



Direktoratet for IKT og fellestjenester
i høyere utdanning og forskning

FORELØPIG RAPPORT ARBEIDSGRUPPE CANVAS DATA

Vedlegg:
Introduksjon til Canvas Data Request-tabell
Canvas Data dokumentasjon (separat PDF)
Canvas Data learning outcome (separat PDF)



Innledning

Arbeidsgruppen for Canvas Data hadde oppstartsmøte 14. februar 2020. Formålet for gruppen var:

1. Kartlegge hva som finnes i Canvas Data (hvilke data, grensesnitt, format, sikkerhetsnivå, tilgangsstyring, hvor lenge det lagres osv.)
2. Starte diskusjon på hva som ønskes oppnådd med en satsing på Canvas Data og hvordan dette kan settes i sammenheng med andre overordnede satsinger på samling av data og/eller analyse – både i bottom-up og top-down perspektiv.
3. Undersøke hva som er gjort i Norden og EU rundt dette og personvernmessige vurderinger tilknyttet Canvas Data (eller andre sammenlignbare kasus).

Arbeidsgruppen har bestått av:

- Tore Jørgensen, teknisk Canvas (UiO)
- Per Bisseberg, teknisk Canvas toppadmin (HiØ)
- Sven Kjetil Haaø, programmering, standarder (USN)
- Karen Bjørndalen, juridisk (UiT) – TRUKKET
- Håvard Pedersen, teknisk programmering (UiT)
- Silje Maria Hansen, faggruppe sikkerhet/personvern (UiT)
- Lars Slettejord, teknisk rådgiver sikkerhet (UiT)
- Ole Martin Nodenes, Datavarehus (Unit)
- Elisabeth S. Guillot, tjenesteansvarlig LMS (Unit)
- Vegard Moen, seksjonsleder Undervisningstjenester (Unit)

Gruppen har hatt faste nettbaserte møter annenhver uke, der man har diskutert mulighetsrommet for bruk av data, samt meldt om status i arbeidet med kartleggingen av Data. Noen av møtene har også blitt brukt til demonstrasjoner og erfaringsdeling fra interessante sammenlignbare kasus (med demonstrasjoner fra eksterne deltakere). For arbeid mellom møtene har gruppen benyttet MS Teams, opprettet og administrert av Tore ved UiO.

Resultat

Kartlegging av data

Arbeidsgruppen har gått gjennom tilgjengelige datatabeller (Canvas Data request-tabeller) og funnet dokumentasjon på Canvas Data. Disse dokumentene er vedlagt rapporten.

Muligheter og behov

Kort oppsummert har arbeidsgruppen opplevd det som utfordrende å få kartlagt behovene/ønskene for læringsanalyse/dataanalyse blant ansatte, da de ansatte ikke selv forstår eller vet hva som er mulighetene. Utfordringen her er todelt, da dette er et nytt område man ikke er kjent med og det dermed blir vanskelig for faglige å oppfatte mulighetene og beskrive sine behov, samtidig som at det for tekniske er vanskelig å kommunisere mulighetsrommet da man mangler den pedagogiske og kontekstuelle forståelsen/erfaringen. Videre er det vanskelig å vite hva man bør/kan sammenstille av data for å svare på gitte spørsmål/problemstillinger.

Juridiske vurderinger

Arbeidsgruppen skulle undersøke om det er noen andre i Europa som har gjort juridiske vurderinger av bruk av data i/fra læringsplattformen Canvas i henhold til EUs personvernforordning. Til tross for



at mange i Europa har tatt i bruk Canvas Data/dataauthenting fra Canvas til analyse, har vi ikke lykkes i å finne noen som har en formell juridisk vurdering av dette.

I uformell samtale mellom tjenesteansvarlig LMS (Elisabeth S. Guillot) og jurist i Unit vurderes det at formål om studentoppfølging dekkes av studieforskriften og dataanalyse i forbindelse med dette i utgangspunktet er tillatt. Det anbefales likevel at en beskrivelse av data og formål sendes til Units jurister for å kunne vurderes i sektorsamarbeid for personvernombud, for å sikre omforent tolkning og vurdering. Dette må inkludere hvilke data som hentes ut og benyttes, samt at formål og behandlingsgrunnlag må klargjøres, hvilket betyr at vurdering som gjøres i dette fora kun vil være gjeldende for de spesifikke data, formål og behandlingsgrunnlag som fremstilles for juridisk vurdering.

Etablere kontakter som er gode på Canvas Data

Vegard Moen har koblet på Sverige og Danmark som har kommet noe lenger enn oss. Det er oppstartsaktiviteter hos Blekinge Tekniske Høgskole og ved Københavns universitet. Disse institusjonene har også kontakt med andre europeiske prosjekter innenfor Canvas Data. Begge institusjoner er interessert i å dele erfaringer videre.

