

# Generering av endringslogger for åpne geografiske datasett

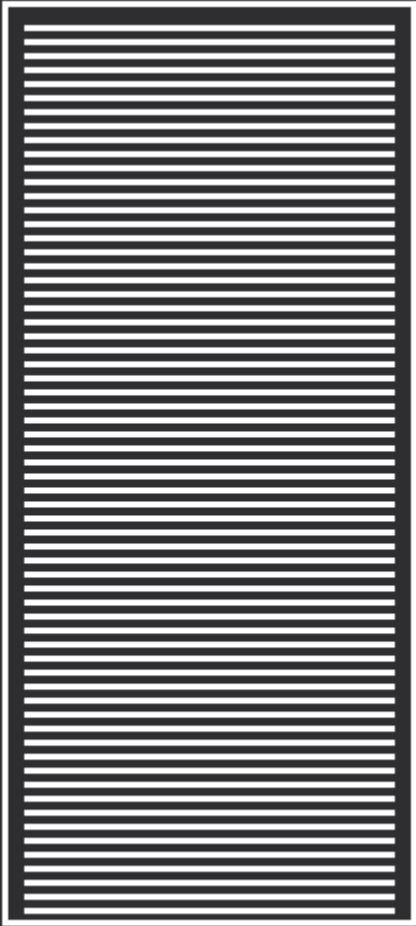
Deriving per-feature change queues from bulk-updated open geospatial datasets

Atle Frenvik Sveen  
Norkart AS / NTNU  
atle.sveen@norkart.no

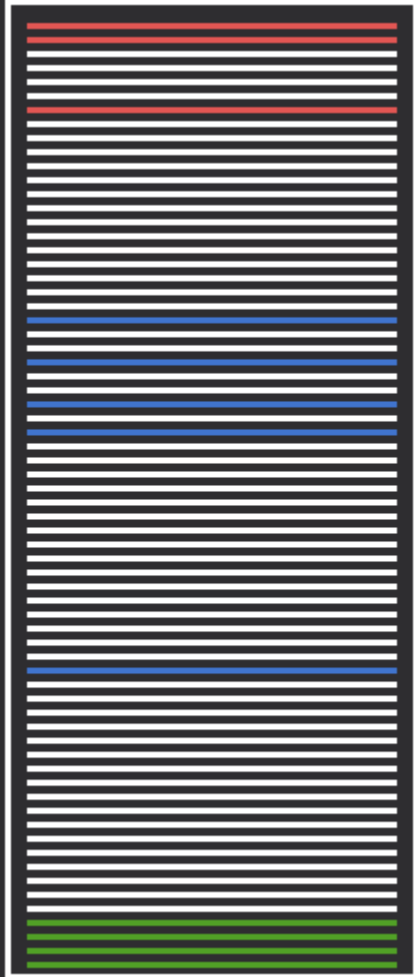
“There should be no need to redistribute an entire data set to its users to propagate changes that are only minor or few in number”

Cooper and Peled (2001)

Versjon 1 (n)



Versjon 2 (n+1)



