


Deall eich dinasyddion gan ddefnyddio data cyfryngau cymdeithasol:

*Adolygiad o ba ddata cyfryngau
cymdeithasol sydd ar gael, sut
i'w gael, a beth mae'n ei olygu
ar gyfer llywodraeth leol*

Cynhyrchwyd gan: Data Cymru
Dyddiad: Gorffennaf 2022

 029 2090 9500

 enquiries@data.cymru

 www.data.cymru

Mae'r adroddiad hwn ar gael yn [Saesneg](#).

Cynnwys

Crynodeb Gweithredol	1
Cyflwyniad.....	3
Ystyriaethau ymarferol a moesegol	5
1 Defnyddio dim ond gwybodaeth sydd ar gael yn gyhoeddus	5
2 Peidio â chyhoeddi cynnwys defnyddiwr yn uniongyrchol heb gydsyniad ar sail gwybodaeth	6
3 Anonymeiddio'r holl ddata	6
4 Dim cysylltu data ymlaen	7
5 Cyfyngu gwneud penderfyniadau ar sail algorithmau	7
Cyrchu data cyfryngau cymdeithasol.....	8
Canfyddiadau cyffredinol	16
Profion penodol a gynhaliom ar COSMOS	19
Casgliad.....	26
Adnoddau pellach	28

Crynodeb Gweithredol

Mae defnydd cyfryngau cymdeithasol yn helaeth. Mae awdurdodau lleol a'r cyhoedd fel ei gilydd yn defnyddio cyfryngau cymdeithasol yn eang at amrywiaeth o ddibenion; o drafod y materion sy'n bwysicaf iddynt, i drafod newid mewn polisi lleol, i leisio eu pryderon. Mae'r holl ryngweithredu hwn yn creu data a allai gynnwys mewnwelediad gwerthfawr.

Mae data cyfryngau cymdeithasol ar ffurf cynnwys postiadau. Ond, yn ogystal, mae'n cynnwys gwybodaeth am y postiad, fel pryd cafodd ei gyhoeddi, ym mha iaith, ym mha ardal ddaearyddol, a chan ba gyfrif.

Nod yr ymchwil hon (a gafodd ei chynnal yn 2021-22) oedd deall a oedd y data hwn ar gael, yn ddefnyddiadwy ac o fudd i'r sawl sy'n gwneud penderfyniadau mewn llywodraeth leol yng Nghymru.

Cyfngom ni'r ymchwil gychwynnol hon i ddulliau o gyrchu **data Twitter** yn unig, a hynny am ei fod yn cael ei ddefnyddio'n eang gan gydweithwyr llywodraeth leol ac mae Twitter yn trefnu bod ei ddata ar gael yn gymharol hawdd. Roeddem yn hyderus, felly, y gallem ni gael gafael ar set ddata a gwneud asesiad cychwynnol o ansawdd a defnyddioldeb. Edrychom ar lawer o opsiynau o ran cyrchu data Twitter, gan gynnwys:

- Rhyngwyneb rhaglennu cymwysiadau (API)¹ Twitter ei hun, dull o ganiatáu i ddau gymhwysiad cyfrifiadur rannu data yn gyflym ac yn rhwydd
- Dau gynnig masnachol ar gyfer rhannu a dadansoddi data cyfryngau cymdeithasol;
 - Synthesio, gwasanaeth gwrando cymdeithasol; ac
 - NCapture for NVivo, offeryn sy'n cael ei ddefnyddio fel arfer am ddadansoddiad ansoddol sydd wedi lansio ychwanegyn yn ddiweddar i gipio data Twitter.
- Platfform a ddatblygwyd gan Labordy Gwyddor Data Prifysgol Caerdydd i wneud data Twitter yn fwy hygyrch i gydweithwyr sector cyhoeddus, sef The Collaborative Online Social Media Observatory (COSMOS).

Yn dilyn adolygiad o'r bedair ffynhonnell, cawsom mai **COSMOS** oedd y dewis gorau am ymchwiliad cychwynnol am ei fod yn **dileu'r rhwystrau technegol a chost** sydd ynghlwm wrth ddulliau eraill o gyrchu data.

