Nordic museology 2022 • 2, s. 73–96

Tema kulturarv

Temamodeller om ABM i svenska statliga offentliga utredningar 1945–89

Pelle Snickars

Title: Topic Cultural Heritage. On Topic Models of ALM in Swedish Government Official Reports 1945–89

Abstract: *In an explorative manner this article uses a data driven digital history set-up to focus on broad cultural heritage issues in Sweden during the second half of the 20th century. By distant reading and topic modeling a dataset of 3100 Swedish Government Official Reports (SOU) between 1945–89, the article gives a new perspective of how the Swedish state examined and discussed national cultural heritage as well as archival, library and museum (ALM) issues. Topic modeling is a computational method to study themes or discourses in texts by accentuating words that tend to co-occur and together create different topics. Via a computational interrogation of the dataset in a Jupyter Lab-environment a number of heritage topics related to the ALM-sector can be detected. They include the most common words for each heritage topic, but also reveal temporal periodisations when archival, library and museum issues were mostly discussed as well as other societal topics in the dataset that heritage was related to.*

Keywords: digital history, Swedish Government Official Reports, cultural heritage, ALM-sector, digital methods, topic modeling, digital humanities.

I november 2016 lanserade den svenska regeringen forskningspropositionen *Kunskap i samverkan – för samhällets utmaningar och stärkt konkurrenskraft* (Prop. 2016/17:50). Där presenterades den nationella forskningspolitikens inriktning i ett längre, tioårigt perspektiv, med särskilt fokus på satsningar i närtid. Ambitionen var (som ofta) att Sverige skulle vara en ledande kunskapsnation och ett av världens främsta forsknings- och innovationsländer. Och i sedvanlig stil kunde man inte läsa specielt mycket om humaniora – och knappt något alls om arkiv-, biblioteks- och museisektorn (ABM). Det enda stället i forskningspropositionen där det stod något om humaniora och ABM handlade om förbättrade förutsättningar för så kallad datadriven forskning.

> För att främja datadriven forskning […] finns ett behov av att digitalisera material som i dag finns vid kulturarvsinstitutionernas arkiv, samlingar och bibliotek. Sådant material utgör en resurs vars

> forskningspotential till stor del är outnyttjad då bara en bråkdel är digitaliserat. En digitalisering av materialet vid kulturarvsinstitutionernas arkiv, samlingar och bibliotek ökar tillgängligheten och skapar nya möjligheter för forskning […] En ökad tillgång till stora digitala datavolymer öppnar för att besvara nya frågeställningar och för metodutveckling (Prop. 2016/17:50:95).

För detta ändamål aviserade forskningspropositionen en ökning av det svenska Vetenskapsrådets anslag med 40 miljoner kronor för att, kort och gott, "stödja utvecklingen mot datadriven forskning" (Prop. 2016/17:50:95).

Dessa medel resulterade så småningom i att Vetenskapsrådet – i samarbete med Riksbankens Jubileumsfond och Kungliga Vitterhetsakademien (bägge centrala humanistiska forskningsfinansiärer i Sverige) – i två omgångar kommit att lysa ut forskningsmedel till Projektbidrag för digitalisering och tillgängliggörande av kulturarvssamlingar, i vad som kommit att kallas för Digarvprogrammet. I dagsläget är 14 större forskningsprojekt finansierade inom denna satsning.[1]

Lyfter man blicken från forskningspolitik och medelstilldelning kan den svenska regeringens ambition att främja datadriven humanistisk forskning med inriktning på digitaliserade kulturarvssamlingar betraktas som en sorts önskan att studera det förflutna via ny teknik. Genom att använda modern medieteknik (digitalisering), liksom sofistikerade digitala metoder, var den uttalade förhoppningen att forskare skulle få möjlighet att analysera, ja rentav förstå det förflutna på ett annorlunda – och kanhända helt nytt sätt. Den här artikeln är skriven inom ramen för ett av de forskningsprojekt som ingår i Digarvprogrammet, Välfärdsstaten analyserad. Textanalys och modellering av svensk politik, media och kultur 1945–89.[2] Det är ett projekt som leds av mig och som använder storskaliga textuella dataset för att analysera tre centrala sfärer i det svenska välfärdssamhällets framväxt: politikens, mediernas och kulturens. Svensk efterkrigstid är en väl utforskad period, men genom att applicera digitala metoder på kurerade dataset kommer politikens, nyhetsmediernas och kulturens sfärer att kunna granskas på nytt. I vårt projekt arbetar vi med olika typer av algoritmisk textanalys av storskalig empiri från politikens sfär (allt digitaliserat riksdagstryck och alla statliga offentliga utredningar under perioden), digitaliserad dagspress (i omfattningen tiotalet dagstidningar) och skönlitteratur (alla svenska romaner utgivna under perioden håller på att digitaliseras). Det rör sig om en historisk empiri grupperad i flera dataset, vilka sammantagna omfattar hundratals miljoner ord. Vårt upplägg påminner om det brittiska forskningsprojektet Living with Machines, där ett storskaligt brittiskt kulturarv analyseras med hjälp av snarlika digitala metoder (Hosseini et al. 2021).

Den här artikeln introducerar hur datadriven digital historievetenskap kan analysera kulturarvsfrågor på ett annorlunda sätt. Genom att distans- eller fjärrläsa – *distant reading* är det samlingsbegrepp som ofta används på engelska – 3154 statliga svenska utredningar (SOU) publicerade mellan 1945 och 1989, är det möjligt att spåra diskurser och tematiker kring kulturarv (förstått i vid bemärkelse) på ett slags makronivå. Under denna långa tidsperiod var det hundratals statliga utredningar som ägnade sig åt och tangerade arkiv-, bibliotek- och museifrågor. Sammantagna utgör alla dessa utredningar en så omfattande empiri att en forskare på egen hand aldrig kan bearbeta den. För en människa är det helt enkelt inte möjligt att läsa flera tusen utredningar och dra slutsatser om teman i dem – men med datorers hjälp går det. Mitt dataset av alla utredningar omfattar cirka 87 miljoner ord (tokens). Det har analyserats

genom temamodellering, ett samlingsnamn på en rad algoritmer som kan klassificera stora textkorpusar baserat på tematiska strukturer i textmassan. Genom olika typer av nätverksillustrationer går det därefter att visualisera och spåra förbindelser mellan utredningar och teman på ett slags aggregerad nivå.

I det följande handlar det därför lika mycket om att visa som att beskriva; artikeln innehåller ett drygt tjugotal grafer. Artikeln närläser *inte* vad som skrevs i utredningar om kulturarv, istället betraktar och bearbetar jag dem som en enda text, som ett (av mig) sammansatt dataset. Den empiriska skalan blir då annorlunda, fjärrläsningens kunskapsformer likaså: "Distant reading […] *is a condition of knowledge*" (Moretti 2013:48). Olika former av *text mining* handlar därför strängt taget inte om att läsa – utan snarare om att mjukvara upptäcker mönster i textmassor, samband som kan vara av olika karaktär beroende på vilken programvara som används. Den här artikeln argumenterar för att det spelar roll i vilken medieteknisk form som historien kommer oss till mötes. Min frågeställning kan, i korthet, formuleras ungefär så här: om alla dessa tusentals utredningar betraktas som en enorm text som den svenska staten skriver, vilka tematiker och diskurser kring kulturarv kan programvara läsa och uppfatta? I vilka specifika utredningar är kulturarvsteman som allra starkast? När under svensk efterkrigstid och välfärdsår var ABM-frågor som mest diskuterade? Vilka andra, större samhällsfrågor sammankopplades de med? Och var det A, B eller M som diskuterades mest?

Data och metod – om temamodellering av utredningar som dataset

Den svenska statens intresse för hur kulturarvet ska samlas in, dokumenteras och representeras har en lång historia. Forskningen på området är mycket omfattande, och de flesta studier om svenskt kulturarv under 1900-talet har använt statliga utredningar som behandlat ABM-sektorn (Bohman & Palmqvist 1997; Aronsson & Hillström 2005). Ibland har utredningar nyttjats för studier av kulturarvspolitik (Grinell & Högberg 2020), ibland har syftet varit att historisera synen på kulturarv (Pettersson 2004). I regel har utredningsmaterial i sådana sammanhang nyttjats genom läsning av ett fåtal SOU:er, emellanåt om några fler. I Elin Nystrand von Unges avhandling, *Samla samtid. Insamlingspraktiker och temporalitet på kulturhistoriska museer i Sverige* rör det sig (bland mycket annat) om analys av fem utredningar (Nystrand von Unge 2019). Två museologiska studier kan illustrera denna forskningspraktik; i bägge avhandlingars litteraturlista finns ett tiotal SOU:er refererade (Wittgren 2013; Näsman 2014).

Mitt syfte här är inte att vara polemisk; jag har själv arbetat med ABM-orienterade utredningar på samma generella sätt (Snickars 2020), och ibland i form av fallstudier som av Dataarkiveringskommitténs betänkanden från 1970-talet (Snickars 2015). Men analyserar man alla utredningar under en viss period som ett enda dataset, då kan helt nya frågor ställas. Vad har exempelvis den svenska staten genom sitt utredningsväsende egentligen haft att säga om museer på ett mer övergripande plan? Ja, en hel del förstås – och då framför allt kring 1950 och en bit in på 1970-talet, åtminstone om man följer den statistiska fördelningen av tema 358 (om museer) i SOU-datan. Biblioteksfrågor däremot skjuter fart först fart i början av 1980-talet (tema 452), medan arkivfrågor har toppar under slutet av 1970- och 1980-talet (tema 343). En rad termer (i form av ordnoder) förbinder dessa tre temamodeller: bestånd, information och urval (mellan bibliotek och arkiv), inköp, deposition och katalog

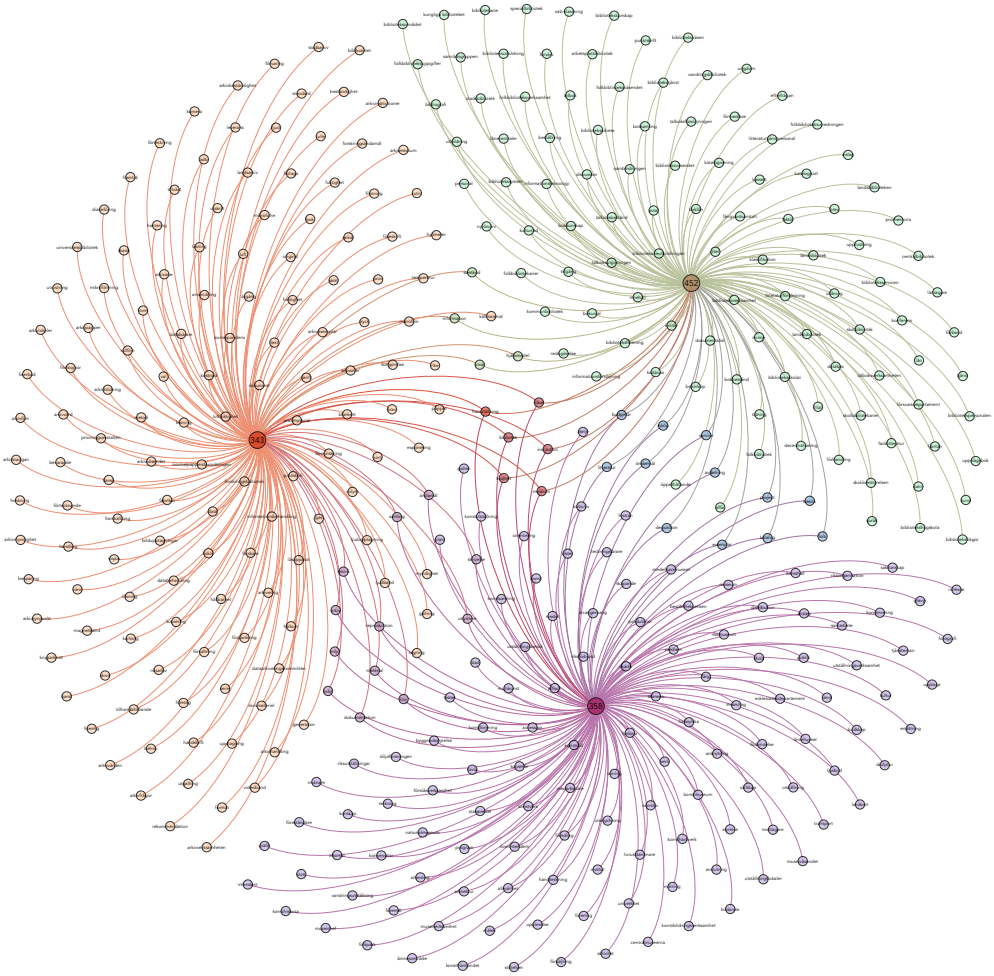


Illustration 1. ABM som nätverk från ett dataset av fler än tre tusen svenska statliga utredningar publicerade mellan 1945–89. Det cirkulärt nätverket i visualiseringsverktyget Gephi (med algoritmen Fruchterman Reingold) kopplar samman ord i tre temamodeller – 343 (om arkiv), 358 (om museer) och 452 (om bibliotek). Nätverket innehåller drygt 440 ordnoder där sex centrala ord (i mörkare rött) förbinder samtliga teman: lokal, framställning, bibliotek, institution, kvalitet och medium.

