

Big data for the Household Budget statistics (HBS)

Gains and lessons learned

Knut Linnerud, Statistics Norway (knl@ssb.no)
Kristin Egge-Hoveid, Statistics Norway (keg@ssb.no)

Agenda

- Background
- Big data sources for Household Budget Survey (HBS) – very good test results!
- Timeline HBS project 2012-2023
- Prerequisites for collecting and using big data
- Privacy issues and complaints

Why utilize big data for HBS?



- HBS statistics not published since 2012
- Great need for modernisation



- Paper diary
- Great response burden
- Manuel coding
- Sample bias and underreporting
- Falling response rates



- What was the potential of transaction data from stores and banks?

Mat- og drikkevarer i butikk

Her registrer du alle mat- og drikkevarer som er kjøpt i butikk, bensinstasjon, kiosk eller liknende. Det er viktig å beskrive varen godt. Vi ønsker for eksempel at det noteres hva slags type brød som er kjøpt: grovbrød, loff, helkornbrød, kneippbrød eller andre typer brød. For andre matvarer er det viktig å vite om de er ferske, dyperfrysede, røkte, saltet, tørket e.l. For fisk og vilt skriver du om vekten gjelder før eller etter fisken/viltet er renses/slaktet.

Hvis varen ikke er merket med vekt, må du om mulig veie varen selv eller anslå vekten. Hvis det er vanskelig å gjøre anslag: Beskriv varen så godt det lar seg gjøre (antall, varemerke, produsent med videre.) Husk å få med alt fra melk, brød og mineralvann til alkoholholdige drikkevarer.

Egenproduserte matvarer

Her registrerer du også egenproduserte matvarer, det vil si matvarer som kommer fra eget jordbruk eller egen hage, ville bær og sopp som man plukker selv, fisk fra eget yrkes- eller sportsfiske, kjøtt fra jakt på vilt og liknende. For slike varer trenger du ikke oppgi pris.

Se for øvrig eksemplene nedenfor.



Dato	Hva slags vare ble kjøpt? Beskriv varen.	Kryss av hvis varen er kjøpt		Menge	Hva kostet varen	
		i utlandet	på internett		Kroner	Øre
Eks.1	2/1 Lammesteik, fersk			2,3 kg	269	90
Eks.2	2/1 Grovbrød			800 g	26	50
Eks.3	5/1 Joröbær selvpilket			4 kg		
Eks.4	6/1 Laks, røkt			250 g	39	90
Eks.5	9/1 Skummet melk			1 liter	10	90
Eks.6	13/1 Rødvin, kartong (kjøpt i Sverige, svenske kroner)		X	3 liter	149	00
1	Frokost skunke			?	29	90
2	Smoothie			0,3	17	90
3	Epler			?	23	90
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

What data have we explored?



- Receipt data from four grocery chains (98 per cent market share)
 - 2,4 million receipts per day



- Payment transactions from NETS Branch Norway
Debit card via BankAxept. Credit card/cash purchases not included
 - 1,6 billion transactions per year

Is it possible to use transaction data?

- Receipt data/bank transaction data can be linked by time, location and sum!
- Linked more than 70 per cent of all grocery purchases (receipts) to persons and households through debit card transactions
- Link to persons and households via bank account and households register
- Mostly for Food and non-alcoholic beverages (COICOP 01)