Sletta



Endra



Ny

# BURGER KING WHOPPER

ADVERTISEMENTS



ACTUAL WHOPPER



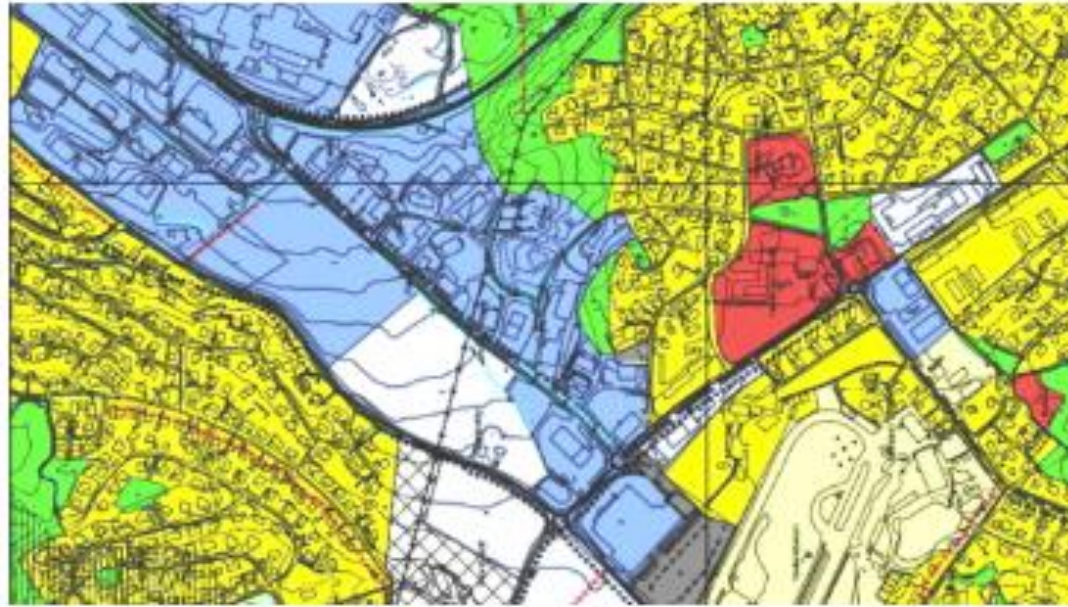
# Case in point



## Det offentlige kartgrunnlaget

Oppdatert: 03.09.18

Det offentlige kartgrunnlaget (DOK) er offentlige geografiske data som er tilrettelagt for kommunenes plan- og byggesaksarbeid. Formålet med det offentlige kartgrunnlaget er å sikre en kunnskapsbasert og effektiv planlegging og saksbehandling.



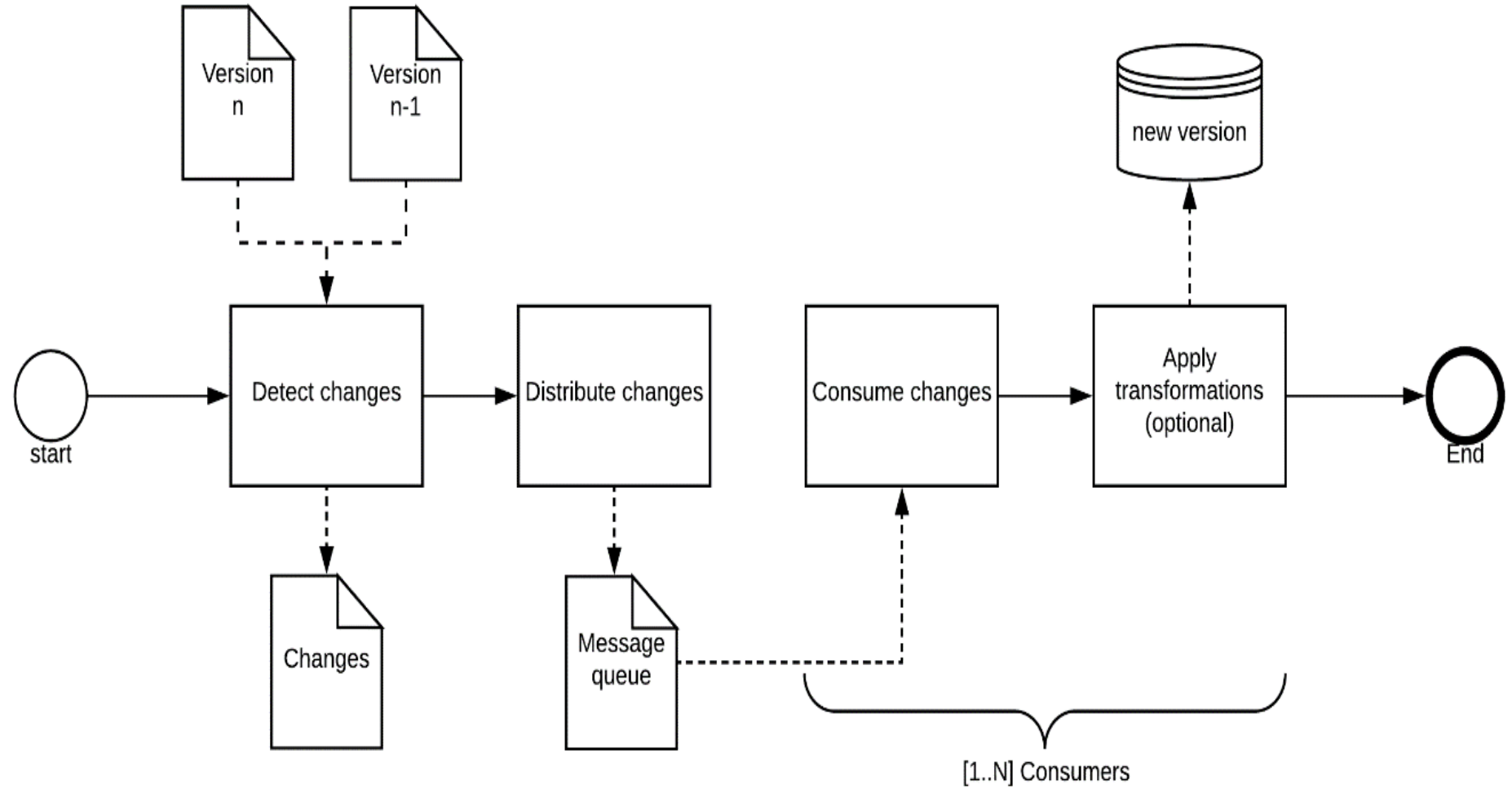
*KARTGRUNNLAG. Det offentlige kartgrunnlaget skal tilrettelegge for plan- og byggesaksarbeid. Utsnitt av arealplan.*