(mellan bibliotek och museum), samling, dokumentation och reproduktion (mellan arkiv och museum).

Det dataset jag använder i denna artikel består av samtliga offentliga utredningar som svenska staten publicerade mellan 1945 till 1989, publikationer som landets nationalbibliotek, Kungliga biblioteket, har digitaliserat. Innan den svenska regeringen lägger fram ett lagförslag tillsätts ofta en särskild utredare el-

ler en kommitté med uppdrag att utreda en viss fråga; dess resultatet samlas i en rapport (som ibland kallas för ett betänkande) och som publiceras i serien Statens offentliga utredningar (SOU). I Norge har Noregs offentlege utgreiingar (NOU) samma funktion. Det svenska utredningsväsendet har en mycket lång historia, och utredningskommittér var ofta en sorts arena för samverkan – och konfrontation – mellan olika samhällsintressen. Redan tidigt var grundlighet i arbetssätt ett kännetecken för det svenska utredningsväsendet; information inhämtades, bearbetades och nytt vetande producerades (Hesslén 1927; Meijer 1956; Zetterberg 1990; Johansson 1992). Kommittéväsendet fungerade, i korthet, som en kunskapsgenererande verksamhet (Åmark 2005), varför svenska historiker brukar hävda att utredningar spelat en "utomordentligt stor roll för utförandet av 1900-talets politiska reformer" (Östberg & Andersson 2013:118). Men naturligtvis har somliga utredningar haft sina brister; även senare som historiskt källmaterial. De har ibland använts som ett teknokratiskt verktyg för att undvika att ta ställning till samhällspolitiska dilemma. Att tillsätta en utredning har stundtals handlat om att skjuta politiska beslut på framtiden. Om ett dataset betraktas som en större textsamling av strukturerad data, är det därför av vikt att känna till hur det är sammansatt, vad dess delar består av, samt hur det eventuellt reproducerar förutfattade meningar. Historiska dataset har inte sällan sin egen specifika etnografi.

Men det är inte utredningarnas kvalitet, kommittéväsendet som politisk aktör eller utredningars eventuella realpolitiska effekter som intresserar mig i det följande, utan själva textmassan i fler än tre tusen utredningar, ett slags *statens röst* om man så vill. Som ofta inom digital humaniora är iordningställandet av data en mödosam verksamhet (Jarlbrink 2020). SOU-datan är dock av god kvalitet; svart tryck på vita standardiserade boksidor förenklar all OCR-tolkning. Humlab på Umeå universitet har först kört alla tusentals utredningar som textfiler genom svenska Språkbankens korpuspipeline Sparv för att annotera och märka varje ord i varje text med dess ordklass och grundform. Resultatet av både annotering och lemmatisering (sammanföring av olika böjningsformer av ett ord) har lagrats som csv-filer – vilka sedan sammanställts till ett enda stort dataset för analys. Lingvister och språkteknologer har under lång tid arbetat med snarlika, omfattande textkorpusar. Inom vårt projekt Välfärdsstaten analyserad använder vi både enklare språkteknologiska metoder som ordfrekvenser och kollokationer, *bigrams* (ordpar som uppträder tillsammans) liksom mer sofistikerade algoritmer som temamodeller eller NER (*Named-entity recognition*) där personnamn, geografi och tidsmarkörer automatiskt kan taggas upp. Men vi är inte språkvetare; det är inte språkets form som intresserar oss utan texters historiespecifika innehåll.

Den digitala metod jag använder i min analys kallas alltså för temamodellering. Grundfrågan inom all sådana modellering av text handlar om vilka diskurser, motiv eller tematiker som kan ha tänkts skapat de dokument som forskaren intresserar sig för. Jo Guldi har i en analys av politiska diskussioner om brittisk infrastruktur under 1800-talet, baserade på *the Hansard* – transkriberade debatter från Storbritanniens parlament, ett dataset på flera hundra miljoner ord – påpekat att temamodellering gör det möjligt att spåra "the invisible categories that structure mind, language, priorities, or prejudice in a given corpus" (Guldi 2019:5). Inom temamodellering finns flera modeller att arbeta med och de kan ställas in på många sätt. Humlab har valt att arbeta med *latent dirichlet allocation* (lda) och programva-

PREPARE Load Topic Model MANDATORY

```
load_gui = gui.create_load_topic_model_gui(corpus_config, corpus_folder, current_state())
display(load_gui.layout())
```

Model gensim_mallet-lda.topics.50.sou_kb-labb_1945-1989_nn Load

Tokens 25 Find Download

topic_id	tokens	alpha	score
0	verksamhet organisation bidrag förening år museum musik ungdom kr lokal anslag utställning skall medlem utredning stöd del studieförbund antal uppgift form statsbidrag sverige bibliotek idrott	0.048889	0.015203
1	skatt inkomst skall belopp avdrag år kr avgift beskattning fall taxering uppgift deklaration krona förslag arbetsgivare skattemyndighet mom länsstyrelse del person make förordning beredning kostnad	0.053441	0.014149
2	barn förälder familj år behov kommun förskola skola hem ungdom personal verksamhet dag daghem samhälle ålder arbete kontakt grupp skall får utredning möjlighet moder del	0.070789	0.016481
3	läkare patient vård sjukhus sjukvård år hälsa sjukdom antal landsting behandling klinik avdelning personal tandläkare behov lasarett del medicinalstyrelse verksamhet skall fall sjuksköterska socialstyrelse vårdplats	0.052828	0.020577
4	fartyg järnväg år hamn ton trafik kostnad sjöman transport gods kr del sjöfart göteborg befälhavare linje utredning redare stat sjöfartsverk km stockholm resa krona båt	0.044623	0.012134
5	elev skola lärare undervisning årskurs ämne grundskola arbete linje gymnasium antal klass skall läsår moment uppgift åk gymnasieskola fackskola högstadium del rektor läroplan läromedel språk	0.041281	0.016453
6	kostnad företag budgetår år krona anslag stat myndighet byggnad anläggning milj arbete ab upphandling fall del belopp utgift antal pris budget affärsverk förslag medel underhåll	0.055338	0.012044
7	kvinna barn make äktenskap fall år ho abort lag fråga död land förälder sverige skall hustru sou förhållande släktnamn namn hemvist fader får liv tid	0.041405	0.009993
8	film tidning radio program sverige reklam tv år skrift verk bild medium del television press sändning allmänhet kostnad företag information land produktion innehåll kr material	0.044877	0.014537
9	år antal tabell procent grupp andel undersökning kvinna person ålder skillnad sou uppgift resultat fördelning material del samband jämförelse svar fråga summa individ siffra förhållande	0.131367	0.030556
10	arbete år arbetsmarknad person arbetskraft invandrare arbetsförmedling arbetslöshet antal företag åtgärd sysselsättning utbildning kvinna anställning arbetsmarknadsstyrelse grupp sverige länsarbetsnämnd plats verkstad bidrag arbetsgivare procent ungdom	0.053072	0.012655
11	sverige land företag år export u-land länd utland sida svenska marknad bistånd stat del samarbete usa utveckling regering projekt investering fall industri värld kr storbritannien	0.060959	0.014754
12	kommun landsting statsbidrag år bidrag kostnad stad verksamhet beslut landstingskommun kr fullmäktig nämnd primärkommun uppgift antal invånare stat stockholm förvaltning styrelse ledamot organ län kommunallag	0.072523	0.013999
13	kommun fastighet mark område skall länsstyrelse byggnad anläggning fall väg bestämmelse plan kostnad kap ersättning bebyggelse åtgärd ägare företag fastighetsägare stad möjlighet fråga intresse hänsyn	0.053687	0.015926
14	skall bestämmelse lag beslut förslag styck fall regel kap får fråga föreskrift myndighet utredning ärende ändring sou paragraf skäl kommitté förhållande uppgift tillämpning lagstiftning ordning	0.209034	0.051603
15	tjänst befattning befattningshavare lönegrad tjänsteman år tjänstgöring förslag lön anställning personal stat bestämmelse mom lärare antal maj arbetsuppgift avlöning fall skall tid skola ersättning utredning	0.061703	0.011665
16	lägenhet bostad år kommun hus hyresgäst län hushåll fastighet kostnad hyra skall rum kr del småhus antal hyresvärd bostadsrätt kök ombyggnad lokal bostadsstyrelse byggnad storlek	0.059157	0.015513
17	län kommun år del antal region ort stockholm stad göteborg länsstyrelse område befolkning land tätort invånare rike industri örebro folkmängd lokalisering uppsala norrbotten malmö person	0.098611	0.020335
18	läkemedel år apotek livsmedel produkt ämne kontroll skall alkohol försäljning preparat tillstånd vin dryck tillverkning laboratorium föreskrift användning undersökning utskänkning öl får sprit bestämmelse spritdryck	0.038403	0.012058
19	år ton nr produktion kg milj import sverige land tillverkning del pris tull industri förbrukning produkt malm kvantitet handel metall export användning material råvara gruva	0.045114	0.015828
20	uppgift register myndighet information system statistik handling bibliotek bok data del skall antal användning år kostnad adb dator förlag registrering behov material arbete företag databehandling	0.049488	0.013043
21	sou undersökning effekt risk samband människa resultat individ metod fall faktor problem committee york information betydelse pojke studie beteende person grupp material ämne form data	0.071520	0.018890
22	skola elev undervisning lärare ämne utbildning kurs klass år skall gymnasium yrkesutbildning antal folkskola lärjunge linje realskola timme läroverk yrkesskola arbete uppgift förslag rektor överstyrelsen	0.046537	0.021957
23	utlänning polis land polismyndighet uppgift person polisman sverige år flykting fall brott ärende räddningstjänst myndighet polisdistrikt stockholm kommission tid medborgare tillstånd åtgärd rike länsstyrelse regering	0.043470	0.010425
24	utbildning kurs studerande ämne lärare skall förslag gymnasieskola linje år kunskap del sou högskola antal undervisning vuxenutbildning studie studium högskoleutbildning behov elev sö betyg grundutbildning	0.054050	0.017729
25	forskning högskola universitet institution ämne institut undervisning fakultet studerande år utbildning förslag professor examen antal del forskare lärare stockholm anslag behov uppsala lund kommitté kurs	0.051159	0.017677
26	vatten fiske område år del fisk jakt län älv sjö mark djur länsstyrelse land utredning vattendrag anläggning friluftsliv antal intresse kust hav km delägare utbyggnad	0.035708	0.012085
27	egendom skall kap fastighet förslag lag fall fordran bestämmelse styck konkurs borgenär gäldenär regel inteckning beredning tid år utmätning ägare försäljning arrendator avtal betalning förvärv	0.040183	0.012201
28	utredning år fall förslag krona förhållande fråga del kommitté hänsyn stat synpunkt betänkande skola maj sida tid möjlighet synes sätt stad betydelse behov åtgärd skäl	0.303003	0.075773
29	artikel des konvention stat og skall patent article le art uppfinning dans par ou une au ansökning bestämmelse convention est patentverk skal regel skydd medlemsstat	0.034880	0.010140
30	utbildning personal försvar krig utredning år förband chef skall behov tjänstgöring armé uppgift krigsmakt civilförsvar del antal organisation fred officer totalförsvar förslag befattning flygvapen dag	0.035438	0.018305
31	riksdag val parti utskott ledamot år valkrets regering förslag kammare antal mandat sd valsedel väljare fråga skall röst rf fp kandidat folkomröstning grundlag regeringsform medborgare	0.040028	0.010024
32	år kr inkomst skall ersättning pension barn försäkring belopp fall regel kommitté tid bidrag kostnad förslag dag förmån avgift ålder make antal lag sjukpenning kassa	0.059742	0.017884
33	jordbruk år skog län skogsbruk område del utredning kostnad areal hektar jordbrukare kr sverige skogsmark jord avverkning produktion lantbruksnämnd mark land norrland rationalisering milj hushållningssällskap	0.045076	0.020524
34	regel fall ersättning försäkring skall köpare lag säljare skada bolag avtal näringsidkare gods konsument sou bestämmelse försäkringstagare ansvar försäkringsbolag skadestånd styck får tid förhållande utredning	0.041050	0.012947
35	företag pris konsument försäljning år produkt bransch ab antal butik fall marknad konkurrens kostnad handel omsättning del kr detaljhandel avtal undersökning tillverkning inköp uppgift hushåll	0.066384	0.019233
36	fordon trafik väg år bil förare körkort utredning kostnad antal flygplats km olycka motorfordon lastbil skall hastighet hänsyn del personbil buss resa kr fall länsstyrelse	0.050045	0.015884
37	sou tabell bilaga zz tt procent ap 00t te ti ol ed se rad oo le ss ve kr wos oc 0t ip sl lt	0.083399	0.014560
38	brott mål domstol fall kap år skall hovrätt åklagare part straff åtal fängelse rätt regel böter förslag rb sou gärning fråga talan påföljd utredning antal	0.051546	0.015788
39	sou verksamhet utredning möjlighet skall fråga utveckling förslag del område behov åtgärd information form krav arbete problem förutsättning resurs organisation samhälle betydelse planering kommitté uppgift	0.245211	0.078315
40	år utveckling företag investering industri period procent produktion tabell ökning förändring sektor del pris effekt andel konsumtion sou sysselsättning efterfrågan bransch prognos alternativ beräkning tillväxt	0.080215	0.036911
41	arbetstagare arbete arbetsgivare arbetstid timme företag lag arbetare avtal dag vecka kollektivavtal semester tid år arbetsplats fall anställda regel part skall anställning antal fråga organisation	0.066883	0.013235
42	bolag företag aktie förening bank län styrelse skall år fond verksamhet aktiebolag revisor sparbank medlem belopp aktieägare kapital riksbank tillgång kredit medel ab får obligation	0.059916	0.018344
43	utredning verksamhet ärende myndighet uppgift organisation styrelse länsstyrelse organ personal skall fråga nämnd arbetsuppgift ledamot verk förslag avdelning arbete byrå stat ledning tjänsteman beslut förvaltning	0.122343	0.026486
44	sr rr ar antal ee nr summa sn se ne ss rn ae fe ra nn tabell te krona al fr le st oo sl	0.375333	0.039885
45	kyrka församling präst år kyrkomöte gud pastorat skall ord domkapitel sou herre stift text dag biskop gudstjänst ordning samfund tid liv psalm trossamfund jesu del	0.038094	0.017862
46	renskötsel sameby ren rennäringen von namn lapp nomadskola betesområde släktnamn lappby lappbyarna valdeltagande rennäringens rennäringslagen sametinget lappbyar kultur renantalet renskötselområdet renskötselrätten renskötare urbefolkning ssr fuvl	0.002234	0.018303
47	år fastighet beskattning värde inkomst avdrag skall regel kr fall taxering del belopp vinst sou rörelse kostnad byggnad företag tillgång skatt fastighetstaxering intäkt får kapital	0.045686	0.013588
48	anstalt vård år fall person åtgärd behandling skall nämnd tid barnavårdsnämnd brott omhändertagande antal undersökning intagning behov klientel övervakning klient uppgift arbete kriminalvård socialvård möjlighet	0.036573	0.012755
49	år olja bränsle anläggning utsläpp energi industri sverige kostnad avfall sou del ton åtgärd effekt kol tabell ämne bilaga användning system twh utveckling skall teknik	0.039211	0.022784

