



SUNET

# From Public Cloud to National Infrastructure

Lessons from Sunet Scribe

TNC 26 – Helsinki – 2026

Erik Lundberg, coordinator Sunet Media Services

# The Challenge

Research and education institutions have large amounts of recorded lectures, interviews, seminars and research material that need transcription.

Many institutions in the R&E sector do not trust commercial providers with sensitive data.

Most commercial speech-to-text services are not optimized for Nordic languages.

# Can Sunet take over?



## ITCF Prototype

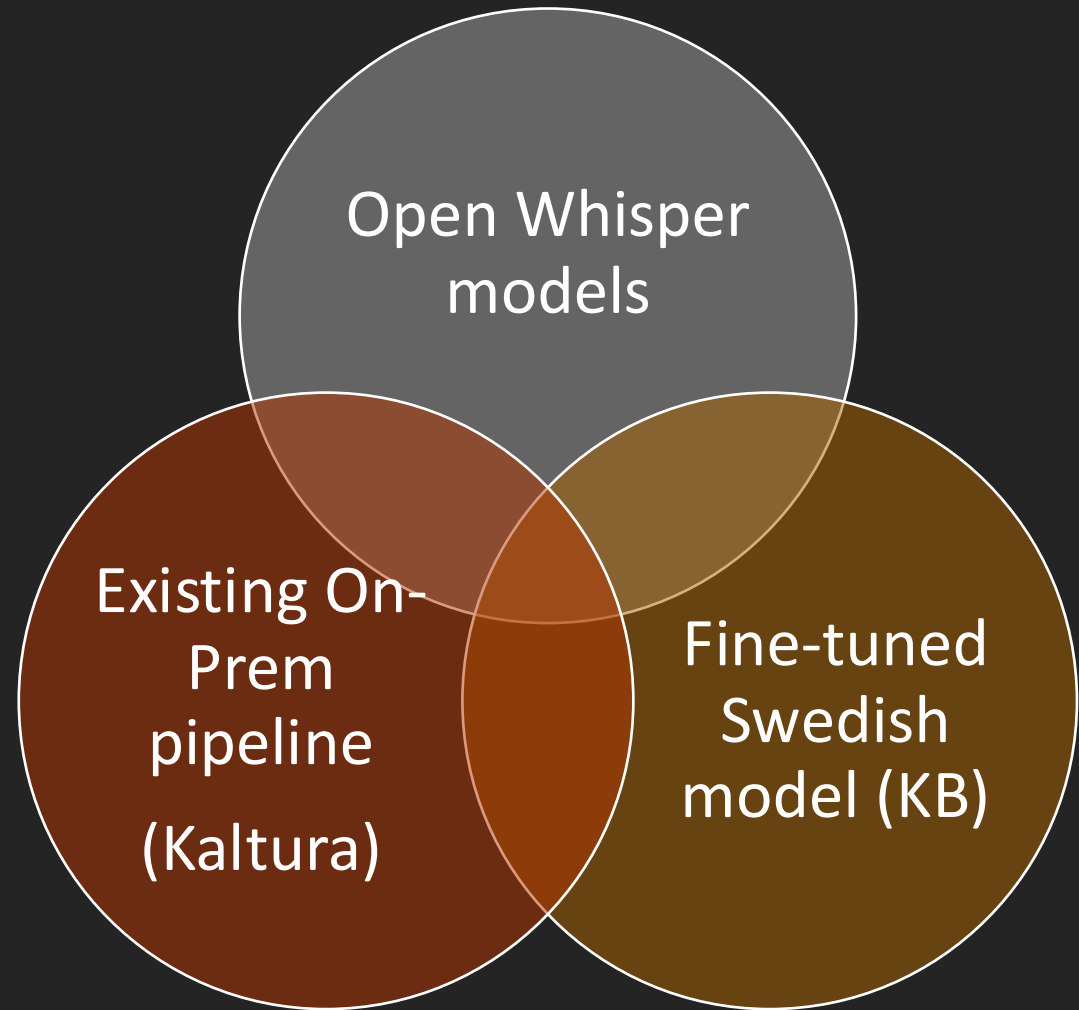
- Azure AI Speech CLI
- Container based Microsoft solution
- English + Swedish
- No functional speaker diarization



## Why we said no

- Heavy Microsoft dependencies
- Outside Sunet's operational model
- Required skills we didn't have in house
- No clear long-term direction


# A different path became possible




# Scale beyond the edge cases

## Two parallel services

Commercial STT service  
General purpose  
Teaching  
Communication



Small internal "secure" STT  
Sensitive material



## One operational platform

Broad internal STT service  
General purpose  
Teaching  
Communication  
Sensitive material  
Stand alone and integrated service