¹ Mae Rhyngwyneb Rhaglennu Cymwysiadau (API) yn dechnoleg sy'n caniatáu i ddau seilwaith TG ryngweithredu. Maent yn cael eu defnyddio'n gyffredin gan sefydliadau – fel Twitter – i agor eu data i ddatblygwyr ac ymchwilwyr.

Gan symud i'r dadansoddiad, roeddem yn gallu defnyddio COSMOS i gynnal tri o'n pum prawf yn llwyddiannus. Yn gyntaf, roeddem yn gallu casglu data am ryngweithiadau â rhestr benodedig o gyfrifon ar draws llywodraeth leol a Llywodraeth Cymru. Yn ail, roeddem yn gallu casglu set ddata ar sail chwilio allweddairiau. Yn drydydd, roedd modd atgynhyrchu'r profion hyn, a hybodd ein hyder yn y data ac yn y fethodoleg.

Fodd bynnag, er ei fod yn hynod o ddefnyddiol, roedd gan COSMOS ei **gyfyngiadau** hefyd. Yn benodol, nid oeddem yn gallu llunio set ddata a oedd yn cynnwys trydarau Cymraeg yn unig; hefyd mae angen allweddair neu hidlydd cyfrif wrth idlo yn ôl iaith, gan gyfyngu'r chwiliad. Yn olaf, ac yn fwy ymarferol, mae COSMOS **ond yn gallu cael ei drwyddedu i sefydliadau academiaidd neu nid-er-elw**.

Problem gyffredin a welsom gyda data Twitter yn ehangach (sut bynnag y cafodd ei gyrchu) yw nad yw'n bosibl casglu "**pob trydar sy'n deillio yng Nghymru**" neu "**pob trydar sy'n crybwyll llywodraeth leol yng Nghymru**". Y rheswm am hyn yw nad yw trydarau, yn ddi-ofyn, yn gael eu **geotagio** gan yr awdur; nodwedd i optio i mewn iddi yw hon.

I grynhoi, adolygom ni bedwar opsiwn am gyrchu data Twitter a gwelsom fod gan bob un ei rinweddau. Fodd bynnag, dewisom ymchwilio i COSMOS yn fanylach oherwydd ei ddiffyg rhwystrau technegol ac ariannol. Cynhaliom bum prawf ar y data Twitter a grynhowyd. Ar gyfer prosiectau sy'n golygu crynhoi a dadansoddi data yn ôl cyrfif neu allweddair, cawsom fod y data ar gael a'i fod o ansawdd da. Fodd bynnag, nid oedd modd i ni grynhoi na dadansoddi data Twitter yn ôl daearyddiaeth neu iaith, oherwydd y ffordd mae'n cael ei gipio wrth y ffynhonnell.

Mae deall y cryfderau a'r cyfyngiadau hyn yn allweddol i benderfynu pryd, a sut, i ddefnyddio data Twitter. Drwy ddangos bod data defnyddiol cyfryngau cymdeithasol yn gallu cael ei gyrchu'n gyflym, yn hawdd, ac yn foesebol, mae'r prosiect peilot hwn wedi cyflawni ei nodau a'r gobaith yw y bydd yn ysgogi defnydd pellach ar ddata cyfryngau cymdeithasol ar draws llywodraeth leol yng Nghymru.

Cyflwyniad

Mae cynghorau lleol yn dibynnu ar ddeall barn dinasyddion wrth gyflenwi a gwella gwasanaethau cyhoeddus yng Nghymru ac mae platfformau cyfryngau cymdeithasol yn chwyldroi sut mae hyn yn digwydd. Mae gan y mwyafrif o gynghorau yng Nghymru o leiaf un presenoldeb cyfryngau cymdeithasol, fel arfer Twitter a / neu Facebook, maen nhw'n ei ddefnyddio i ledaenu a chasglu gwybodaeth, yn ogystal ag i ymgysylltu â defnyddwyr.