Ellers er det funnet at JISC¹ i England har jobbet noe med læringsanalyse, men fra hva vi har funnet, har de ikke gjort grundige personvernmessige vurderinger.

Videre har gruppen funnet følgende nettsider som kanskje inneholder nyttig informasjon:

- <https://acla.amsterdam/>
- <https://www.surf.nl/en/report-learning-analytics-in-education-design-a-guide>

Demonstrasjon av bruk av Canvas-data i Power BI ved Høyskolen Kristiania (HK) 30/4-2020

Høyskolen Kristiania viste resultatene av sitt arbeid med å hente Canvas-data via et datavarehus inn i presentasjonsverktøyet Power BI.

HK er i oppstarten med å bruke Canvas-data og har foreløpig konsentrert seg om innlogginger i Canvas pr. emne. For de emnene med færre enn 80% innlogging vises det røde tall for emnet, dette betyr at det er mindre enn 80% innlogginger og da vil man foreta aktiviteter knyttet til tallene.

Spesifikt benytter de Canvas API for uthenting av data, der data hentes 1 gang per døgn og sendes til Power BI. Power BI kjører analyse (spørring) 1 gang per time. Power BI ved HK brukes «on-prem», altså ikke i sky-løsning.

Hovedsakelig benyttes anonymiserte data, der de ser på aktivitet i Canvas. For de situasjonene som er identifiserende er vurderingen til HK at dette er berettiget interesse for å hindre frafall (avd. For studentoppfølging har fulgt opp studentene). De har også lagt inn tillegg i sin Personvernerklæring om at innloggingsdata vil bli brukt for studentoppfølging.

I fremtiden ønsker HK å utvide løsningen og hente mer data fra Canvas, men sier også at dette er ressurskrevende da det krever mye kompetanse og tid.

Møte med SLATE (UiB) 15/4-2020

Møtet med SLATE fokuserte på erfaringsutveksling på hvordan man kan begynne med dataanalyse og hvilke behov man ser for seg. Foreløpig har SLATE begynt med en rapport om

¹ <https://www.jisc.ac.uk/>



påloggingsinformasjon og datasett fra studieavdelingen. De baserer databehandlingen per i dag på samtykke fra studentene (omlag 70% av studentene gir samtykke).

SLATE var særlig interessert i at analyse av bruk av Canvas kan være nyttig, gjerne også data/analyse som kan si noe om emnedesign. Spesielt spørsmål rundt bruk av Zoom, om studenter fortsetter å bruke dette på egenhånd, og om det har gitt noen effekt på deltakelse blant studenter ble trukket frem som interessante.

Videre anbefales det at man knytter sammen ulike datakilder om samme tema for å gi mer komplett/utfyllende analyse.

Demonstrasjon av Tableau v/STAR Datavarehus 19/5-2020

Ole Martin Nodenes gikk gjennom studieprogramlederrapporten i STAR Datavarehus. STAR tilbyr rapporter med studiedata i presentasjonsverktøyet Tableau, hovedsakelig fra FS.

<https://www.fellesstudentsystem.no/applikasjoner/star/>

I studieprogramlederrapporten kan studieprogramledere finne tilrettelagte dashboards for sitt program. Rapporten inneholder data om Søker- og studenttall, gjennomføring og frafall, score i Studiebarometeret (NOKUT), samt strykprosent og snittkarakter.

<https://www.fellesstudentsystem.no/applikasjoner/star/studieprogramledere.html>

Arbeidsgruppen for Canvas-data mente at det ville være ønskelig med et eget dashboard med Canvas-data for studieprogramledere, som kunne bli en del av studieprogramlederrapporten i STAR. En mulig use-case i oppstarten kunne være å se på andel av studentene innenfor et program og emnene i programmet, som har vært innlogget Canvas.

Andre mulige use-case som kom opp som aktuelle:

- Hvilke emner har stor aktivitet i Canvas?
- Hvis stor aktivitet, er det noen forskjell i hva som ligger i Canvas?
- Hvilken funksjonalitet blir brukt?
- Bruk av LTE-plugins

Veien videre

Arbeidsgruppen synes det er viktig å komme i gang med å prøve ut analyser, for å se mulighetsrom og hvordan det fungerer i praksis (så langt det er innenfor lovverket).