Parallel operational pipelines  
Fragmented adoption  
Full development overhead  
Separate business models

✓ One service ecosystem  
✓ Shared support structure  
✓ Shared development & scaling  
✓ Unified business model



# Usage and Traction

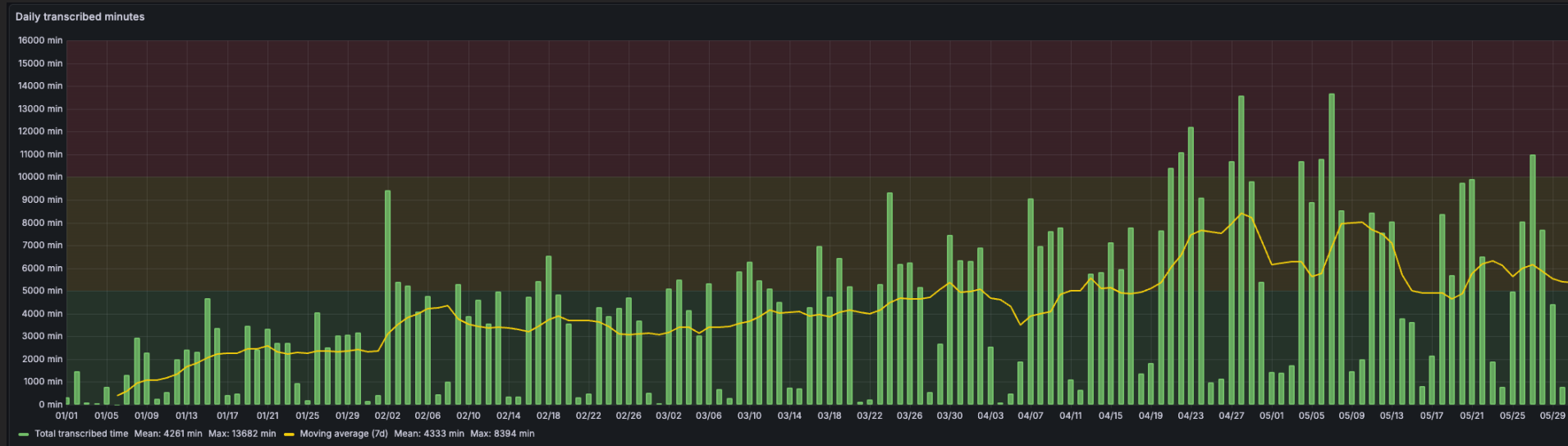
30 onboarded universities

3,000 users

~650,000 min transcribed 2026

3-4 x usage vs 2025

Handled by 3 GPU servers



# Designed for security, integration and evolution

## Security



- Self-hosted
- Federated login
- Zero-access model
- Auto retention

## Integration



- Kaltura
- NextCloud (2026)
- Published API

## Evolution



- Open source
- Modular architecture
- Swappable models and components

# From media to usable text

### Export transcript

**Format**  
Text (.txt)

**Timestamps**  
 Include timestamps  
Timestamp range: Start & End  
Format: SRT (00:00:00,000)

**Text options**  
 Include speakers  
 Include block numbers  
Separator: Double newline

**Preview**

```
(00:00:08,300 - 00:00:14,919) Speaker_01:
Välkomna till det 43 avsnittet av sociologipodden med dig Åsa Wettergren.

(00:00:14,919 - 00:00:16,920) Speaker_02:
Och med dig Håkan Thorn.

(00:00:16,920 - 00:00:45,920) Speaker_01:
Idag gästas vi av Lotte Schack som nyligen disputerade i sociologi här på Goteborgs universitet med en avhandling som hetar Children of the crisis, political subjectivity, responsibility and justice in the Swedish climate movement. Blir det ungefär krisens barn, politisk subjektivitet, ansvar och rättvisa i den svenska klimatrörelsen.

(00:00:45,920 - 00:01:00,920) Speaker_02:
Just det. Och senare ska jag presentera en bok idag som heter Passion och paranoia på svenska. Om akademins emotionskultur som är skriven av en sociolog som heter Charlotte Block.

(00:01:00,920 - 00:01:35,799) Speaker_01:
Spännande. Men först alltså Lottes avhandling. Tidmässigt då så fokuserar den ju egentligen på den perioden, kan man väl säga, i klimataktivismen som började 2018-2019. Goran Terborn som vi hade här tidigare i en podd, han gjorde observationen att den här rörelsen då, 2018-2019, som ju egentligen leddes av Fridays for Future, det var den snabbast växande sociala rörelsen i världshistorien.

... 260 more captions
```

Total 265 | Showing 5

File: 43. Framtidens barn (MP3 Audio).txt

[CLOSE](#) [EXPORT](#)



Upload



Transcribe



Edit



Export

# Questions?