Yn 2020, canfu arolwg gan y Swyddfa Ystadegau Gwladol² fod gan 96% o aelwydydd fynediad i'r rhyngwrld bellach, a bod tua 70% o'r bobl hyn yn ymgysylltu â chyfryngau cymdeithasol. Mewn astudiaeth wahanol, canfu OfCom³ fod oedolion y DU sy'n defnyddio'r rhyngwrld yn treulio tua phedair awr y dydd ar-lein ar gyfartaledd. Roedd 18% o'r amser oedolion y DU ar-lein a fesurwyd ym Medi 2021 ar apiau a safleoedd fel Facebook, Instagram a WhatsApp (sydd i gyd ym mherchnogaeth y cwmni rhiant Meta) a bod 14% yn cael ei dreulio ar Google a YouTube (sydd ym mherchnogaeth y cwmni rhiant Alphabet).

Mae data sy'n cael ei gasglu trwy gyfryngau cymdeithasol yn gallu bod yn ffynhonnell amhrisiadwy gwybodaeth 'ar y llawr', sy'n cynnig adborth crai yn y fan a'r lle. Mae'r data hwn ar ffurf cynnwys postïadau⁴, ond hefyd gwybodaeth am y postïad, fel amser ei bostio, iaith, data daearyddol, a data proffil defnyddwyr. Mae'r metadata⁵ hwn yn ddefnyddiol i ganfod y patrymau sydd y tu ôl i gynnwys postïadau. Mae modd ei ddefnyddio i raglennu algorithmau sy'n gallu nodi meta-tueddiadau, fel:

- y rhwydweithiau o bostiadau a phostwyr cysylltiedig;
- eu lleoliad tebygol a phroffiliau demograffaid; a'r
- mathau o bostiadau byddant o bosibl yn eu gwneud yn y dyfodol.

Ond er mwyn i ddadansoddiad o'r fath gael ei gynnal, mae'n rhaid i'r data fod o ansawdd a maint digonol, mae heriau strwythurol ynghlwm wrth gasglu a defnyddio data cyfryngau cymdeithasol. Yn benodol, mae cyfryngau cymdeithasol yn mynnu caledwedd ddigidol, cysylltiad â'r rhyngwrld, a sylfaen defnyddwyr gweithredol. Mae rhannau arwyddocaol o'r boblogaeth sy'n llai actif yn ddigidol nag eraill, sy'n gallu cyflwyno tuedd i'r data.

²[Internet access – households and individuals, Great Britain, 2020'. Y Swyddfa Ystadegau Gwladol \(2020\)](#)

³['Communications Market Report 2022'. OfCom \(2022\)](#)

⁴ Mae postïad yn cael ei ddiffinio fel "cynnwys a rennir ar gyfryngau cymdeithasol trwy gyfrwng proffil defnyddiwr".

⁵ Mae metadata yn ddata am y data. Er enghraifft, byddai metadata postïad cyfryngau cymdeithasol yn cynnwys gwybodaeth am yr awdur, y math o neges, dyddiad ac amser ac yn y blaen.

Ar ben hynny, mae'n bosibl nad yw ymddygiad ar-lein yn adlewyrchu ymddygiad all-lein. Mae defnyddwyr cyfryngau cymdeithasol yn deall eu bod yn cynhyrchu ac yn cyhoeddi cynnwys am gynulleidfa. Rhaid i ymchwilyr fod yn ymwybodol o'r duedd bosibl hon wrth feddwl am ddefnyddio'r cyfryw ddata.

I ymchwilyr a datblygwyr, mae mynediad i ddata cyfryngau cymdeithasol fel arfer trwy gyfrwng Rhyngwyneb Rhaglennu Cymwysiadau (API), technoleg sy'n caniatáu i ddau seilwaith TG ryngweithredu. Mae APIs yn ddulliau grymus o rannu symiau mawr o ddata. Fodd bynnag, eu perchennog sy'n rheoli ym mha fodd maen nhw'n darparu data. Er enghraifft, mae Twitter, Facebook, neu YouTube yn gallu defnyddio APIs i gyfyngu a siapio'r data maen nhw'n ei ryddhau. Mae hyn hefyd yn gallu gwro'r data sy'n cael ei ryddhau i'r cyhoedd drwy gapio maint a math y data a ddangosir yn artiffisial.