Mulige analyser å begynne med:

- Andel studenter som har logget på Canvas
- Når på døgnet er studenter påkoblet
- Aktivitet på emner
- Bruk av LTI-plugins
- Bruk av funksjonalitet i Canvas
- Type enhet/nettleser

Anbefaling fra gruppen

- Det bør gjøres en felles RoS-vurdering av hvor sensitive opplysninger som finnes i Canvas Data



- Foreløpig rapport bør sendes til personvernombud v/Unit slik at sektorsamarbeidet for personvernombud kan diskutere og vurdere formål og behandlingsgrunnlag
- Utvide studieprogramlederrapporten fra STAR til å inkludere data fra Canvas (se foreslåtte analyser over)
- Det bør utnevnes en arbeidsgruppe for videre arbeid med å implementere utvalgte kasus. Denne gruppen bør være forankret i flere ledd for å sikre at det blir avsatt tilstrekkelig med ressurser til arbeidet
 - Gruppen bør også inkludere ressurser med juridisk kompetanse for gjennomføring av RoS, vurdering av studieforskrift o.l.

Introduksjon til Canvas Data Request-tabell

av sven.k.haao@usn.no

Request-tabellen inneholder "alle" handlinger utført i Canvas, både de interaktive vha nettlesere og apper fra mobile enheter, og de interne handlingene i Canvas og de integrerte systemene.

Når vi skal ta uttrekk fra tekstfilene som Canvas Data genererer, må vi

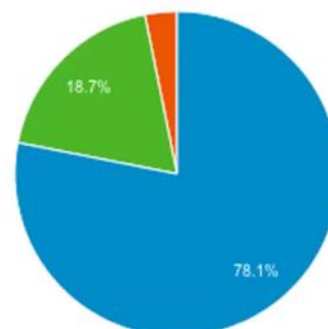
- kjenne dataene og hva de representerer
- ha konkrete bestillinger på hva dataene fra et uttrekk skal løse eller understøtte
- teste relevans opp mot evt andre liknende dataanalyser

Vi begynner med eksempler på bestillinger:

Eksempler på hva vi ønsker med uttrekk fra Canvas Data

- *Når jeg legger ut fagstoff,*
 - *må det tilpasses de enhetene studentene i virkeligheten bruker - og kommer til å bruke? (videre kursing av lærere)*
- *Hvilke LTI-er benyttes mest og av hvilke studieprogram? (økonomi)*
- *Hvilke studieprogram har mye frafall av studenter? (økonomi, trivsel, merkevarebygging)*

<input type="checkbox"/>	Device Category	Sessions	Sessions
		371,491 <small>% of Total: 100.00% (371,491)</small>	371,491 <small>% of Total: 100.00% (371,491)</small>
<input type="checkbox"/>	1. ■ desktop	290,043	78.08%
<input type="checkbox"/>	2. ■ mobile	69,614	18.74%
<input type="checkbox"/>	3. ■ tablet	11,834	3.19%



Hvilke nettlesere benyttes?

1.	Chrome	219,090	58.98%
2.	Safari	55,001	14.81%
3.	Safari (in-app)	36,385	9.79%
4.	Android Webview	18,786	5.06%
5.	Edge	17,991	4.84%
6.	Firefox	17,051	4.59%
7.	Internet Explorer	5,692	1.53%
8.	Samsung Internet	996	0.27%
9.	Amazon Silk	230	0.06%
10.	Opera	151	0.04%

Prosessering av request-tabellen fra Canvas Data

- Skille ut brukerinteraksjoner fra Canvas' interne handlinger
- Sortere data etter informasjon i adressefeltet

Brukerinteraksjoner foretas av en "User agent" - feks en nettleser som åpner en wiki-side

Request-tabellen inneholder over 100 forskjellige User agents så vi må lete etter de som er relevante for vår analyse.

Case:

Vi vil finne hvor stor andel av interaksjonene som foretas vha mobile enheter. Dette kan gi en lærer s tøtte for å droppe tabeller på sine Canvas sider, siden disse er lite egnet for mobile enheter.

User agents i request-tabellen for mobile enheter er pt.:

Canvas app for iOS	Canvas app for Android	Nettleser på telefon/nettbrett:	Nettleser på Windows 10 (Chrome 69.0)
<ul style="list-style-type: none"> • CanvasKit/ • iCanvas/ 	Candroid/	<ul style="list-style-type: none"> • Mozilla/5.0 (iPad; CPU OS 7_0 like Mac OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11A465 Safari/9537.53 • Mozilla/5.0 (Linux; U; Android 4.0.3; ko-kr; LG-L160L Build/IML74K) AppleWebKit/534.30 (KHTML, like Gecko) Version/4.0 Mobile Safari/534.30 	Mozilla/5.0 (Windows NT xx.x_Win64_x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/xx.xx.xx Safari/537.36



Eksempler for User agents for Windows og MAC :

Nettleser på Windows 10 (Firefox 64.0)	Nettleser på macOS (Safari 12.0)	Nettleser på Ubuntu Linux (Firefox 64.0)
<ul style="list-style-type: none"> Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:64.0) Gecko/20100101 Firefox/64.0 (Edge 17.0): Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/64.0.3282.140 Safari/537.36 Edge/17.17134 	<ul style="list-style-type: none"> Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_2) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/12.0.2 Safari/605.1.15 	Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:64.0) Gecko/20100101 Firefox/64.0

