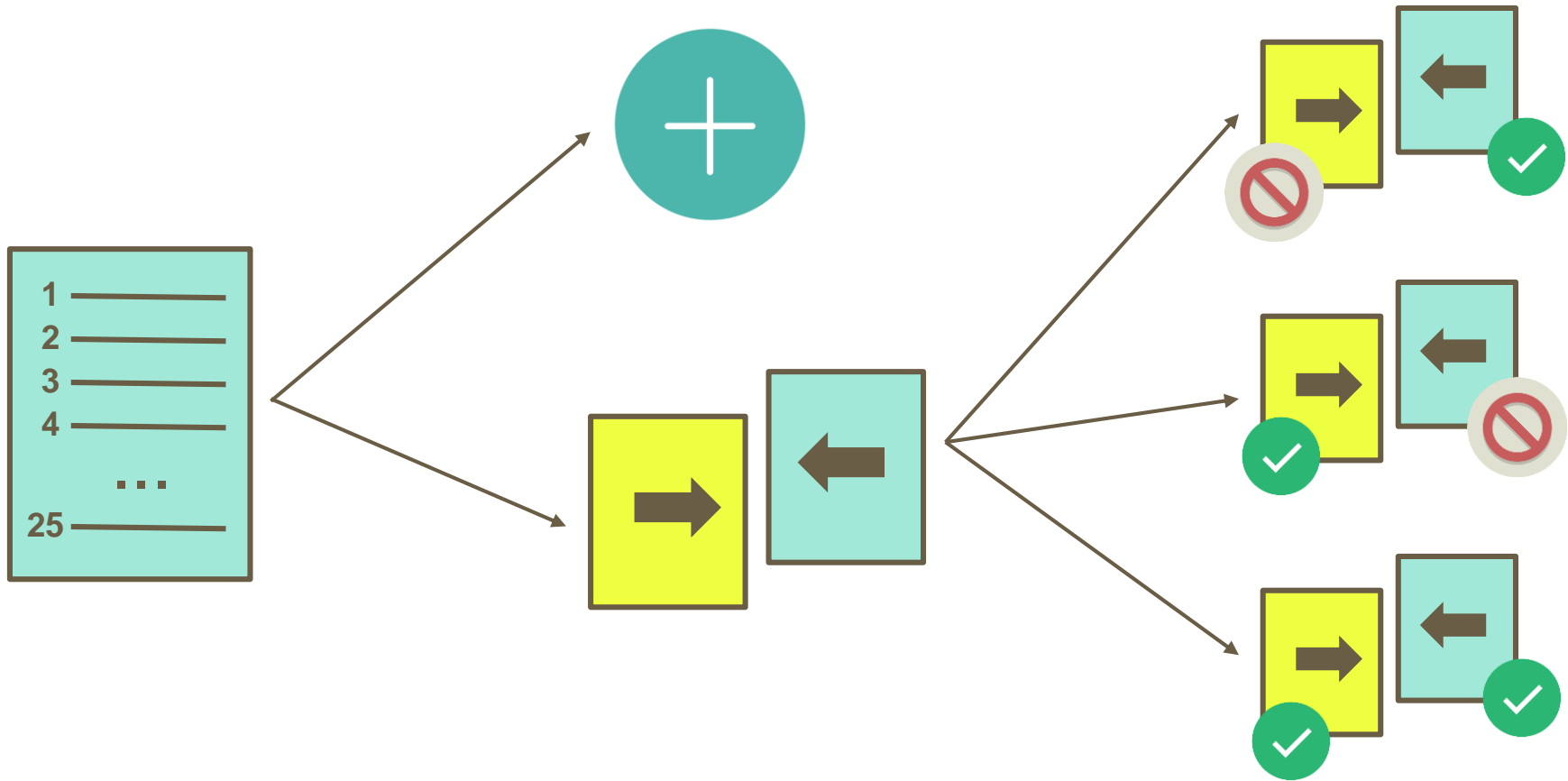

**The Use of Google Maps Geocoding API
and Google Places API Web Service Data
for Automation
of Updating & Matching Processes in SBR**