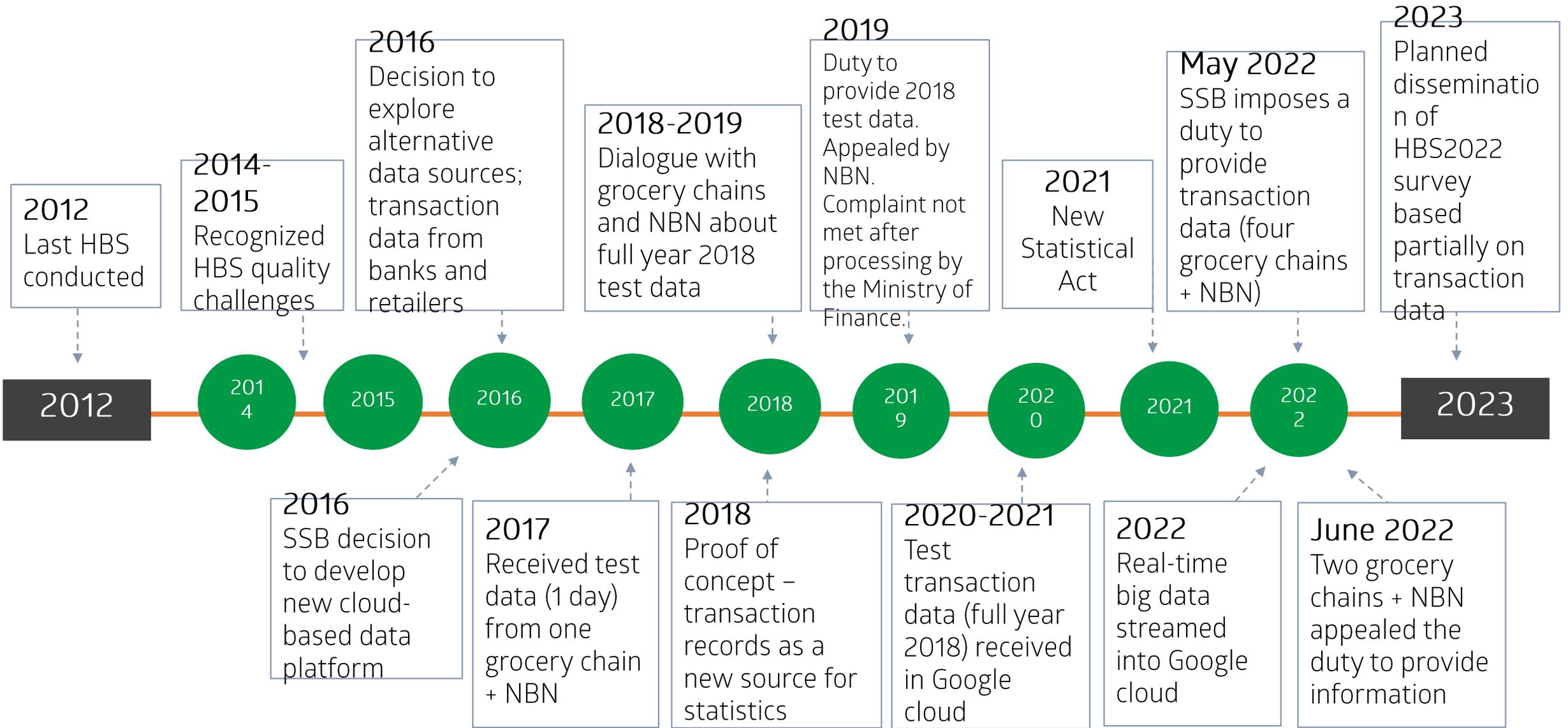


Benefits of using transaction data



- Data in large quantities over the whole year
- Effective data transfer through real time streams
- Data for the whole population
- Improved quality
- Reduced response burden for COICOP 01
- Reduced costs compared to surveys
- Potential to combine with register data and survey data

Long and time-consuming process...





SSB vil ha matkvitteringen din – dagligvarekjeder protesterer

Norgesgruppen mener SSBs ønske om innsamling av «bongdata» fra dagligvarekjedene utgjør en form for systematisk overvåking



Samfunn SSB krever å få vite nøyaktig hva nordmenn kjøper i matbutikken

Skrevet av Martin Gundersen 28. mai 2022 146



2,4 MILLIONER KVITTERINGER: SSB krever å få flere millioner kvitteringer daglig fra matvarekjedene. Foto: Martin Gundersen/NRK

– Hvor detaljerte kart over våre liv trenger statsapparatet egentlig for å ta gode beslutninger? spør NTNU-forsker.



Aktuelle nyheter 2022

Vil føre kontroll med Statistisk Sentralbyrå

Statistisk sentralbyrå (SSB) har besluttet at dagligvareaktørene Norgesgruppen, Coop, Rema og Bunnpris skal levere ut bongdata om sine kunder. Datatilsynet vil nå gjennomføre en kontroll med virksomheten.

– Vi vil følge opp saken, og ber om en nærmere redegjørelse om hvilke vurderinger som er gjort, sier konstituert direktør i Datatilsynet Janne Stang Dahl.

Bongdata er i praksis kvitteringen som viser hvilke varer hver enkelt kunde har kjøpt, og i hvilket omfang. Ved å koble en bong til en betalingstransaksjon, kan SSB knytte varekjøp til person og husholdning via data fra Skatteetaten og folker registret. SSBs vedtak vil innebære en utlevering av data om alle som betaler dagligvarer med bankkort. Dette kan også omfatte personer under 18 år med eget bankkort.