Yng ngoleuni pwysigrwydd cyfryngau cymdeithasol a'u rôl mewn cysylltu cynghorau lleol â dinasyddion, penderfynom ni ymchwilio i ddulliau o gyrchu data cyfryngau cymdeithasol. Nod yr ymchwil hon (a gafodd ei chynnal yn 2021-22) oedd deall a oedd data ar gael, yn ddefnyddiadwy ac o fudd i'r sawl sy'n gwneud penderfyniadau mewn llywodraeth leol yng Nghymru.

Ystyriaethau ymarferol a moesegol

Mae cyfryngau cymdeithasol yn ffordd gymharol newydd i bobl a sefydliadau ryngweithio. Mae'r 15 mlynedd diwethaf wedi gweld newid aruthrol yn nifer a math y rhyngweithiadau cymdeithasol ar-lein. Gyda'r chwyldro hon daw risg newydd a rheoleiddio newydd. Mae'n bosibl y bydd y rheoleiddio newydd yn methu â digwydd ar yr un cyflymder ag arloesi, normau'r diwydiant data a barn y cyhoedd. Mae'n hanfodol, felly, bod llywodraethiant cryf o amgylch prosiectau sy'n defnyddio data cyfryngau cymdeithasol. Mae hyn yn arbennig o wir yn achos y sector cyhoeddus, lle mae ymddiriedaeth yn gallu cael ei thorri'n hawdd ac yn anodd ei hadennill.

Am hynny, yn ogystal â'n prosesau Rheoliad Diogelu Data Cyffredinol y DU (UK GDPR), treulïom ni gryn dipyn o amer yn trafod a ddylem ymchwilio'n foesegol i ddefnydd data cyfryngau cymdeithasol a sut mae gwneud. Manteisiom ar sawl egwyddor allweddol i'n llywio, ar sail canllawiau presennol, gan gynnwys rhai Gwasanaeth Ymchwil Gymdeithasol (GSR)⁶ y Llywodraeth.

1 Defnyddio dim ond gwybodaeth sydd ar gael yn gyhoeddus

Mae safleoedd cyfryngau cymdeithasol yn gadael i unigolion a sefydliadau gyhoeddi cynnwys i'r cyhoedd. Fodd bynnag, maen nhw hefyd yn hwyluso cyfathrebu uniongyrchol rhwng defnyddwyr ar ffurf negeseuon (er enghraifft drwy swyddogaeth sgwrsio, Negeseuon Uniongyrchol ac ati). Mae'r rhain yn hygyrch trwy APIs cyfryngau cymdeithasol mewn rhai achosion.

Gallai fod heriau moesegol wrth ddefnyddio cyfathrebiadau uniongyrchol a chynnwys a gyhoeddir mewn fforymau sydd ag aelodaeth gaeedig (h.y. nid ydynt yn gyhoeddus). Er bod platfformau cyfryngau cymdeithasol yn gwneud yn glir yn eu telerau ac amodau y gall y cyfryw ddata gael ei rannu at ddibenion ymchwil a datblygu, efallai na fydd defnyddwyr yn deall hyn yn llawn neu fod ganddynt ddisgwyliadau gwahanol am sut y gall data sy'n cael ei rannu mewn grwpiau cyfyngedig gael ei drin. Felly, efallai ei bod yn bosibl amau a wnaeth defnyddwyr cyfryngau cymdeithasol roi eu cysdyniad llawn ar sail gwybodaeth⁷ i ddata sy'n cael ei gasglu drwy'r llwybrau hyn gael ei ddefnyddio mewn ymchwil.

Gan nad oedd arnom ni angen data o'r fath am y prosiect hwn, dewisom ni beidio â chynnwys unrhyw gyfathrebiadau uniongyrchol, negeseuon o fewn grwpiau caeedig, na data o unrhyw fforwm aelodau cyfyngedig arall yn ein hymchwil. Ailadroddwn mai

⁶

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524750/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf.