# Løsning?



# Idé



# 1. Endringsdeteksjon

- Hva er nytt?
- Hva er endret?
- Hva er slettet?
- Hva er uendret?



# Endringsdeteksjon – med id

Versjon n-1

Attrs	Geom
Id: 1	[[1, 1], [1, 2], [1,3]]
Id: 2	[[1, 1], [2, 1], [3,1]]
Id: 3	[[1, 1], [2, 2]]
Id: 4	[[4, 5], [4, 8]]
Id: 5	[[4, 5], [4, 8]]
Id: 6	[[2, 2], [4, 4], [8, 8]]

Versjon n

Attrs	Geom
Id: 1	[[1, 1], [1, 2], [1,3]]
Id: 2	[[1, 1], [2, 1], [3,1]]
Id: 3	[[1, 1], [2, 2]]
Id: 5	[[5, 5], [5, 8]]
Id: 7	[[5, 4], [8, 1], [8, 2]]
Id: 8	[[0, 0], [2, 4]]

1	Uendret
2	Uendret
3	Uendret
4	Slettet
5	Endret
6	Slettet
7	Ny
8	Ny

# Endringsdeteksjon – uten id

Versjon n-1

Attrs	Geom
	[[1, 1], [1, 2], [1,3]]
	[[1, 1], [2, 1], [3,1]]
	[[1, 1], [2, 2]]
	[[4, 5], [4, 8]]
	[[4, 5], [4, 8]]
	[[2, 2], [4, 4], [8, 8]]

Versjon n

Attrs	Geom
	[[1, 1], [1, 2], [1,3]]
	[[1, 1], [2, 1], [3,1]]
	[[1, 1], [2, 2]]
	??
	[[5, 5], [5, 8]]
	[[5, 4], [8, 1], [8, 2]]
	[[0, 0], [2, 4]]

Slettet?

Ny?

Ny?

## 2. Endringsdistribusjon

- Format
  - Hvordan uttrykker vi en endring?
    - «Diff»
- Teknologi
  - Hvordan distribueres endringen?
    - «event sourcing»

# Diffing – Forskjeller

atle -> Atle

Gammel verdi: atle

Ny verdi: Atle

=> Mer data enn  
man trenger?

POINT(1 1) -> POINT(1 2)

Gammel verdi: POINT(1 1)

Ny verdi: POINT(1 2)

Brukes av:

- OpenStreetMap
- GeoGig

# Diffing – diff-match-patch

atle -> Atle

@@ -1,4 +1,4 @@

-a

+A

tle

POINT(1 1) -> POINT(1 2)

@@ -2,10 +2,10 @@

POINT (1

-1

+2

)



left.json

Swap Clear examples

right.json

```
{
  "type": "Feature",
  "properties": {"type": "test"},
  "geometry": {
    "type": "Point",
    "coordinates": [
      0,
      1
    ]
  }
}
```

```
{
  "type": "Feature",
  "properties": {"type": "test2"},
  "geometry": {
    "type": "Point",
    "coordinates": [
      1,
      1
    ]
  }
}
```