Illustration 2. Skärmdump av analysmiljön för temamodellering i Jupyter Lab – utvecklad vid Humlab på Umeå universitet – som visar de 25 vanligast förekommande orden i 50 teman (topic-id 0–49) i ett dataset av fler än tre tusen svenska statliga utredningar (SOU) publicerade mellan 1945–89. 50-modellen delar in SOU-datan i femtio teman – det vill säga, algoritmen instrueras att ta fram femtio stycken teman baserade på ords statistiska sammanhang. I SOU-datan återfinns teman i procentuell omfattning (alpha-värdet till höger); tema 2 (om barn och familj) förekommer exempelvis i cirka sju procent av alla utredningar. Vad teman handlar om är upp till forskaren att bestämma, numrering är också godtycklig. Men som synes är det ofta tämligen uppenbart: tema 0 (om kultur, museum och bibliotek), tema 5 (om utbildning), tema 8 (om medier), tema 25 (om forskning och universitet) eller tema 41 (om arbete). Temamodeller fångar också upp brus och förkortningar i datan; uppenbart ovan i tema 37 och 44.

ran Mallet där modeller ställts in för att sortera fram 50, 100, 200 och 500 teman (härefter kallat 100-modellen etcetera). Teman har inga inherenta beteckningar utan listas enbart i stigande nummerordning (från 0 till 49 eller 0 till 499). All namngivning av teman baserar sig därför på subjektiva tolkningar, varför jag endast använder siffror i mina illustrationer.

Samtliga temamodeller har gjorts åtkomliga för mig i utvecklingsmiljön Jupyter Lab via så kallade *notebooks*. Genom ett användarvänligt gränssnitt kan man i dessa notebooks köra (och ändra på) Pythonkod direkt i webbläsaren. I Jupyter Lab finns kodceller (i varje notebook) där man som forskare med knappar och manuella reglage kan styra parametersättningen av koden; de kan modifieras och exempelvis plotta valfritt antal ord i varje tema (från 25 till 200), lista teman över tid (eller en specifik tidsperiod), jämföra eller vikta teman med andra snarlika tematiker etcetera. Resultatet förevisas som tabeller, diagram eller enklare nätverk. Arbetsgången handlar i regel om att söka fram teman, se efter vilka utredningar som ingår, ladda ned tabelldata och därefter visualisera resultaten i Gephi – standardprogramvara för nätverksvisualiseringar – för bättre överblick. I Gephi är det också möjligt att interaktivt se exakt vilka noder som hänger samma.

En erfarenhet av sådant forskningsarbete, där empiri och metod sammanförs i ett slags infrastrukturellt gränssnitt, är att det ibland är svårt att hålla isär vad som är vad. Körningar med utredningar omvandlade till dataset framträder först när metoden som används genererar ett resultat. En annan är att man som forskare sitter fast i det svenska språket; temamodeller i denna artikel innehåller uppemot hundratusen svenska ord. Att översätta dem hade varit ogörligt (och det främsta skälet till att artikeln inte är skriven på engelska). Noterbart är också att trots att data kurerats och iordningställts så innehåller de flesta textuella dataset likväl ett kvardröjande brus; tabelldata behöver därför ofta rensas på ovidkommande ord såsom utredningsrelaterade termer (bilaga, figur, kapitel), förkortningar (cm, km, kg), samt tids- och datumangivelser. Stoppord filtreras också bort. Den historievetenskaplig empiri som ligger till grund för datadriven digital historieforskning är med andra ord långt ifrån identisk med den som återfinns i arkiven (Snickars 2022).

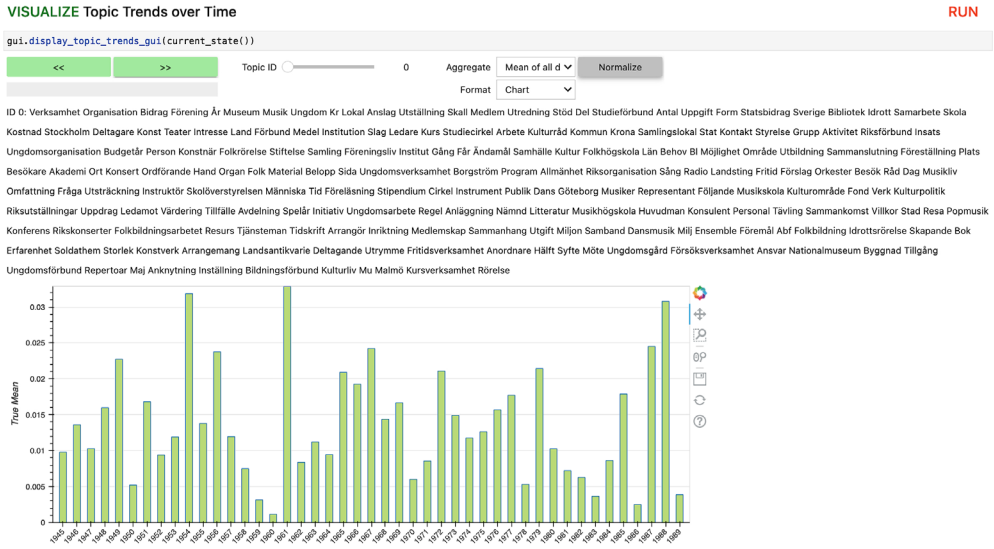
Temamodeller handlar om probabilitet. I modeller som listar 200 eller 500 teman är sannolikhetsvärden mycket låga (i promille, eftersom det handlar om många teman), medan de i modeller om femtio teman är något högre (ofta i procent). Temamodeller arbetar med statistisk analys av distribution av ord i form av sannolikhetsfördelning över alla ord i en textmassa. Modellerna kan urskilja vilka ord som ingår i ett tema (baserat på ords statistiska fördelning), vilka statliga utredningar som innehåller detta tema, hur det fördelar sig över tid liksom hur temat – eller ord i ett tema – anknyter till andra teman och utredningar. Grundprincipen är att ord samförekommer. Det hela kan förefalla komplicerat men är inte speciellt konstigt; det är mer sannolikt att ord som dokumentalist och informationsförsörjning förekommer i en SOU om bibliotek än termer som fartyg eller jordbruk. Som André Baltz och Fredrik Norén emellertid påpekat i en pedagogisk introduktion av temamodellering så är det "ords fördelning, inte deras ordning i dokumenten (eller det språk orden tillhör) som har en betydelse för hur teman genereras" (Baltz & Norén 2021:222). Temamodeller bygger nämligen på antagandet om så kallade *bag-of-words*; modellerna betraktar inte ett dokument som en sekvens av ord, utan som en påse med ord – där alla ords förekomst adderas, exempelvis som en frekvens.

Eftersom temamodellering handlar om statistisk sannolikhet går det att i en Jupyter Lab-miljö laborera med parametrar och de resultat som modellerna genererar, exempelvis beträffande relationer mellan teman. Sätter man till exempel ett tröskelvärde på 0,3 då listas alla utredningar som innehåller minst trettio procent av ett specifikt tema. Om tröskelvärdet sänks, så kommer följaktligen fler teman att hänga samman. Omvänt minskar kopplingen mellan teman om samma värden höjs. Men om tröskelvärden sätts alltför lågt så finns risken att teman egentligen handlar om något annat. Exakt vilket värde som bör användas beror på hur många teman en modell har. I en 500-modell är det konkurrens, där behövs låga tröskelvärden för att få korn på relationer. I en 50-modell kan de vara betydligt högre.