Samfunn

Miljøpartiet ut mot økt datafangst: – SSB trenger en realitetsorientering

Skrevet av Martin Gundersen 13. juni 2022 70



Norway DPA investigates statistics agency

Bethan John
06 June 2022

Print article



credit: Shutterstock.com/SiljeAO

Stopped receiving data in June 2022 pending the complaint from the grocery chains and NETS

Prerequisites to access privately held big data



- The statistics act
- Thorough cost-benefit assessments
- Technical infrastructure
- Good internal coordination

The statistics act

- Authorizes Statistics Norway to impose a *duty to provide information* for any data, public or private, that may be necessary, to develop, produce or disseminate official statistics.
- The statistical need must be anchored in the national statistical program adopted by the Government.
- Emphasis placed on adapting the new Act to also reflect the ever-increasing amount of data in society and the need to utilize new types of data



Cost-benefit assessment

Before the duty to provide information can be imposed, SSB must conduct a thorough cost-benefit assessment of the planned data collection.

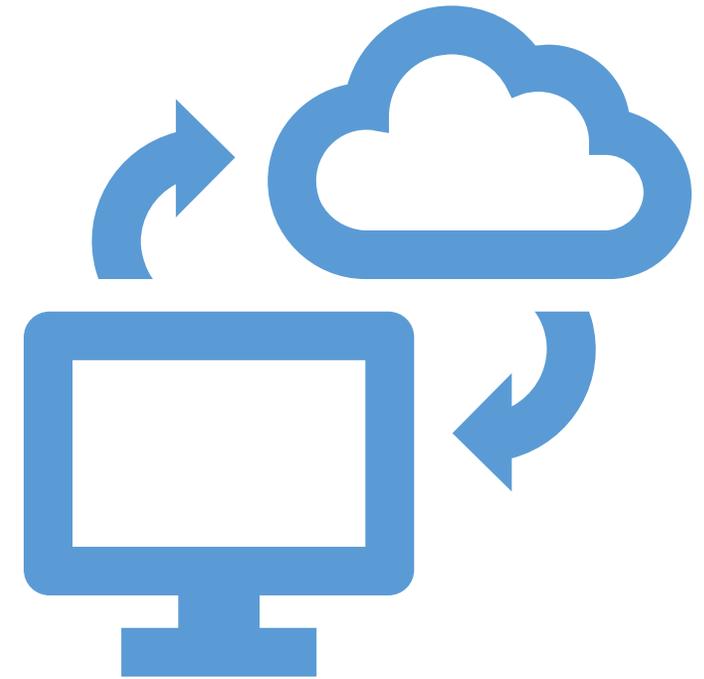
The assessment must include:

- the purpose of collecting the data
- Justification of the benefits of collecting the data
- The cost for those required to provide information, including privacy costs for individuals



Technical infrastructure

- Large amount of data (billions of transactions) requires sufficient capacity for storage and processing
- Google Cloud Platform in SSB
- New tool landscape (R, Python and Jupyter notebooks)
- Requires building new competence
- Challenge to access competent staff to process data with code



Privacy issues

- The use of privately held big data with personal identifiable information raises privacy issues.
- Two chains and Nets appealed the imposed duty to deliver data.

Main objections:

- Privacy concerns, or lack of respect for privacy
- Questioning data security, could data be stolen or shared?
- Requirement for considerable data minimization.

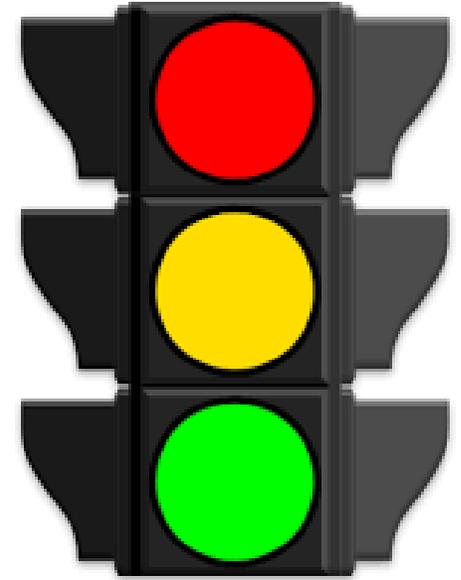


How can we meet these objections?

- Concerns from companies and private persons must be seriously met
- A prepared communication strategy recommended
- Improved methods for protecting privacy and data security
- Pseudonymization and aggregating data to groups

Next steps

- Processing of appeals
- Public debate welcome
- Establish better understanding of SSBs safety and privacy routines
- Accept if the decision gets in our disfavour
- Further improve methods for data minimization and confidentiality



Back up in a full scale survey in 2022