— BPS - Statistics Indonesia —

Current SBR

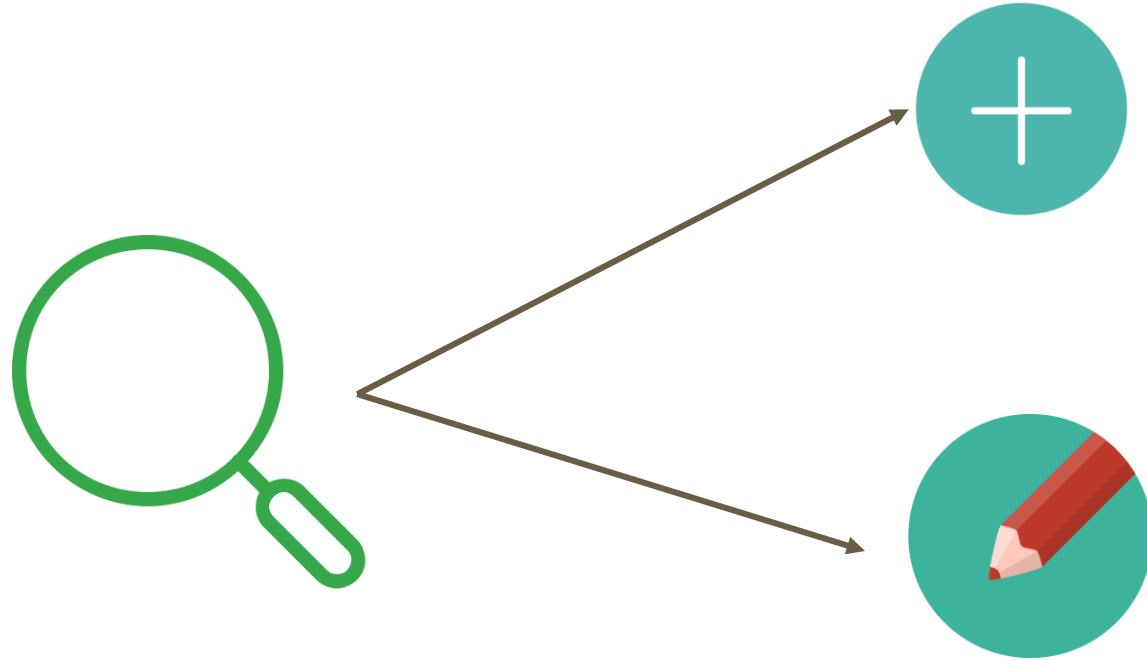




Current Matching Process
(Simplified Version)

Uploading + Matching

Current Updating Process (Batch Process)
(Simplified Version)



Current Updating Process (Non-Batch Process)
(Simplified Version)

Problems

- Current updating and matching operations are resource intensive.
- There are currently only a few people dedicated for SBR.
- SMAs have already had some high burden other than SBR.
- Interns only helps in limited times.

Process
automation
might help



How potential are
Google Maps Geocoding API and
Google Places API Web Service Data
for SBR automation?



Geocoding API

Google

Convert between addresses and geographic coordinates.



Places API

Google

Get detailed information about 100 million places

Source: Google

About Google APIs

The Experiment 1

Updating Automation
with Google API

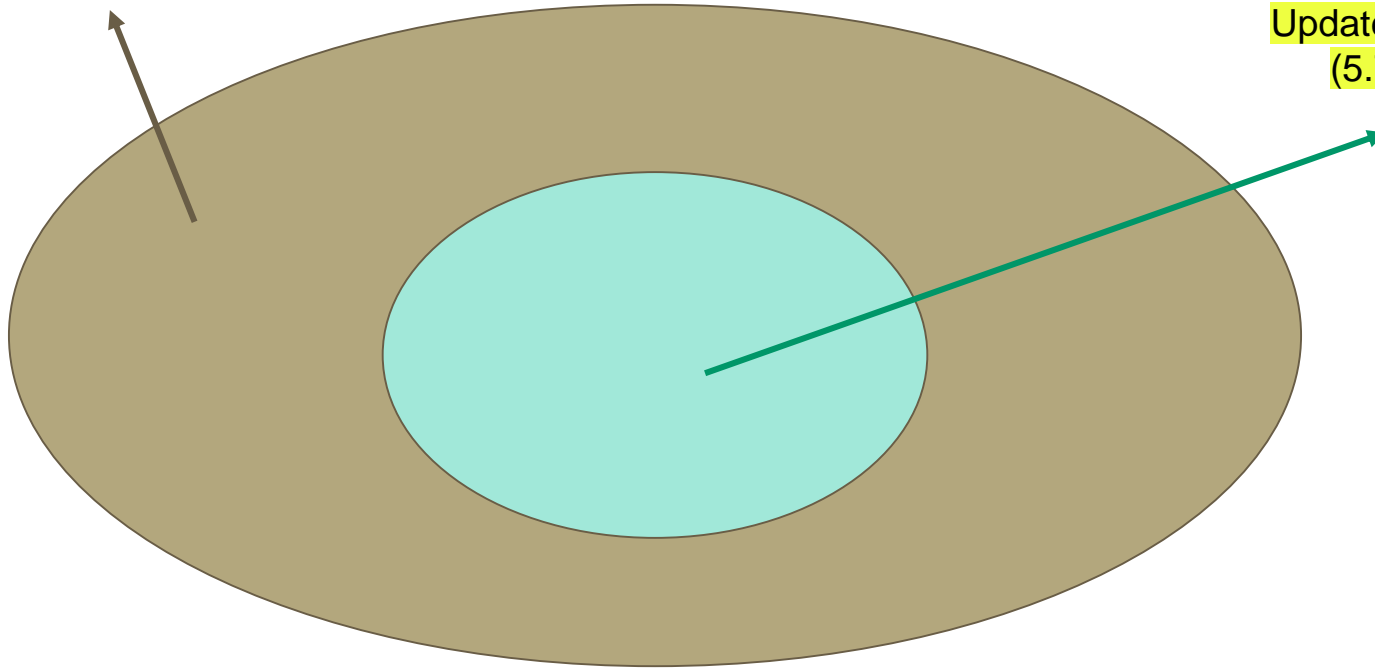
- Knowing number of updates
 - Knowing number of correct updates
 - Finding a way to filter the correct updates
 - Knowing what variables are potential for update
-