Om temamodellering ger indikationer på hur övergripande tematiker fördelar sig över specifika utredningar, hur de är relaterade och samförekommer, så är det av vikt att notera vilken typ av modell man som forskare använder. Modeller som delar in textmassan av utredningar i 50 teman ger mer generella resultat, medan modeller med 500 teman är mer specifika. Som digital metod kan temamodellering framstå som abstrakt, men i en Jupyter Lab-miljö är det inte speciellt komplicerat att utföra. Som forskare söker man fram ett tema med i mitt fall koppling till kulturarvsfrågor; man väljer tidpunkt eller tidsspann, man kan studera relationer mellan utredningar eller mellan termer – eller båda. Det är helt enkelt möjligt att vrida och vända på det dataset man arbetar med; att laborativt experimentera och dela in och modellera data på olika sätt, testa hypoteser – bekräfta eller avfärda dem. I allra högsta grad rör det sig om en explorativ forskningsverksamhet. Och eftersom alla enskilda utredningar finns tillgängliga på Kungliga bibliotekets webb är det dessutom enkelt att gå från mer övergripande frågeställningar till att ladda hem och kontrollera de specifika utredningar som behandlas.

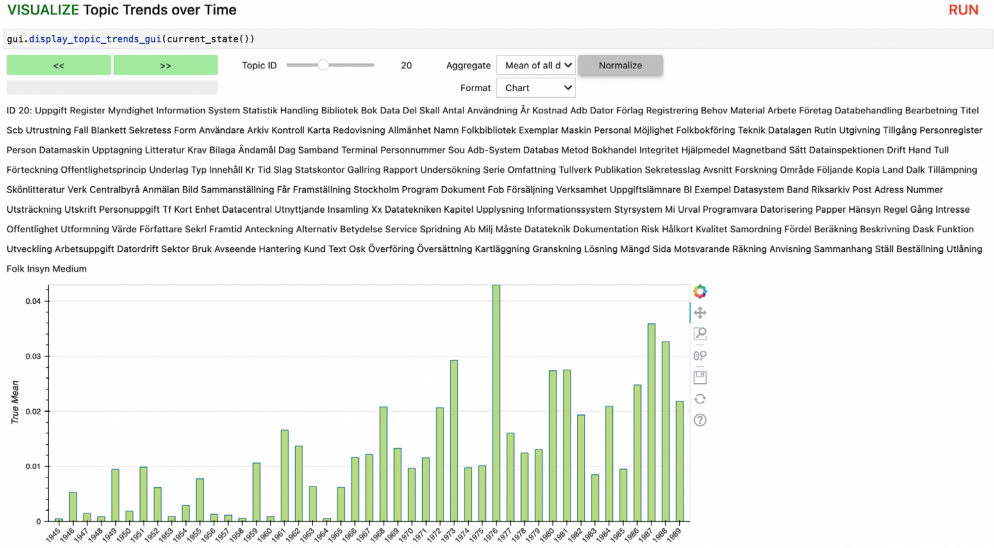
Modeller av ABM

Ett vanligt sätt att som forskare påbörja arbetet med temamodellering är att initialt använda en modell som delar in ett dataset i relativt få teman, i mitt fall 50 stycken. Som framgår av illustration 2 ovan så fördelas SOU-datan då i mer övergripande samhällsfrågor – om transport och infrastruktur (tema 4), om arbetsmarknad (tema 10), om bostadsfrågor (tema 16) eller om myndighetsinformation och data (tema 20). Men där finns också ett tydligt tema som innehåller ord som museum, bibliotek och kultur (tema 0). Sannolikheten för att detta tema ska förekomma i utredningarna är knappt fem procent, det kan låta lite – men teman om skatter eller sjukvård är ungefär lika vanligt förekommande. Mer intressant är hur detta kulturtema fördelar sig över tid. Några tydliga uppgångar i temat är dock svåra att upptäcka, däremot förefaller kulturfrågor i utredningarna varit mindre diskuterade under sent femtiotal och tidigt åttiotal. Om man som forskare är intresserad av mer generella tematiker i SOU-datan och breda samhällsfrågor, så är 50-modellen alltså användbar. Genom att periodisera utredningsdata kan man exempelvis snabbt få överblick av empirin. Temamodellen om myndighetsinformation, data och arkiv (nummer 20) ger exempelvis besked om svenska staten fram till 1965 ägnade föga uppmärksamhet åt datafrågor – men att de därefter tilltar markant. Det är knappast förvånande, men om inte annat ett tecken på att 50-modellens algoritm fungerar och genererar ett plausibelt resultat.



Illustrationer 3–5. Stapeldiagram för tema 0 (om museum, bibliotek och kultur) och för tema 20 (om myndighetsinformation, data och arkiv) i 50-modellen av SOU-datan med de 200 vanligast förekommande orden, det vill säga när lda-algoritmerna ställs in för att sortera datan i femtio teman så återfinns ett tydligt kulturtema och ett informationstema. Nätverket ovan plottar cirka 300 ord i båda dessa teman – samt tjugotalet ordnoder som förbinder dem (i lila färg), bland dem bok, bibliotek och program.

Illustration 4.



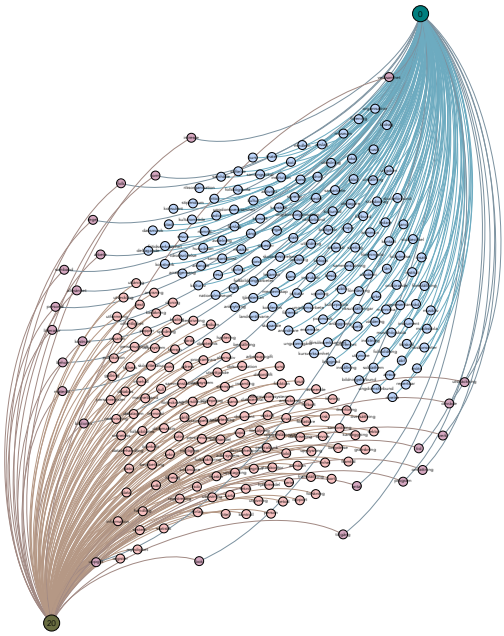


Illustration 5.

För att upptäcka frågeställningar som är mer explicit relaterade till arkiv-, biblioteks- och museisektorn i SOU-datan behöver man växla till en 100-modell. Då framträder ett tydligt ABM-tema (nummer 39) som behandlar bibliotek, museer och kulturminnesvård (tydligt i illustration 6 och 7), och där den senare kategorin också är länkad till naturvård och byggnadsfrågor (tema 8). Arkiv ryms delvis inom detta ABM-tema, men arkivrelaterade frågor förekommer också i en annan temamodell om data, information och sekretess (nummer 69) liksom i ett specifikt medietema (nummer 62) kring audiovisuella bevarandefrågor. Separerar man de olika bokstäverna i ABM och schematiskt undersöker dem var för sig, kan man notera att svenskt utredningsväsende mest ägnat tid och kraft åt biblioteksfrågor. De har i regel utretts i relation till museirelaterade spörsmål, medan arkivfrågor ofta behandlats separat i andra utredningar.

I Jupyter Lab-miljön är det enkelt att söka efter snarlika termer (som ingår i ABM-tema 39) och på så vis kan man för 100-modellen av SOU-datan få fram en rad teman med förbindelser sinsemellan – exempelvis med tema 41 (om forskning och universitet), med tema 62 (om medier), med tema 69 (om data, information och sekretess) och med tema 93 (om förlagsbranschen). Om man i i Jupyter Lab-miljön skapar ett nätverk mellan dessa fem temamodeller – och undersöker det vid fyra olika tidpunkter (se illustration 13) – framträder en tämligen komplex grafisk illustration av SOU-datan. Inte bara ändrar förbindelser mellan dessa teman karaktär över tid, de uppstår också en sorts utredningsblomster (i varierande omfång) under varje period, ofta med ett tilltagande antal utredningar (vilket delvis är en följd av det publicerades något fler utredningar per år mot slutet av den period som undersöks). Mellan 1945–55 är ABM-tema 39 tämligen isolerat, alltmedan forskningsnoden 41 är vitt förgrenad. Detsamma gäller för nästa tioårsperiod, men då har ABM-tema 39 en tydligare koppling till tema 69 om data och information, liksom till medietema 62. Under nästa period, 1965–75, så återfinns en rik flora av utredningar kring alla teman, men noterbart är att det har få förbindelser sinsemellan. Under denna period utreds ABM-frågor separerat från exempelvis medie-, forsknings- eller datafrågor. Endast en utredning, *Arkiv inom hälso- och sjukvård* (SOU 1968:53) förbinder tema 39 med tema 69. Under den sista perioden, mellan 1975–85, är bilden snarlik även om förbindelserna mellan teman då är något tydligare än tidigare.

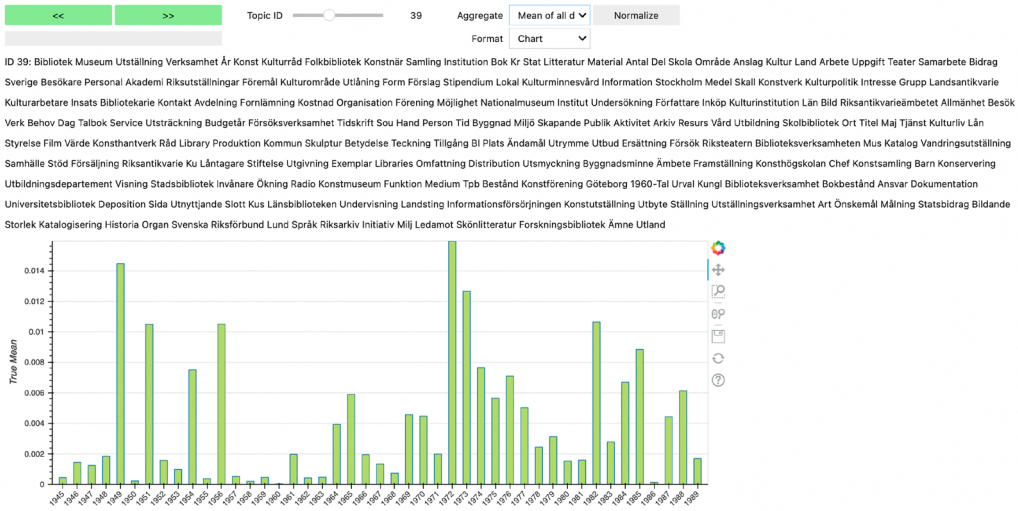


Illustration 6.

Illustration 7.

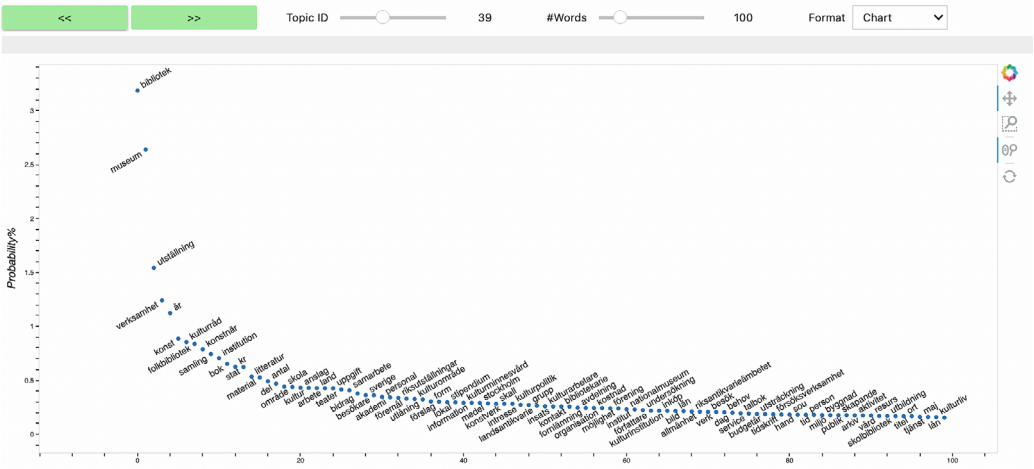
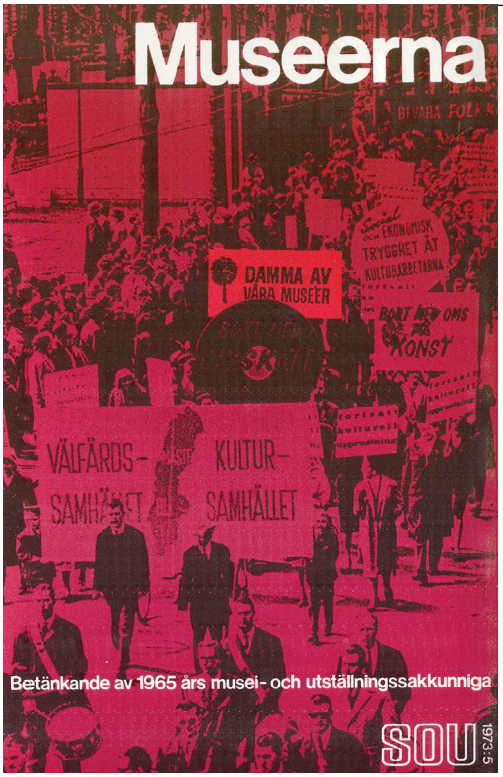
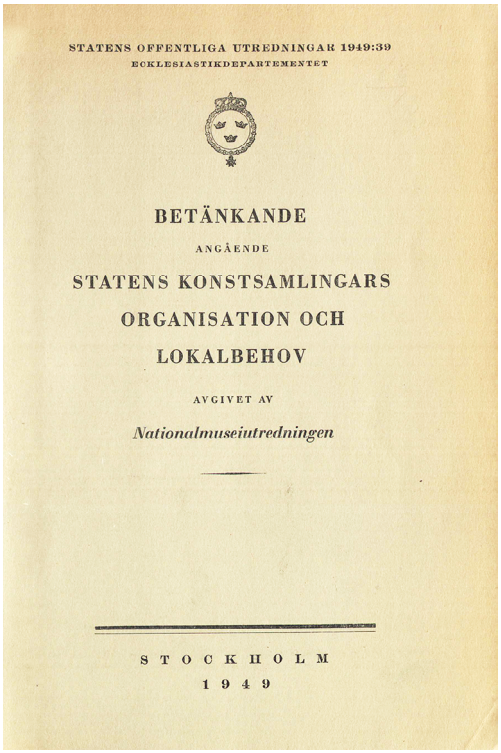


