

Replacement with Kuana Exporter + Solution of "Unknown Channel" issues

Den originale larm-doms-exporter er i master branch på <https://github.com/statsbiblioteket/larm-doms-exporter> .

Den nye larm-kuana-exporter ligger i larm-kuana-branch .

Den underliggende software model er at larm-chaos har givet os en template fil hvori vi propper nogle values fra vores doms eller kuana metadata poster. Sideløbende er der en upload script som tage dagens eksport, pakker den ind i en zip-fil, og uploader gennem ftp.

Der er stor akitektur-mæssigt forskel mellem de to kodebase:

DOMS export code trækker direkte fra DOMS og bruger en persistence layer (database) til at holde styr på hvilke programmer den har fundet i DOMS og om de er blevet eksporteret til Larm.

Kuana koden trækker det hele fra Summa indeks og transformerer den til den endelige form mha xslt. Kuana pakken har ingen Java, kun scripts. Dens "persistence layer" består udelukkende af a en timestamp i en fil, som refererer til den sidste-opdateret timestamp af den sidste eksporteret program.

"Ukendt" Problematiken

Larm har sendt os en liste over kanaler i deres system med tilhørende filnavne for logo-filer. Disse har vi matchet med vores egne interne kanalnavne og lagt i en ad hoc xml file. I DOMS eksportereren ligger de i https://github.com/statsbiblioteket/larm-doms-exporter/blob/master/src/main/config/chaos_channelmapping.xml .

I kuana eksportereren er de indlejret i xslt filen: <https://github.com/statsbiblioteket/larm-doms-exporter/blob/larm-kuana-exporter/src/main/config/XIPToLarm.xsl> . (Scripstene er skrevet af Colin Rosenthal og xslt af Dorete.)

Når vi eksporter et program fra en kanal som ikke findes i listen, so får den channel name "Ukendt" og filnavn "unknown_logo.png".

Det er problematisk fordi det betyder at Larm har ikke tilstrækkeligt information for at vide om de har ret til at vise programmet. Problemet omfatter omkring 128000 udsendelser. Vi skal løse problemet både i den nye Kuana eksport og der skal også rettes til i den gamle DOMS eksporter så vi re-eksporter relevante programmer.

Løsnings-model

En mulig løsnings model ser sådan ud:

1. Vi udarbejder en liste af alle de påvirkede kanaler
 - a. enten med at bruge larm api: <https://api.prod.larm.fm/v6/EZSearch/Get?q=&facets={Search}.Kanal%3AUkendt&pageIndex=0&pageSize=100&format=json2>
 - b. eller med at sammeligne mapping-filen med en liste over alle vores kanaler
2. LARM-Chaos sletter alle nuværende programmer med "ukendt" fra deres system - og sender en liste over disse programmer til os
3. Vi sender kanalisten til LARM-Chaos og de fortæller os hvordan mapping af disse kanaler skal se ud fremover - dvs. hvilke nye entries der skal være i mapping xml.
4. Vi modificerer både den gamle og den nye eksporter sådan at de ikke eksporter "ukendte" kanaler
 - Den ville være bedste at indføre en nye betegnelse "frataget" i mapping-xml filen for kanaler hvor der er taget et bevidst valg at de ikke skal med, og så at man kan identificere egentlig "ukendte" kanaler i tilfælde at sådan nogle programmer dukker op senere.
 - Canal 8 Sport, Eurosport Danmark, Investigation Discovery, TV2 Sport HD, TV3 Max skal være blandt de "frataget"
5. Vi re-eksporter de slettede programmer - dvs. kun dem som ikke er "frataget"
6. Proceduren for fremtidigt behandling af nye kanaler er allerede lagt ind i <https://sbprojects.statsbiblioteket.dk/pages/viewpage.action?pageId=37594160>
7. Der er nogle timing issues da vi ikke skal blive ved at uploade "ukendte" programmer efter at Larm har slettet alle dem som findes p.t.. Det nemmeste er bare at sette en pause for den daglige kørsel af doms-exporter scriptsene mens vi retter på den.