Delta  Visual  JSON  JSON (Annotated)

```
{
  "geometry": {
    "coordinates": {
      "_t": "a",      Array delta (member names indicate array indices)
      "_0": {
        [
          0,
          0,         remove index 0 ([previousValue, 0, 0])
          0
        ]
      },
      "0": {
        [
          1         insert at index 0 ([newValue])
        ]
      }
    }
  },
  "properties": {
    "type": {
      [
        "test",     modify property "type" ([previousValue, newValue])
        "test2"
      ]
    }
  }
}
```

# Diffing på binær-nivå?

Versjon 1:  $[[0, 1], [1, 1], [2, 1]]$

Versjon 2:  $[[0, 1], [0, 1], [2, 1]]$

Forskjell:  $[[0, 0], [-1, 0], [0, 0]]$  <- må kunne uttrykkes kompakt?

WKB, TWKB?







# Utfordringer

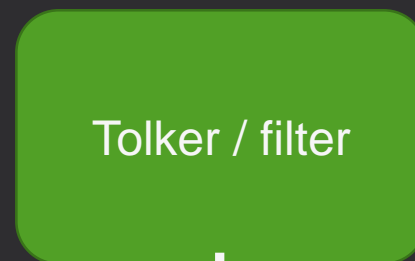
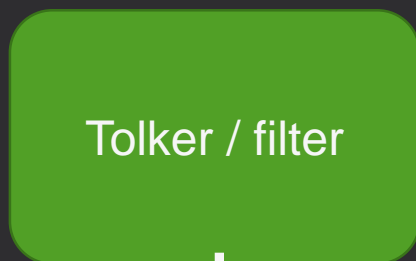
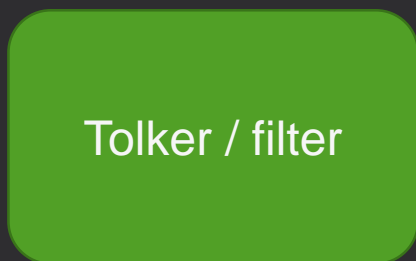
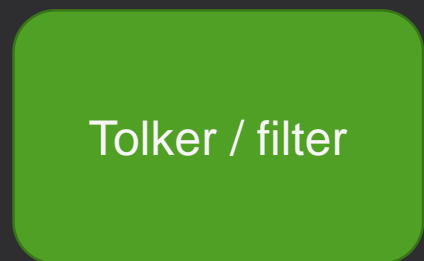
- En event pr feature?
- Kan havne i «mellomstate»
- Hva med nye abonnenter?
- «spol fremover fra start»
- Checkpoints
- Mye data?
- Tja..?

salg

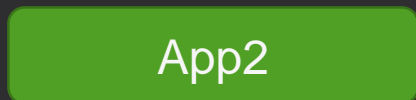
1971	Kassa		kredit. for.			Varekjøp		Pliktig	Fritt	kunder		Bank 723		Smle	Pile
	DEBET A <sub>1</sub>	KREDIT A <sub>2</sub>	DEBET B <sub>1</sub>	KREDIT B <sub>2</sub>		DEBET C <sub>1</sub>	KREDIT C <sub>2</sub>	DEBET D <sub>1</sub>	KREDIT D <sub>2</sub>	DEBET E <sub>1</sub>	KREDIT E <sub>2</sub>	DEBET F <sub>1</sub>	KREDIT F <sub>2</sub>	DEBET G <sub>1</sub>	KREDIT G <sub>2</sub>
Beholdning															
Motorverktø		1070													1070
Wenvertge	7.000	184	87590	87590	83520	83520	21	69600							87590
Øfo - Bukt øls		185					13	75000							2000
Karnt Sparb	28500	186													249825
Alf. Hærrid	23/2	187		52800			19								52800
Øls Hærrid		189					1	10255							
Guinar T. Hærrid							4								120379
Øls Hærrid		191					7	24000							
Øls Hærrid		192					5	27000							
Øls Hærrid		193					7	111221							
Øls Hærrid		194					24								400200
Øls Hærrid		196					9								200
Øls Hærrid		197					21	78340							340532
Øls Hærrid		198					25	649950							779940
Øls Hærrid		12490	12490				8								
Øls Hærrid		3500		3500			2								
Øls Hærrid		11800	11800				2		9900						
Øls Hærrid		33240	33240				2	27700							
Øls Hærrid		4737	4737				2	1579							
Øls Hærrid		2900		2900			2								
Øls Hærrid		21158	21158				2	17634							
Øls Hærrid		20130													
Øls Hærrid		70544					3	58788							
Øls Hærrid		100000													100000
Øls Hærrid		100000													100000
Øls Hærrid		311750													311750
Øls Hærrid		200000													200000
Øls Hærrid			29568				9	24640							