Illustration 6–12. Stapeldiagram för ABM-tema 39 (om bibliotek, museer och kulturminnesvård) i 100-modellen av SOU-datan med de 200 vanligast förekommande orden – med en tydlig uppgång under tidigt sjuttiotal. Grafen ovan plottar hundra av de vanligaste orden i samma tema – i fallande frekvens. Tema 39 om ABM förekommer allra främst i fem utredningar (på kommande sidor); att två av dem utgör betänkanden från 1965 års musei- och utställningssakkunniga (vilka publicerades under 1970-talet) är inte iögonfallande. Vad som är förvånande att temat förekommer allra mest i Talböcker. Utgivning och spridning (SOU 1982:7), liksom att Nationalmuseiutredningen 1949 och Sigurd Curmans betänkande kring organisationen av statens arkeologiska samlingar från 1951 också tillhör topp fem.

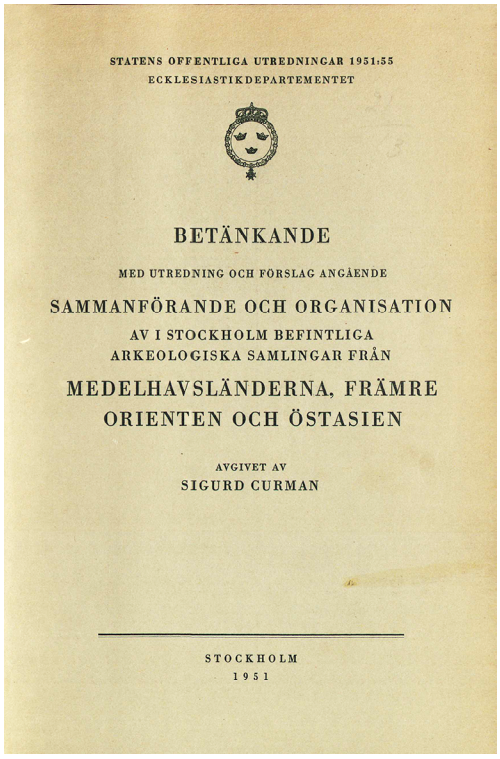




STATENS OFFENTLIGA UTREDNINGAR 1949:39
ECKLESIASTIKDEPARTEMENTET

BETÄNKANDE
ANGÅENDE
STATENS KONSTSAMLINGARS
ORGANISATION OCH
LOKALBEHOV
AVGIVET AV
Nationalmuseiutredningen

STOCKHOLM
1949



STATENS OFFENTLIGA UTREDNINGAR 1951:55
ECKLESIASTIKDEPARTEMENTET

BETÄNKANDE
MED UTREDNING OCH FÖRSLAG ANGÅENDE
SAMMANFÖRANDE OCH ORGANISATION
AV I STOCKHOLM BEFINTLIGA
ARKEOLOGISKA SAMLINGAR FRÅN
MEDELHAVSLÄNDERNA, FRÄMRE
ORIENTEN OCH ÖSTASIEN
AVGIVET AV
SIGURD CURMAN

STOCKHOLM
1951

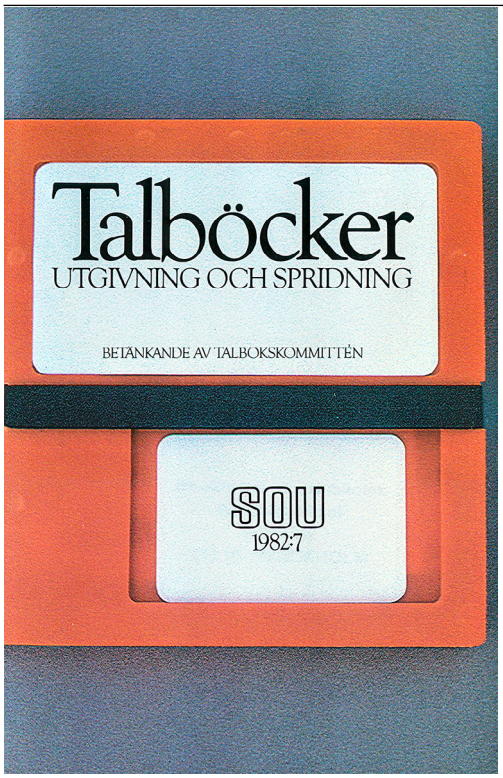
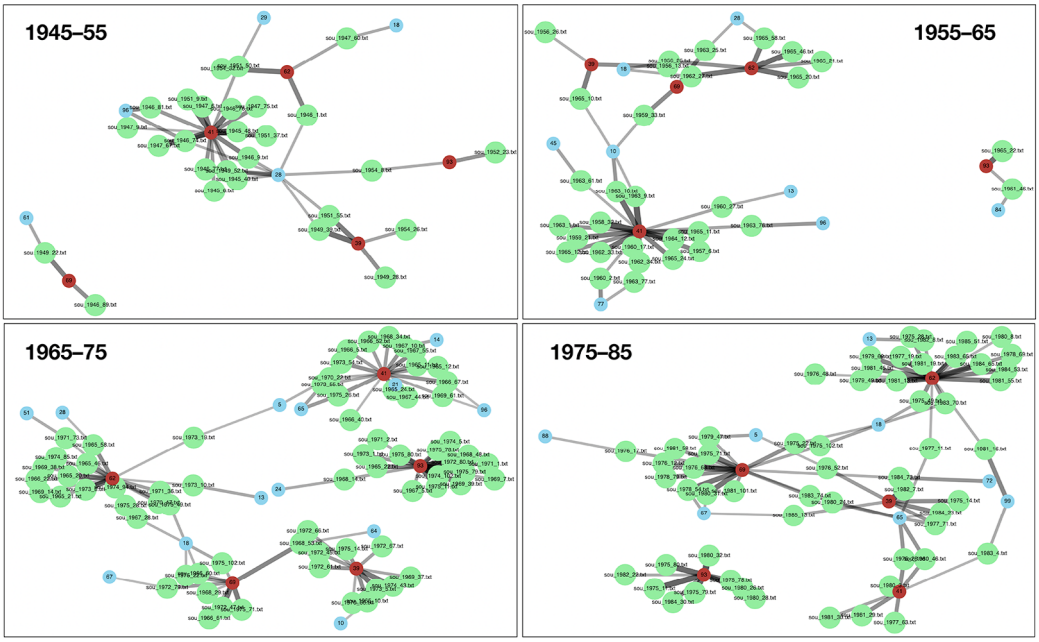


Illustration 8-12.

På ett snarlikt sätt kan man fortsätta att explorativt redogöra för hur arkiv-, biblioteks- och museifrågor förändras över tid och relaterar till andra temamodeller i utredningarnas textväv. På grund av utrymmesskäl kommer jag dock framgent att koncentrera mig på 500-modellen av SOU-datan. För det är i sådana mer differentierade modeller som det är enklare att få syn på mer specifika frågeställningar kring hur arkiv-, bibliotek- och museifrågor utreddes och diskuterades. Eftersom

Illustration 13. Nätverk mellan utredningar och fem (rödfärgade) temamodeller under fyra olika tidsperioder i 100-modellen av SOU-datan med ett tröskelvärde på 0,2 (där tjocka linjer anger ökade samband) – tema ABM (nummer 39), tema 41 (om forskning och universitet), tema 62 (om medier), tema 69 (om data, information och sekretess) och tema 93 (om förlagsbranschen). Gröna noder utgörs av enskilda utredningar och blå noder är diverse teman med viss förbindelse till de röda huvudnoderna.



det i 50-modellen fanns ett tydligt kulturtema (med ord som museum och bibliotek) är det inte förvånande att sådana termer återkommer när lda-algoritmerna ska dela upp SOU-datan i 500 teman. Modellen urskiljer då fyra tydliga temamodeller med koppling till ABM-sektorn: tema 343 (om arkiv), tema 358 (om museer), tema 386 (om kulturminnesvård) och tema 452 (om bibliotek). Vid sidan av dessa fyra huvudteman kring kulturarv finns det också teman där samma eller snarlika ord figurerar – som ordet fornlämning i bebyggelsetema 463, liksom böcker och bibliotek i både universitetstema 68 och kulturtema 125. I den senare temamodellen ingår också museirelaterade termer som utställning och kulturförmedling. Men framför allt är det termen arkiv som figurerar i många teman i 500-modellen: i journaltema (nummer 50) om sjukhusjournaler och arkivhandlingar i vårdsektorn, liksom i registertema (nummer 182) där de sex mest frekventa orden är uppgift, register, myndighet, datalagen, personregister och information – ett tema i utredningarna som för övrigt knappt existerar innan 1970 men som sedan växer kraftigt i styrka.

ABM-förbindelser i SOU-datan går alltså att spåra i flera riktningar, men här kommer jag att fokusera på de fyra temamodeller där kulturarvsdiskursen är som tydligast. Som framgår av stapeldiagrammen nedan är arkivtema 343 som starkast under 1970-talet; det är i utredningen *Moderna arkivmedier* (SOU 1976:68) som det är allra mest förekommande. Museitema 358 är i sin tur som starkast kring 1950 och (liksom arkivtemat) vid mitten av 1970-talet. Temat kring kulturminnesvård 386 är mer utspritt över tid än de tre andra, med toppar vid mitten av 1950-, 1960- och 1970-talet. Temat är (föga förvånande) som allra starkast i utredningen *Kulturminnesvård* (SOU 1972:45), ett resultat av kommitén 1965 års musei- och utställningssakkunniga. Lite mer iögonfallande är att bibliotekstema 452 är mest frekvent under tidigt 1980-tal. Här kunde man förledas att tro att det är i folkbiblioteksutredningens betänkande *Folkbibliotek i Sverige* (SOU 1984:23) som temat är som starkast – men det stämmer inte, delvis eftersom ord som informationsteknologi också inryms. Istället toppas temat av utredningen *Telubaffären* (SOU 1981:48) som utredde eventuella missförhållanden på det statliga företaget Telub. Bolaget hade 1979 skrivit kontrakt med Libyen för att utbilda teletekniker till nya informationssystem – som det visade sig också inbegrep militära användningsområden. Skandalen var ett faktum. I tema 452 samsas därför termer som stadsbibliotek, biblioteksförening och bokinköp med (lätt förvånande) ord som Telub och libyer.