⁷ Mae cysdyniad ar sail gwybodaeth yn gysyniad pwysig mewn moeseg data ac ymchwil. Mae'n sicrhau bod unigolyn yn gwbl ymwybodol o oblygiadau a risgiau rhannu eu data neu gymryd rhan mewn ymchwil. Mae cysdyniad dilys yn syniad cysylltiedig yn neddfwriaeth GDPR.

cynnwys postïadau cyhoeddus a oedd o ddiddordeb i ni, nid cyfathrebiadau uniongyrchol rhwng defnyddwyr.

Defnyddiom ni ddim ond data cyfryngau cymdeithasol a oedd ar gael yn gyhoeddus.

2 Peidio â chyhoeddi cynnwys defnyddiwr yn uniongyrchol heb gydsyniad ar sail gwybodaeth

Cynnwys defnyddiwr yw unrhyw wybodaeth mae'n ei phostio ar gyfryngau cymdeithasol, gan gynnwys postïadau, lluniau, fideos, sylwadau ac ati. Os bydd yr achos dros ei ddefnyddio yn ei gyfiawnhau ac mae'r prosesau cyfreithiol priodol wedi'u dilyn, gall fod yn bosibl dyfynnu cynnwys unigolion os:

- cysylltwyd â'r defnyddiwr am gydsyniad ar sail gwybodaeth; a / neu
- deilliodd y cynnwys o gyfrifon sefydliadau cyhoeddus neu ffigurau cyhoeddus uchel eu proffil, fel y Prif Weinidog neu Arweinydd cyngor.

Mae hyn yn cyd-fynd â thelerau ac amodau platfformau cyfryngau cymdeithasol fel Twitter a'r offerynnau casglu data cyfryngau cymdeithasol a drafodir yn yr adroddiad hwn.

Fodd bynnag, nid yw hyn yn atal data cyfryngau cymdeithasol rhag cael ei ddefnyddio i adrodd gwybodaeth gryno. Er enghraifft, niferoedd trydarau i ac oddi wrth rai cyfrifon llywodraeth leol neu nifer y trydarau sy'n cynnwys geiriau neu hashnodau (#) penodol.

Ni chafodd unrhyw gynnwys defnyddwyr ei gyhoeddi yn yr adroddiad hwn am nad oedd yn angenrheidiol at ddibenion ein hymchwil.

3 Anonymeiddio'r holl ddata

Mae anonymeiddio data yn cyd-fynd â'r arfer gorau o fewn cymuned ehangach ymchwil cyfryngau cymdeithasol^{8,9}. Anaml iawn y byddai angen i ddata wedi'i ddad-anonymeiddio gael ei gyhoeddi. Pe bai'n angenrheidiol, byddai ar yr ymchwilydd angen cydsyniad ar sail gwybodaeth gan yr awdur neu reswm cyfreithiol arall (fel a amlinellir yn neddfwriaeth UK GDPR) am wneud yr wybodaeth adnabyddadwy'n gyhoeddus.

Yn yr ymchwil hon, cafodd yr holl ddata ei anonymeiddio fel nad oedd modd adnabod unrhyw unigolyn.

⁸ https://www.gla.ac.uk/media/Media_487729_smxx.pdf.

⁹

<https://www.york.ac.uk/media/staffhome/research/documents/researchgovernance/codeofethics/Guidelines%20for%20the%20Use%20of%20Social%20Media%20Data%20in%20Research%20for%20web%20page.pdf>.

4 Dim cysylltu data ymlaen

Gall cysylltu data cyfryngau cymdeithasol defnyddiwr â ffynonellau gwybodaeth eraill gynnig dirnadaeth gyfoethog. Ond gallai cysylltu'r ffynonellau data ddatgelu mwy am unigolyn nag yr oedden nhw wedi bwriadu o gynnwys eu postiad yn unig.