### 3. Endringskonsumpsjon



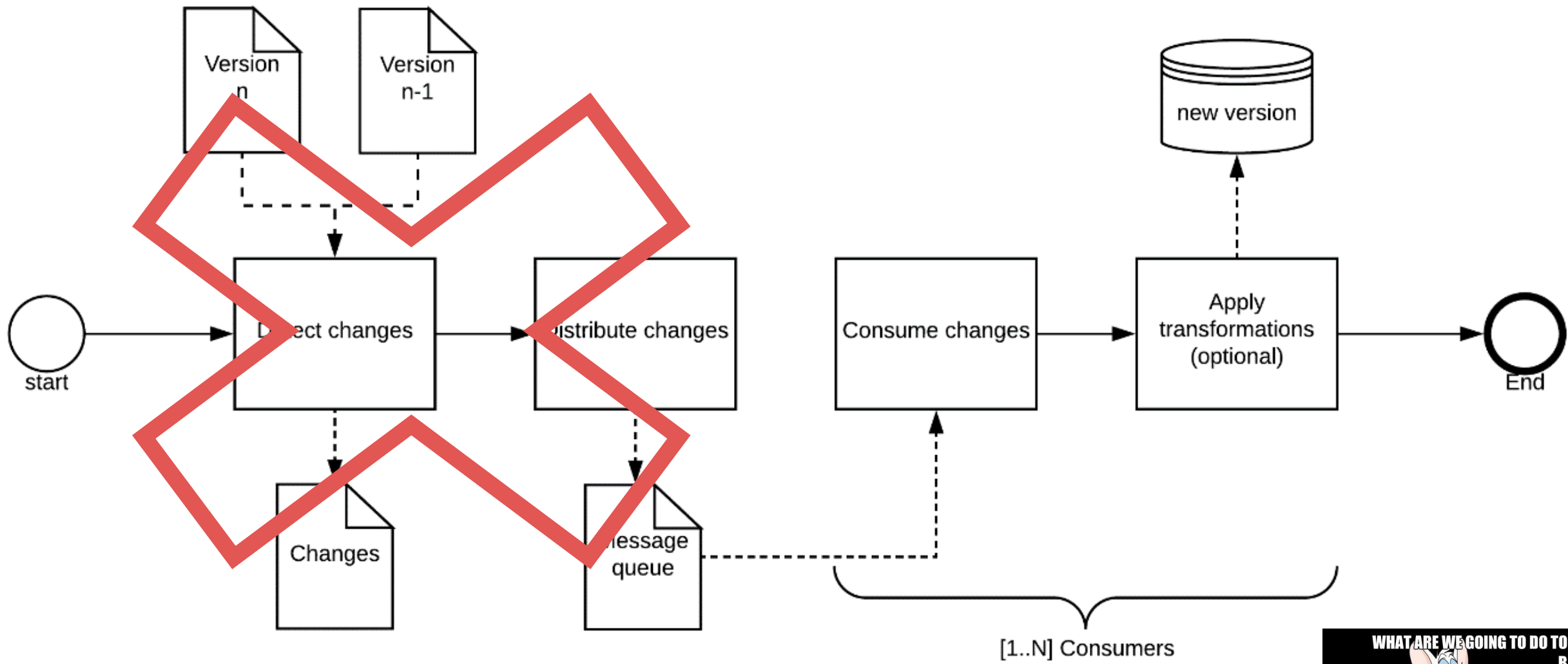


+++



# Jammen?!?

Hva om data-eier begynner å distribuere endringer?



# Takk?

 NTNU

Norwegian University of  
Science and Technology

NORKART 