Telubaffärens roll inom den svenska ABM-sektorn tillhör kulturhistoriens marginal, men eftersom temamodellering baserar sig på ords statistiska fördelning är den likväl synbar. Granskar man de fyra kulturarvsteman över tid är det annars påfallande hur få utredningar som sammanbinder dem. Visserligen är sannolikhetsvärden mycket låga i 500-modellen – eftersom det handlar om väldigt många teman – ändå är dessa fyra teman under perioden sammanlänkade med endast ett fåtal utredningar och knappt alls sinsemellan. Plottar man förbindelser under olika tidsperioder (se illustration 18) med ett tröskelvärde på 0,05 – det vill säga, ett tema ska förekomma till åtminstone fem procent – så är det egentligen bara under perioden kring 1970 som dessa fyra teman relaterar till varandra via tiotalet utredningar. Bibliotekstema 452 förekommer dessutom inte alls under perioden 1955–65, och detsamma gäller för museitema 358 under perioden 1975–85. Analyserar man samförekomsten av alla teman under hela perioden 1945–89 är förbindelserna naturligtvis flera, och noterbart är där att museitemat 358 och

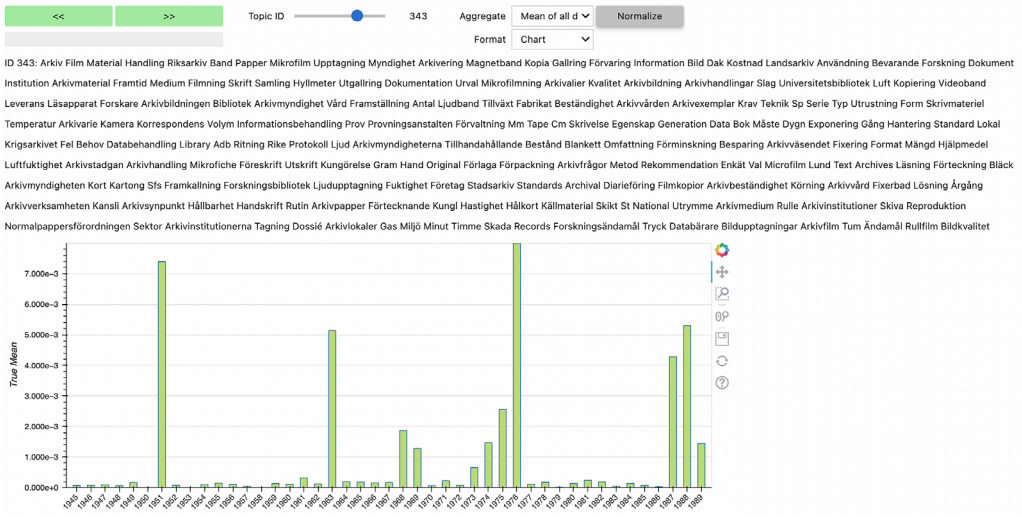


Illustration 14.

Illustration 15.

Illustration 14–17. Stapeldiagram för tema 343 (om arkiv), tema 358 (om museer), tema 386 (om kulturminnesvård) och tema 452 (om bibliotek) i 500-modellen av SOU-datan med de 200 vanligast förekommande orden i varje tema.

88

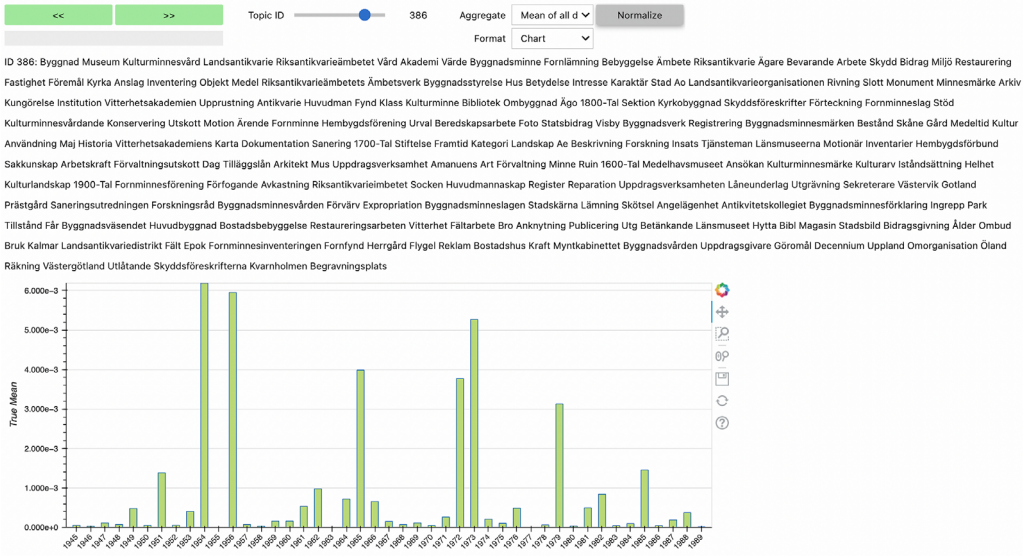


Illustration 16.

Illustration 17.

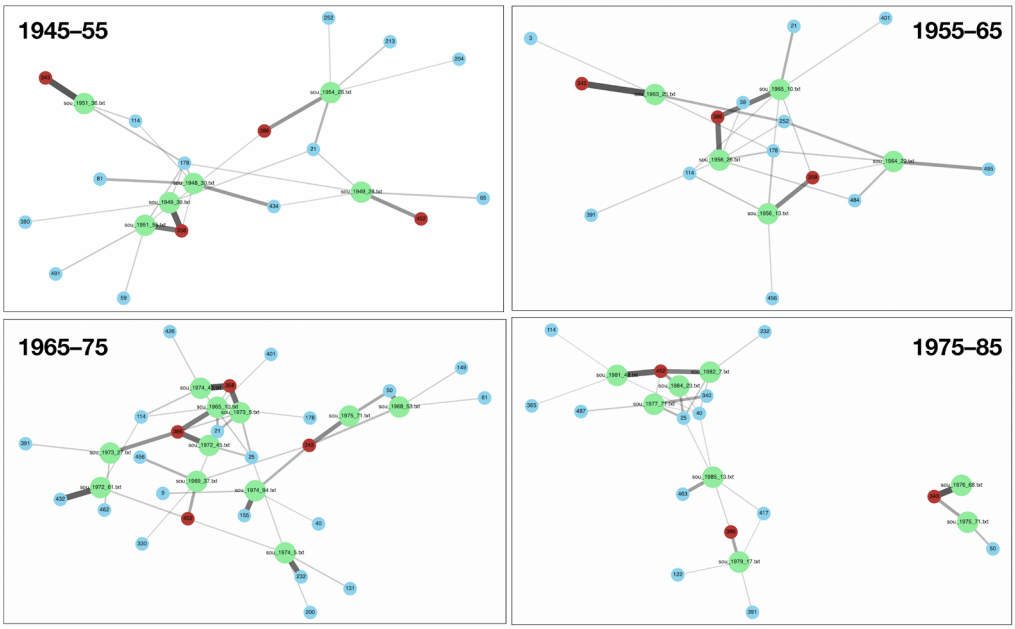


Illustration 18. Nätverk mellan utredningar och fyra (rödfärgade) temamodeller under fyra olika tidsperioder i 500-modellen av SOU-datan med ett tröskelvärde på 0,05 (där tjocka linjer anger ökade samband) – tema 343 (om arkiv), tema 358 (om museer), tema 386 (om kulturminnesvård) och tema 452 (om bibliotek). Gröna noder utgörs av enskilda utredningar och blå noder är teman med viss förbindelse till de röda huvudnoderna.

kulturminnesvårdstema 386 är allra tydligast sammanflätade.

Ett annat sätt att studera relationer i SOU-datan mellan dessa fyra kulturarvsteman är att plotta de trettiotalet ord som är vanligast förekommande i varje tema under hela perioden 1945–89 och jämföra dem med varandra (se illustration 19). Ord som föremål och museum är då gemensamma för museitemat 358 och kulturminnesvårdstema 386, medan ord som medium och information länkar samman arkivtemat 343 med bibliotekstemat 452. Vad som emellertid är slående är hur få termer som de olika temamodellerna delar och har gemensamt. Faktum är inte ett enda begrepp sammanbinder alla fyra teman, och endast två ord (institution och bibliotek) är gemensamma i tre modeller. Och ökar man jämförelsen till 200 ord inom varje temamodell – ja, då är det likväl enbart institution och bibliotek som samförekommer i alla fyra. Att den vokabulär som används inom ABM-sektorn inte är densamma när arkiv-, bibliotek- och museifrågor utreds är något förvånande. Visserligen är det flera ordnoder som förbinder två modeller med varandra, men analysen av SOU-datan antyder att kulturarvsfrågor – liksom medierelaterade frågor (Norén & Snickars 2017) – har en tydlig tendens att utredas separat, med ett språkbruk som inte alltid är gemensamt och som dessutom inte heller nyttjar identiska begrepp.

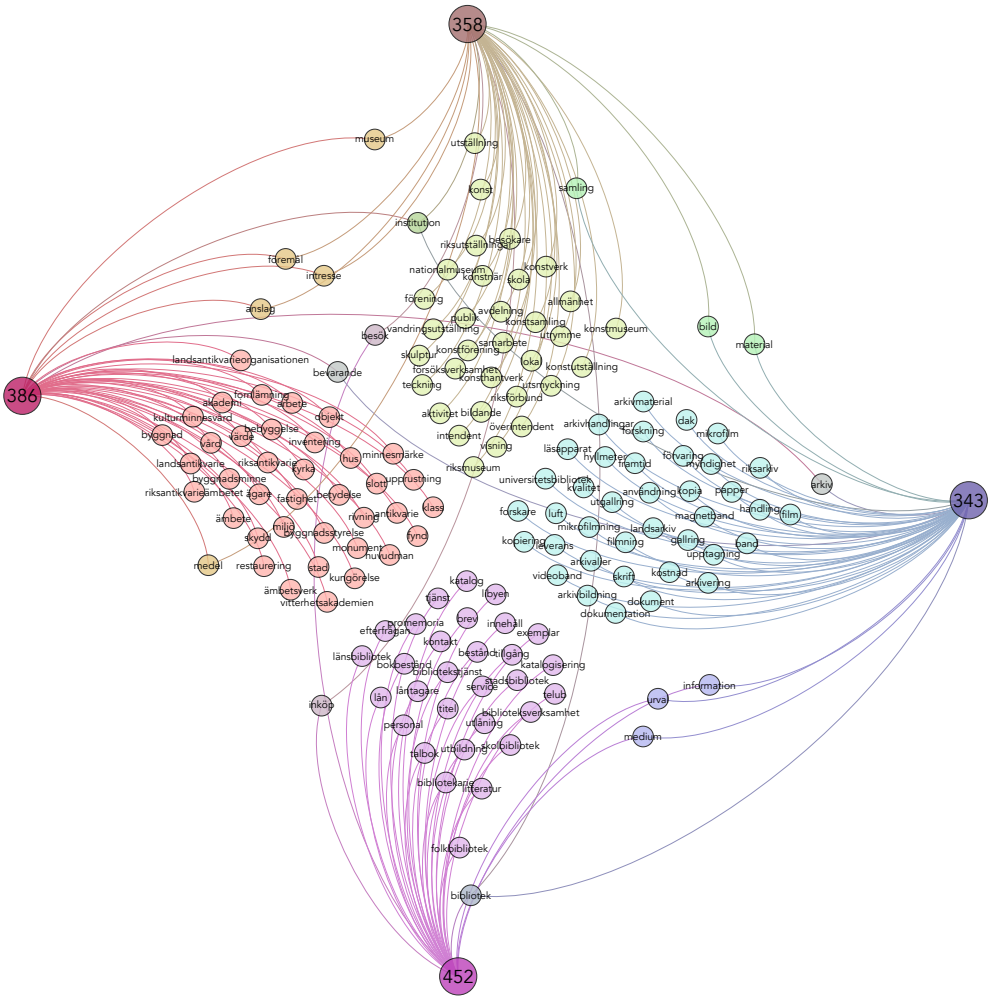


Illustration 19. Nätverk i Gephi (med algoritmen Force Atlas) av 120-talet ordnoder som är vanligast förekommande i respektive kulturarvsteman – arkivtema 343, museitemat 358, kulturminnesvårdstema 386 och bibliotekstema 452 – under perioden 1945–89 i SOU-datan. Noterbart är att terminologi skiljer sig markant; inte ett enda begrepp förbinder samtliga fyra teman.