SBR Statistical Unit Table Data
(2.4 M)



Geocoding API Places API

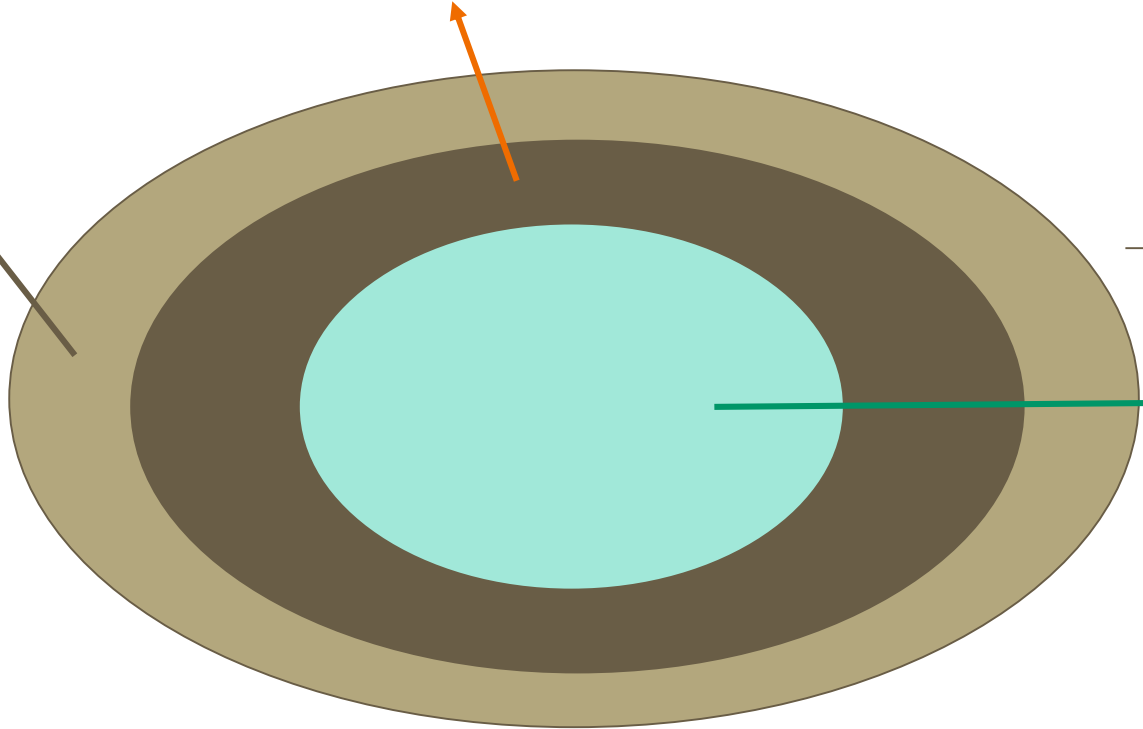
Updated Data
(5.7 K)



The Data between Experiment Stages - SBR Statistical Unit Table

SBR Incoming Unit Data
(88.3 K)

Incoming Unit: Detected as Enterprises
(2.7 K)



Geocoding API Places API

Updated Data
(2.3 K)

The Data between Experiment Stages - SBR Incoming Unit Table

Classification		Number of Entity
Correct Update		32
Incorrect Update	Incorrect Update, but the Correct Place	26
	Incorrect Update and Incorrect Place	43

The number of incorrect update is larger than the correct one.

