

TDM in Sweden – a few practical notes



Eric Luth, Wikimedia Sverige
1/10 2024

Structure

1. Short background
2. Swedish TDM implementation
3. Practical issues with TDM implementations
4. Proposed new provision to deal with issues
5. Practical issues that new proposed provision will not solve



- Researchers have mined for years.
- The National Library is the legal deposit. It has all kinds of materials since centuries.
- Their collections are the most valuable collections for TDM research on Swedish language text, audio and film
- With little or no basis in Swedish or EU copyright law, National Library started to develop AI models for research and a Swedish LLM



- The EU CDSM provisions on TDM took effect 1/1 2023 in Sweden.
- This was long awaited in research communities.
- However already on new year's day, issues with the new law became apparent.
- To understand these issues, we need to look into the way that the provisions are implemented.

Framställning av exemplar för text- och datautvinning

15 b § Forskningsorganisationer, sådana bibliotek och museer som är tillgängliga för allmänheten, arkiv samt institutioner för film- eller ljudarvet får **framställa exemplar** av verk som de har **lovlig tillgång till**, dock inte datorprogram, för att utföra text- och datautvinning för **forskningsändamål**. Exemplaren får inte behållas längre än vad som är nödvändigt för ändamålet och får inte användas för andra ändamål. Exemplaren ska lagras på ett sätt som hindrar obehörig användning.

Första stycket hindrar inte att upphovsmannen vidtar proportionerliga åtgärder för att säkerställa integritet och säkerhet i nätverk och databaser som innehåller verk.

Avtalsvillkor som inskränker rätten att använda verk enligt denna paragraf är ogiltiga. Lag (2022:1712).

Making copies for text and data mining

copies

research purposes

**make
lawful access**



Practical issues

- Only copying, no CTPP
- What is lawful access in regards to legal deposits?
- No remote usage for universities
- Unclear on cross border usage
- How long can copies be stored?
- Blurry lines on commercial use

Proposals in SOU 2024:4



Användning av verk i vetenskaplig forskning

15 d § Forskningsorganisationer och kulturarvsinstitutioner får **tillgängliggöra verk** som de har lovlig tillgång till, dock inte datorprogram, om det sker

1. för forskningsändamål **till deltagare i ett vetenskapligt forskningsprojekt** som forskningsorganisationen eller kulturarvsinstitutionen också ingår i, eller
2. **För verifiering** och granskning av resultatet av vetenskaplig forskning som forskningsorganisationen eller kulturarvsinstitutionen har utfört.

Use of works in scientific research

Section 15 d. Research organisations and cultural heritage institutions may make works to which they have legal access **available to the public**, except computer programs, if this is done

1. for research purposes to **participants in a scientific research project** in which the research organisation or cultural heritage institution is also involved, or
2. **for the verification** and examination of the results of scientific research carried out by the research organisation or cultural heritage institution.

Problems in the proposal

- CTPP exception 15 d § very narrow. Solves two problems, but there might be more.
 - Why not verbatim introduction of 5.3.a in InfoSoc?
- More focus on institutions than researchers.

Still unclarity on what lawful access means.
- Few solutions for cross border or digital usage.

Other issues



- Difficult to get access to data from social media platforms
 - GDPR makes TDM research difficult in many cases
 - Ethical principles vary greatly across EU countries
 - CMOs trying to argue that TDM \neq AI
- Digital Knowledge Act

Question

- Do you know of Finnish examples where cultural heritage institutions have restricted access to their material for TDM purposes?
 - Either through the opt out mechanism in CDSM art. 4
 - Or as part of out of commerce licenses
- Send to eric.luth@wikimedia.se

TACK!



WIKIMEDIA
SVERIGE