I Jupyter Lab-miljön och i Gephi kan man fortsätta att på snarlikt manér vrida och vända på SOU-datan. Ytterligare ett sätt att studera relationer inom ABM-sektorn är att fokusera på de utredningar där kulturarvsteman (se illustration 20) är mest frekvent förekommande, liksom vilka andra teman som dessa utredningar då knyter an till. Om man sätter tröskelvärdet mycket lågt till att utredningar ska innehålla ett tema till åtminstone fem procent

– som sagt, i 500-modellen måste värden vara låga – så går det att plotta ett nätverk i Gephi med cirka 130 utrednings- och temanoder, med koppling till de fyra kulturarvsteman. På ett mer generellt plan visar nätverket – där grafisk placering alltså är betydelsebärande – att museitema 358 och kulturminnesvårdstema 386 ligger relativt nära varandra, medan arkiv- och biblioteksmodellerna är mer separat placerade. Studerar man utredningars placering, kan man notera att tätt intill arkivtema 343 återfinns ånyo utredningen *Moderna arkivmedier* (SOU 1976:68), men även utredningen *Landstingens arkiv* (SOU 1975:71) liksom *Öppenhet och minne* (SOU 1988:11), ett betänkande av den arkivutredning som dåvarande kulturminister Bengt Göransson tillsatte i mitten av 1980-talet. I nätverket går det också att bli varse hur äldre utredningar sammanlänkar kulturarvsteman: så förbinder exempelvis utredningen *Arkiv- och biblioteksfilmning* (SOU 1951:36) – om behovet av att mikrofilma arkiv- och tidningsmaterial – arkiv- och museitemat. På samma sätt är utredningen *Utbildning för bibliotek, arkiv och informatik* (SOU 1969:37) en länk mellan arkiv- och bibliotekstemat. Nätverket ger också en tydlig illustration av de temamodeller som är länkade till kulturarvsfrågor och närliggande utredningar. Invid bibliotekstema 452 återfinns till exempel temamodell 232 som handlar om bokförlag och bokhandel, där den främsta förbindelselänken utgörs av 1970-talets litteraturutredning och dess huvudbetänkande, kort och gott, kallad *Boken* (SOU 1974:5). Och strax intill den i nätverket återkommer utredningen *Telubaffären* från 1981 – vilket gör att temamodell 365 om vapenproduktion – med termer som krigsmateriel, export, försvar, materiel, försvarsindustrin försvarsmateriel, bofors, nedrustning och vapen – något oväntat här sammanlänkas med kulturarvsfrågor.

Vid det här laget torde det stå klart för läsaren att teman är som allra starkast i SOU-datan vid tidpunkter när statliga utredningar publicerats vilka explicit behandlat kulturarvsfrågor. Det är förstås inte förvånande, men bekräftar på nytt att algoritmerna utför sitt numeriska arbete väl. I nätverket ovan befinner sig till exempel Nationalmuseiutredningens betänkande från 1949 liksom 1970-talets utredningar om *Museerna* (SOU 1973:5) och *Utställningar* (SOU 1974:43) – bägge betänkanden från 1965 års musei- och utställningssakkunniga – strax intill museinoden 358. Men som exemplet med Telub antyder så återfinns i detta dataset också oväntade förbindelser, vilka är vanligare än man kan tro. Det är en av temamodelleringens fördelar att algoritmerna kan upptäcka och avslöja statistiska relationer som går den mänskliga blicken förbi. Så är exempelvis kulturminnesvårdstemat genom utredningen *Förbud mot utförsel av kulturföremål* (SOU 1964:22) sammankopplat med en temamodell om äganderätt, egendom och fideikommiss (tema 495). På samma sätt är det något överraskande att det är utredningen *De botaniska trädgårdarnas verksamhet och finansiering* (SOU 1988:68) som förbinder museitema 358 med forsknings- och universitetstemat 330. Det är i regel först genom visualiseringar i Gephi som det går att skönja sådana oväntade förbindelser; de kan därefter i nästa steg leda forskaren att explorativt undersöka andra datarelationer. I nätverket ovan kan därför i princip varje nod tas som utångspunkt för vidare efterforskning beroende på intresse. Mellan museitemat 358 och forsknings- och universitetstemat 330 finns det just ett antal förbindelser, som även inbegriper en temamodell som handlar om humaniora (tema 396) – med frekventa termer som språk, historia, universitet, latin och litteratur. Det är en temamodell som mest förekommer i SOU-datan fram till

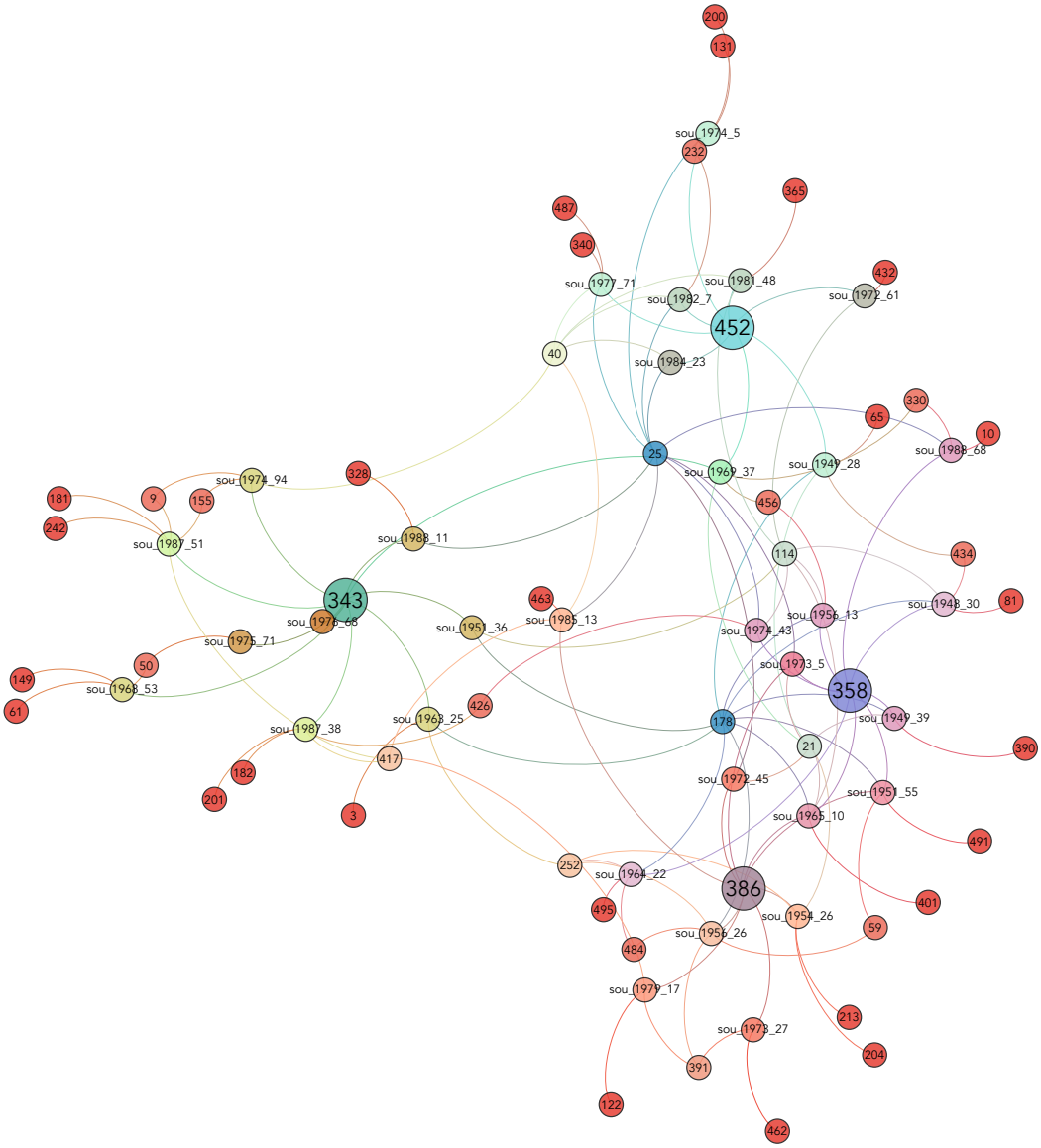


Illustration 20. Nätverk i Gephi (med algoritmen Force Atlas 2) av 130-talet utrednings- och temanoder i SOU-datan med ett tröskelvärde på 0,05 för arkivtema 343, museitemat 358, kulturminnesvårdstema 386 och bibliotekstema 452. Rödaktiga noder utgörs av angränsande temamodeller – som ibland handlar om något annat än kulturarv, som exempelvis tema 365 (om vapenproduktion) eller tema 495 (om egendom och fideikommiss).

1970, en indikation om att humaniora därefter sätts på undantag inom utredningsväsendet.

Avslutning

Att ägna sig åt datadriven kulturhistorisk forskning är annorlunda och skiljer sig från traditionell historievetenskaplig praktik. Men även data måste tolkas; hermeneutiska övningar förblir en gemensam nämnare mellan kvalitativa och digitala metoder. Ofta brukar det hävdas att dessa metoder bör kombineras; forskning når bäst resultat om fjärrläsning kompletteras med närläsning heter det då. Men som den här artikeln visat går det bra att endast använda sig av digitala metoder för att analysera ett omfattande textuellt dataset av 44 årgångar med statlig utredningsprosa. Som artikeln påvisat återkommer arkiv-, bibliotek- och museifrågor explicit och implicit i hundratals SOU:er – även om vän av ordning säkert noterat att just termen kulturarv lyser med sin frånvaro. Den är visserligen belagd sedan 1887 och förekommer i flera betänkanden om kulturminnesvård redan under 1920-talet, men den är inte speciellt vanlig i mitt dataset. I 500-modellen förekommer den endast i kulturminnesvårdstema 386 – och då först på plats 120 (mätt i frekvens).

Något explicit tema kulturarv existerar alltså inte, trots artikelns titel – men att ABM-relaterade spörsmål återfinns i mitt dataset är uppenbart. På ett introducerande och explorativt manér har artikeln resonerat om de möjligheter (och tillkortakommanden) som temamodellering av ett historiskt dataset kan innebära. Fokus har legat på kulturarvsfrågor, men det kunde faktiskt lika gärna ha handlat om bostads- eller skattefrågor. I så måtto utgör den sammanställda SOU-datan en rikhaltig historievetenskaplig empiri. Temamodelleringens styrka ligger i första hand inte i att urskilja exakt vilka statliga utredningar som behandlat ABM-frågor; det känner museologisk och kulturhistorisk forskning till sedan tidigare. Att termen kulturarv inte figurerar i SOU-datan betyder inte att de idéer och värderingar som ingår i en kulturs historia och fungerar som en gemensam referensram – det vill säga kulturarv – inte kan studeras. Tvärtom. Temamodelleringens förtjänst är att den kan få korn på besläktade tematiker (som i de mest framträdande utredningarna) men i andra delar av datasetet, exempelvis i utredningar som tidigare har förbisetts av forskningen. Som påtalats är det i utredningen *Talböcker. Utgivning och spridning* (SOU 1982:7) som ABM-tema (nummer 39) förekommer allra mest frekvent i 100-modellen, inte inom betänkanden från 1965 års musei- och utställningssakkunniga (välciterade i den museologiska forskningen).