The Results of Automated Updated Businesses
(Sampled)

Classification		Number of Entity
Correct Update		24
Incorrect Update	Incorrect Update, but the Correct Place	0
	Incorrect Update and Incorrect Place	0

All unit updates are now correct.

The Results of Automated Updated Businesses
Filtered with Names Similarity Constraint per Variable (Sampled)

Variable	Upgraded	Downgraded	Same Positive	Same Negative	Significant Update?*
Name	4	1	27	0	No
Address	7	3	22	0	No
Latitude	31	1	0	0	Yes
Longitude	31	1	0	0	Yes
Telephone	8	1	19	4	Yes
Website	11	2	10	9	Yes
Status	0	0	32	0	No

* assumed that we can differentiate between correct and incorrect results

The Results of Automated Updated Businesses per Variable
(Sampled)

Variable	Upgraded	Downgraded	Same Positive	Same Negative	Significant Update?*
Name	0	0	24	0	No
Address	5	3	16	0	No
Latitude	23	1	0	0	Yes
Longitude	23	1	0	0	Yes
Telephone	6	1	14	3	No
Website	8	2	6	8	No
Status	0	0	24	0	No

* assumed that we can differentiate between correct and incorrect results

The Results of Automated Updated Businesses
Filtered with Names Similarity Constraint per Variable (Sampled)


The Experiment 2

Matching Automation with
Google API

- Knowing what variable combination is the best for the automation
 - Finding another way to improve the automation result
 - Knowing the effect of Place ID in the automation
-

Query	Place ID	Name	Address	Phone	Facsimile	Website	Name & Address	Name & Phone	Name & Facsimile	Name & Website	Address & Phone	Address & Facsimile	Address & Website	Phone & Facsimile	Phone & Website	Facsimil & Website	Industrial Category	2 Digits of ISIC
A	✓																	
B	✓																✓	
C	✓																✓	✓
D	✓	✓	✓	✓	✓	✓												
E	✓	✓	✓	✓	✓	✓											✓	
F	✓	✓	✓	✓	✓	✓											✓	✓


 The update & the original data must have the same values on this variable

 The update & the original data must have the same values on at least one the variables/combination of variables with this background

The Composition of the Matching Queries (1)

Query	Place ID	Name	Address	Phone	Facsimile	Website	Name & Address	Name & Phone	Name & Facsimile	Name & Website	Address & Phone	Address & Facsimile	Address & Website	Phone & Facsimile	Phone & Website	Facsimile & Website	Industrial Category	2 Digits of ISIC
G	✓	✓		✓	✓	✓												
H	✓	✓		✓	✓	✓											✓	
I	✓	✓		✓	✓	✓											✓	✓
J	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
K	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
L	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

 The update & the original data must have the same values on this variable

 The update & the original data must have the same values on at least one the variables/combination of variables with this background

The Composition of the Matching Queries (2)

	Query											
	A	B	C	D	E	F	G	H	I	J	K	L
Correct Match	260	187	152	244	175	140	237	171	136	180	129	109
Incorrect Match	106	30	15	19	8	3	15	7	2	6	3	1

Using the place ID with the combination of industrial category code and two digits of ISIC plus having at least two of the constraints (business name, address, telephone number, facsimile, and website) gave the best results.

The Results of Query Trials for Matching Automation

	Query											
	A	B	C	D	E	F	G	H	I	J	K	L
Correct Match	224	160	132	212	151	121	206	147	117	153	107	90
Incorrect Match	88	24	9	17	6	2	13	5	1	5	2	0

We found a result with zero error by filtering the establishment.

The Results of Query Trials for Matching Automation with Establishment Filtering

Did Place ID Have a Contribution to the Result?

The answer is yes.

	Best Query (Query L)	Control Query (Disregarding Place ID)
Correct Match	90	119
Incorrect Match	0	6

Conclusion

Generally speaking..

The update using Google Maps Geocoding API and Google Places API Web Service data was significantly effective.

Conclusion

Furthermore..

- Latitude and longitude were statistically significant for the update.
- If we can differentiate which API responses are right or wrong, telephone and website updates are also statistically significant.
- Using place ID with combination of industrial category code and two digits of ISIC plus having at least two of five constraints (name, address, telephone number, facsimile, and website) gave the best results for enterprise matching automation.