Én oppføring i request-tabellen kan se slik ut:

```
5a97f077-2d96-4d2d-82e5-8c076d649c472017-02-10 07:20:07.76920172017-022017-02-1032146627200619346\N10231000000000001\N\N\N\N\N\users/88Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/46.0.2486.0 Safari/537.36 Edge/13.10586GET128.39.204.86328610usersshowUser88\N73f4f5285763a0f5e4531da30897f6ab198698370476673637200HTTP/1.1\N
```

Dataene den representerer er:

id	guid	Request ID assigned by the canvas system to the request.
timestamp	datetime	Timestamp when the request was made in UTC.
timestamp_year	varchar	Year when the request was made.
timestamp_month	varchar	Month when the request was made.
timestamp_day	varchar	Day when the request was made.
user_id	bigint	Foreign key in user_dim for the user that made the request. If the request was made by one user masquerading as another, then this column contains the ID of the user being masqueraded as.
course_id	bigint	Foreign key in course_dim for the course that owned the request. Set to NULL if not applicable.
root_account_id	bigint	Foreign key in account_dim for the root account on which this request was made.
course_account_id	bigint	Foreign key in account_dim for the account the associated course is owned by.
quiz_id	bigint	Foreign key in quiz_dim if the request is for a quiz, otherwise NULL.
discussion_id	bigint	Foreign key in discussion_dim if the request is for a discussion, otherwise NULL.
conversation_id	bigint	Foreign key in conversation_dim if the request is for a conversation, otherwise NULL.



assignment_id	bigint	Assignment foreign key if the request is for an assignment, otherwise NULL.
url	text	URL which was requested.
user_agent	text	User agent header received from the users browser/client software.
http_method	varchar	HTTP method/verb (GET, PUT, POST etc.) that was sent with the request.
remote_ip	varchar	IP address that was recorded from the request.
interaction_micros	bigint	Total time required to service the request in microseconds.
web_application_controller	varchar	The controller the Canvas web application used to service this request.
web_application_action	varchar	Controller the Canvas web application used to service this request.
web_application_context_type	varchar	Containing object type the Canvas web application used to service this request.
web_application_context_id	varchar	Containing object's ID the Canvas web application used to service this request.
real_user_id	bigint	If the request was processed by one user masquerading as another, then this column contains the real user ID of the user.
session_id	varchar	ID of the user's session where this request was made.
user_agent_id	bigint	(Not implemented) Foreign key to the user agent dimension table.
http_status	varchar	HTTP status of the request.
http_version	varchar	HTTP protocol version.
developer_key_id	bigint	ID of the developer who accessed this resource if the request was made by a developer.

Disclaimer: The data in the requests table is a 'best effort' attempt and is not guaranteed to be complete or wholly accurate.

Her er et eksempel på bruk av python for å trekke ut relevante user_agents:

```
# by sven_k_haao@usn.no
substring0 = "iPhone"
substring1 = "CanvasKit"
substring2 = "iCanvas"
substring3 = "candroid"
substring4 = "Mozilla/5.0 (iPad"
#substring5 = "Windows"
substring5 = "intel mac"
#substring5 = "Linux"

f = open("requests-00000-1099e4ac.", "r")
n0 = 0
n1 = 0
n2 = 0
n3 = 0
n4 = 0
n5 = 0
ant = 0

for x in f:
# print(x),
# print(x[20:25] + x[len(x)-7:len(x)].lower())
# print(x[172:300])
# print(x[x.find(x, " "):300])
# print(x.split(","))
```

Vedlegg



```
# print(x.find(substring1))
# print(x[x.find(substring1):300])
ant += 1

if substring0 in x:
    n0 += 1
    # print(str(n0) + " " + x + " " + substring)
elif substring1 in x:
    n1 += 1
# print(str(n1) + " " + x + " " + substring)
elif substring2 in x:
    n2 += 1
# print(str(n2) + " " + x + " " + substring)
elif substring3 in x:
    n3 += 1
# print(str(n3) + " " + x + " " + substring)
elif substring4 in x:
    n4 += 1
# print(str(n4) + " " + x + " " + substring)
elif substring5 in x.lower():
    n5 += 1
# print(str(n5) + " " + x + " " + substring)
# if ant == 300000:
#     break

else:
    print(str(n0) + " " + substring0)
    print(str(n1) + " " + substring1)
    print(str(n2) + " " + substring2)
    print(str(n3) + " " + substring3)
    print(str(n4) + " " + substring4)
    print(str(n5) + " " + substring5)
    print(str(ant) + " totalt")
```