En annan uppenbar förtjänst med temamodellering som metod – vilket artikeln framför allt illustrerat genom grafer och nätverk – är att påvisa hur teman utvecklar sig över tid och på vilket sätt förbindelser uppträder vid olika tidsperioder. Genom att växla mellan modeller med olika antal teman går det också att studera hur en tematik (som den om kulturarv) förgrenar sig i mer specifika temamodeller (om exempelvis kulturminnesvård). På ett relativt enkelt sätt kan sådana modeller dessutom relateras till andra samhälleliga teman som återfinns inom samma dataset. De kan handla om forskning och universitet, om humaniora eller om data och information. Som artikeln återkommande påvisat är poängen med att lokalisera sådana förbindelser att ABM-frågor inte utretts i enskildhet. Snarare kan SOU-datan indikera relationer mellan en rad olika samhällsteman, vilka ibland kan förefalla något aparta – från vapenexport till fideikommiss. Samtidigt har artikeln också lyft fram att terminologi inom respektive

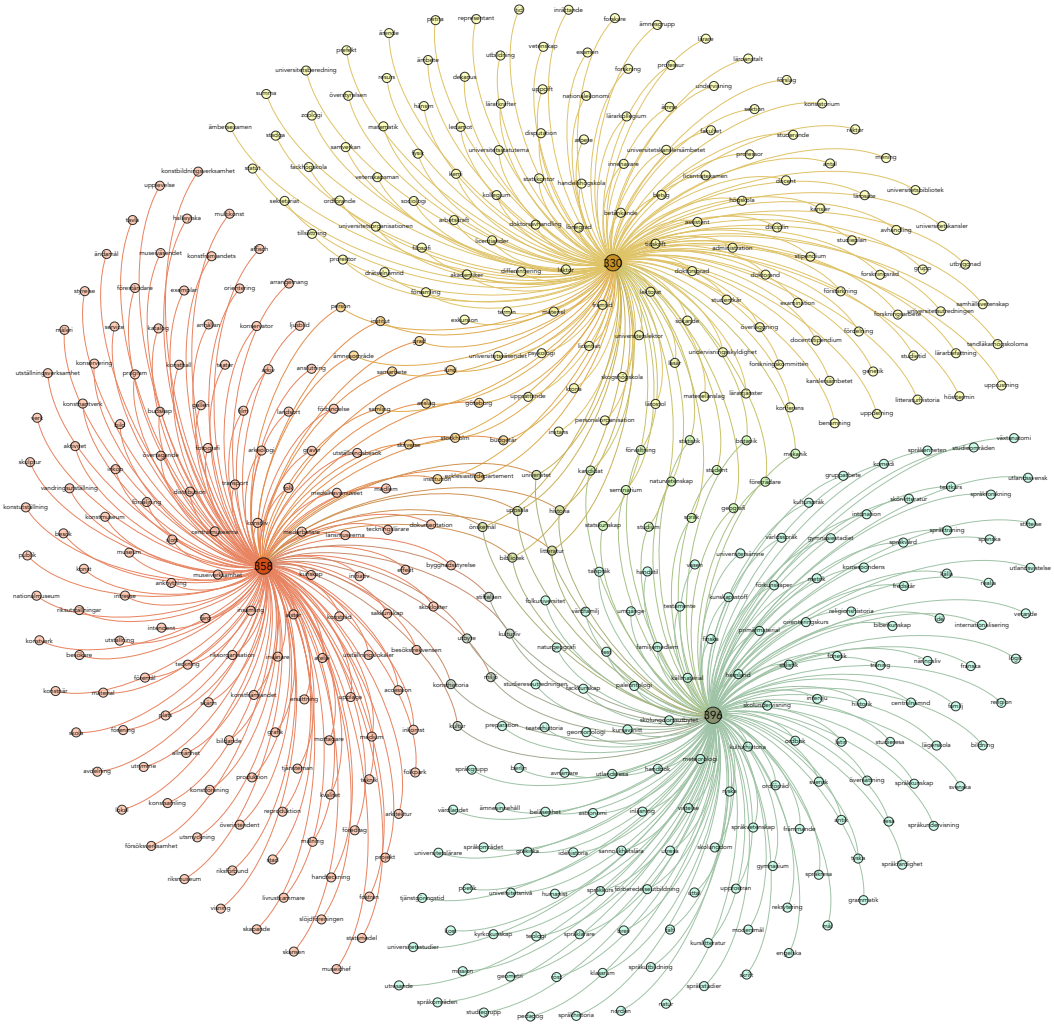


Illustration 21. Det cirkulära nätverket (med algoritmen Fruchterman Reingold) kopplar samman ord i tre temamodeller – 330 (om forskning och universitet), 358 (om museer) och 396 (om humaniora). Nätverket innehåller drygt 500 ordnoder där ord som konsthistoria och kultur förbinder musei- och humanioratemat. Sex centrala ord i mitten (i mörkare grönt) länkar till samtliga teman, bland dem historia, bibliotek och litteratur.

temamodell (främst beträffande de fyra kulturarvstematikern i 500-modellen) ofta är specifik. Inte sällan är det endast ett fåtal termer som i realiteten förbinder teman.

Grunden för all datadriven humanistisk forskning är att arkivens, bibliotekens och museers samlingar görs digitalt tillgängliga, det är en av utgångspunkterna för Vetenskapsrå-

dets Digarvprogram. Förutsättningen för att använda en metod som temamodellering är iordningställande av dataset, ett ofta resurskrävande arbete där algoritmisk maskininlärning samsas med manuell kurering av data. Inte sällan omstöper sådana processer ABM-sektorns empiri till dataset – det gäller även för äldre statliga utredningar. Att sammanställa alla SOU:er mellan 1945–89 till en enda text kan måhända förefalla överilat, ja rentav ett slags ahistoriskt tillvägagångssätt för att bedriva kulturhistorisk forskning. Ingen har ju haft för avsikt att läsa alla dessa utredningar. Men utifrån ett mediehistoriskt perspektiv behöver det inte vara fallet. Arkivens, bibliotekens och museernas samlingar är för mig som mediehistoriker alltid först mediala – därefter diskursiva. I den forskning jag bedrivit under mer än ett decennium, som bland annat handlat om att historisera dagens digitaliseringsverksamhet, har jag intresserat mig för de diskursiva systemens medialitet inom ABM-sektorn (Snickars 2020). Digital access och datadriven forskning utgör här den senaste utvecklingen. Det finns därför goda skäl att påminna om att nya mediebruk – som analysverktyg i Jupyter Lab-miljö – gör att synen på det förflutna ständigt förändras. I vilken medieteknisk form som historien kommer oss till mötes spelar roll, det gäller även för kulturarvsrelaterade frågor i storskaliga textuella dataset.

Noter

1. Tillsammans med konstprofessor Anna Dahlgren vid Stockholms universitet koordinerar jag Digarvprogrammet. För mer information om de forskningsprojekt som ingår, se https://www.digarv.se/deltagande-forskningsprojekt/>(senast kontrollerad 17/4, 2022).
2. Svensk efterkrigstid är en väl utforskad period, men genom att applicera digitala metoder på kurerade dataset är tanken med projektet att granska politikens, nyhetsmediernas och kulturens sfärer på nytt. Projektperiodens slut, 1989, är tämligen godtyckligt vald och handlar främst om att begränsa omfånget av de dataset vi använder. Fär mer information, se https://www.westac.se (senast kontrollerad 17/4, 2022).

Referenser

Statliga offentliga utredningar (SOU)

Arkiv inom hälso- och sjukvård (SOU 1968:53).
Arkiv- och biblioteksfilmning (SOU 1951:36).
Boken (SOU 1974:5).
De botaniska trädgårdarnas verksamhet och finansiering (SOU 1988:68).
Folkbibliotek i Sverige (SOU 1984:23).
Förbud mot utförsel av kulturföremål (SOU 1964:22).
Kulturminnesvård (SOU 1972:45).
Moderna arkivmedier (SOU 1976:68).
Museerna (SOU 1973:5).
Kunskap i samverkan – för samhällets utmaningar och stärkt konkurrenskraft (Prop. 016/17:50).
Landstingens arkiv (SOU 1975:71).
Talböcker. Utgivning och spridning (SOU 1982:7)
Telubaffären (SOU 1981:48).
Utbildning för bibliotek, arkiv och informatik (SOU 1969:37).
Utställningar (SOU 1974:43).
Öppenhet och minne (SOU 1988:11).

Litteratur

Aronsson, Peter & Magdalena Hillström 2005. *Kulturarvens dynamik: det institutionaliserade kulturarvets förändringar.* Norrköping: Tema Kultur.

Baltz, André & Fredrik Norén 2021. "Tematisk innehållsanalys med temamodellering." I Johan Jarlbrink & Fredrik Norén (red.). *Digitala metoder i humaniora och samhällsvetenskap.* Lund: Studentlitteratur, 211–234.

Bohman, Stefan & Lennart Palmqvist 1997 (red.). *Museer och kulturarv*. Stockholm: Carlsson.

Grinell, Klas & Anders Högberg 2020. "Lagstadgad kunskap. Om svensk museipolitik och forskning." *Nordisk Museologi* 29:2, 41–49. DOI: 10.5617/nm.8449

Guldi, Jo 2019. "Parliament's debates about infrastructure. An exercise in using dynamic topic models to synthesize historical change." *Technology and Culture* 60:1, 1–33.

Hesslén, Gunnar 1927. *Det svenska kommittéväsendet intill år 1905: dess uppkomst, ställning och betydelse.* Uppsala: Lundequistiska bokhandeln.

Hosseini, Kasra, Katherine McDonough, Daniel van Strien, Olivia Vane & Daniel C. S. Wilson 2021. "Maps of a nation? The digitized ordnance survey for new historical research." *Journal of Victorian Culture* 26:2, 284–299. DOI: 10.1093/jvcult/vcab009

Jarlbrink, Johan 2020. "All the work that makes it work. Digital methods and manual labour." I Mats Fridlund, Mila Oiva & Petri Paju (red.). *Digital Histories. Emergent Approaches Within the New Digital History.* Helsinki: Helsinki University Press, 113–126.

Johansson, Jan 1992. *Det statliga kommittéväsendet: kunskap, kontroll, konsensus,* Stockholm: Stockholm studies in politics.

Meijer, Hans 1956. *Kommittépolitik och kommittéarbete: det statliga kommittéväsendets utvecklingslinjer 1905–1954 samt nuvarande funktion och arbetsformer.* Lund: Gleerup.

Moretti, Franco 2013. *Distant reading.* London: Verso.

Norén, Fredrik & Pelle Snickars 2017. "Distant reading the history of Swedish film politics – in 4,500 governmental SOU reports." *Journal of Scandinavian Cinema* 7:2, 155–175.

Nystrand von Unge, Elin 2019. *Samla samtid. Insamlingspraktiker och temporalitet på kulturhistoriska museer i Sverige.* Stockholm: Stockholms universitet.

Näsman, Olof 2014. *Samhällsmuseum efterlyses. Svensk museiutveckling och museidebatt 1965–1990.* Umeå: Institutionen för kultur- och medievetenskaper.

Pettersson, Richard 2004. *Blick för kultur. Idéhistoriska aspekter på etnologisk och arkeologisk kulturforskning i Sverige under 1900-talet.* Umeå: Kulturgräns.

Regeringens proposition 2016/17:50. *Kunskap i samverkan – för samhällets utmaningar och stärkt konkurrenskraft.*

Snickars, Pelle 2015. "Mediestudiets infrastruktur. Om etableringen av Arkivet för ljud och bild." I Mats Hyvönen, Pelle Snickars & Per Vesterlund (red.). *Massmedieproblem. Mediestudiets formering.* Lund: Mediehistoriskt arkiv, 55–103.

Snickars, Pelle 2020. *Kulturarvets mediehistoria. Dokumentation och representation 1750–1950.* Lund: Mediehistoriskt arkiv.

Snickars, Pelle 2022. "100 miljoner ord. Reflektioner kring forskningsarbete med storskaliga textuella dataset som historisk empiri." *Historisk tidskrift* 142:3, 320–352.

Wittgren, Bengt 2013. *Katalogen: nyckeln till museernas kunskap? Om dokumentation och kunskapskultur i museer.* Umeå: Institutionen för kultur- och medievetenskaper.

Zetterberg, Kent 1990. "Det statliga kommittéväsendet." *Att styra riket: regeringskansliet 1840–1990.* Stockholm: Allmänna förlaget, 753–763.

Åmark, Klas 2005. *Hundra år av välfärdspolitik.* Umeå: Boréa.

Östberg, Kjell & Jenny Andersson 2013. *Sveriges historia 1965–2012.* Stockholm: Norstedt.

Pelle Snickars, professor
pelle.snickars@kultur.lu.se
Institutionen för kulturvetenskaper
Avdelningen för ABM & digitala kulturer
Lunds universitet
https://portal.research.lu.se/sv/persons/pelle-snickars