Ukendte kanaler (fundet med larm api og opslag i pbcore på naiad)

sb kanal id	kanal navn	kommentar
drp4oj	DR P4 Østjylland	
tv3sport	TV3 Sport	
tv2mv	TV/MIDT-VEST	
drk	DR K	
tv2sport1hd	TV 2 Sport HD	

dr k	DR K	defekt pbcore? (08dd761a-e769-4db6-91b6-db6ab45f170d)
tv3max	TV3 MAX	
dr	DR	defekt pbcore? (01c48b87-2b40-42b4-86fc-7d908553ea84)
boomerang	Boomerang	
mtv	MTV	
idinvestigation	Investigation Discovery	
eurosportdk	Eurosport 2	
rakl	Radio Klassisk	
tv2sport	TV 2 Sport	
canal8sport	CANAL8 Sport	
voice	Radio The Voice Århus	
tv2	TV 2	
tlcdanmark	TLC	
AKTV	AKTV	
EfterskoleTV	EfterskoleTV	
Canal International	Canal International	defekt pbcore? (70ef33df-a8f1-4d2c-9b87-b9d75abc6cb1)
TVGaderummet	TVGaderummet	
Kaos TV	Kaos TV	defekt pbcore? (f8bcce7a-85ee-4f3a-ad39-5e6d3f91a73d)
SBS Net Syd	SBS Net Syd	defekt pbcore? (4546a5f1-17f6-4464-aa56-5df2612a7511)
tvd2	TV 2	defekt pbcore? (f61807c6-7ead-46b5-a677-7a2eb2b087bf)
tv 2	TV 2	defekt pbcore? (913ddf56-8f84-479f-a733-6e08a55fadde)
tvdk2	TV Danmark 2	
TVS - Den Danske Sportskanal	TVS - Den Danske Sportskanal	defekt pbcore? (164e97c0-66b7-47a2-ab34-ed884e033e4f)
TV 2	TV 2	defekt pbcore? (cbf137ad-dda4-4e2f-bfe2-e8b3b0187c5b)

Ukendte kanaler (fundet ved sammeligning af kanaler for ritzau programmer og mapping-filen)

sb kanal id	kanal navn	kommentar
drp4oj	DR P4 Østjylland	Samme kanal som drp4o, så skal nok have samme mapping
drx	DR X	
tv3sport	TV3 Sport	
bbcw	BBC World	
idinvestigation	Investigation Discovery	"Frataget"
bbce	BBC Earth	
ard	ARD	
tv3max	TV3 Max	"Frataget"
rakl	Radio Klassisk	
ndr	NDR	
disney	Disney	
nrk1	NRK 1	

3sat	3 SAT	
mtv	MTV	
eurosportdk	Eurosport Danmark version 2	
tv2f	TV2 Film	
kanalsyd	Kanal Syd	
toondisney	Toon Disney	
tv2mv	TV2 Midt & Vest	
arte	ARTE	
tv5mon	TV5 Monde	
svt2	SVT2	
zdf	ZDF	
drk	DRK	
tv2sport1hd	TV 2 Sport HD	"Frataget"
tlcdanmark	TLC Danmark	
rtl	RTL	
discov	Discovery	
bbck	BBC Knowledge	
viasport1	Viasat Sport 1	
nickdan	Nickelodeon Danmark	
svt1	SVT1	
boomerang	Boomerang	
natgeo	National Geographics	
animal	Animal Planet	
radiocity	Radio City	
bbcb	BBC Brit	
eurosportdan	Eurosport Danmark	"Frataget"
canal8sport	Canal 8 Sport	"Frataget"
drp4sj	DR P4 Sjælland	Burde være samme kanal som drp4re, så skal nok have samme mapping
voice	The Voice	Samme kanal som thevoice, så bør have samme mapping
drp4esb	DR P4 Esbjerg	
tlc	TLC	Samme kanal som tlcdanmark, så bør have samme mapping