Mae cysylltu data yn gallu cael ei ddefnyddio lle bo'n briodol ac yn gyfiawnadwy. Yn yr un modd â gwneud penderfyniadau ar sail algorithmau, mae'n rhaid deall a dogfennu risgiau, effaith a chyfiawnhad cyn bod prosesu'n gallu dechrau.

At ddibenion ein hymchwil ni, nid oedd angen unrhyw gysylltu data ymlaen.

5 Cyfyngu gwneud penderfyniadau ar sail algorithmau

Mae gwneud penderfyniadau ar sail algorithm yn derm technegol am greu a defnyddio model ystadegol i wneud penderfyniadau'n awtomatig. Gall y dechneg fod yn ffordd rymus o brosesu a dosbarthu llawer o wybodaeth yn gyflym ac yn effeithlon. Er enghraifft, rhagfynegi rhywedd, lleoliad, neu farn wleidyddol defnyddiwr Twitter ar sail eu proffil a chynnwys eu trydarau.

Mae'n amlwg bod pŵer mewn gallu prosesu symiau mawr iawn o ddata yn gyflym a chynhyrchu data a mewnwelediadau newydd o ddata sy'n bodoli'n barod. Ond mae risg nad yw algorithm wedi cael ei bennu neu ei brofi'n briodol, a allai arwain at ddata allbwn annibynadwy neu ragfarnllyd.

Ar ben hynny, gallai'r dechneg hon gael ei defnyddio i gynhyrchu gwybodaeth na wnaeth defnyddiwr cyfryngau cymdeithasol ei rhannu o'u gwirfodd. Er enghraifft, drwy brosesu data am defnyddiwr Twitter, gallai fod yn bosibl rhagfynegu eu credoau gwleidyddol. Ni fyddai'n briodol gwneud y rhagfynegiad hwn ac yna cyhoeddi'r canlyniadau ar sail algorithm yn unig.

Oherwydd y risgiau hyn, mae'n bwysig bod effeithiau, a chyfiawnhad am ddefnyddio, prosesu algorithmig yn cael eu deall a'u dogfennu cyn i brosesu ddechrau. Yn yr un modd, mae'n bwysig bod gwneud penderfyniadau ar sail algorithm yn cael ei gyflwyno a'i ddefnyddio fel offeryn i gynorthwyo penderfyniadau gan fodau dynol, nid i gymryd eu lle.

Byddem yn argymhell eich bod yn trafod y materion hyn â'r staff llywodraethu gwybodaeth a moeseg perthnasol yn eich sefydliad cyn casglu unrhyw ddata.

Ni wnaethom benderfyniadau ar sail algorithmau fel rhan o'r ymchwil hon.

Cyrchu data cyfryngau cymdeithasol

Mae llawer o blatfformau cyfryngau cymdeithasol a llawer o ddulliau o gyrchu eu data. Dengys **Tabl 1** amcangyfrif o nifer defnyddwyr deg platfform cyfryngau cymdeithasol mwyaf poblogaidd ar lefel Cymru a'r DU (mae rhestr lawn ar gael o'r ffynhonnell). Cyhoeddir y data hwn yn Adroddiad y Farchnad Cyfathrebu OfCom 2022 ac mae wedi'i seilio ar ddata a gafwyd gan Ipsos Iris¹⁰.

Tabl 1 – Nifer defnyddwyr platfformau cyfryngau cymdeithasol yn ôl ardal yn ystod Medi 2021

Platfform cyfryngau cymdeithasol	Nifer defnyddwyr cyfryngau cymdeithasol yng Nghymru (miliynau)	Nifer defnyddwyr cyfryngau cymdeithasol yn y DU (miliynau)
Facebook a Messenger	2.1	46.7
YouTube	2.1	45.6
Instagram	1.5	36.3
Twitter	1.4	30.8
LinkedIn	0.9	18.8
Pinterest	0.9	16.6
TikTok	0.7	15.4
Snapchat	0.7	11.8
Nextdoor	0.3	9.2
Google	0.4	7.2

Ffynhonnell: Communications Market Report 2022, OfCom (2022)

Wrth ymchwilio i ddulliau o gyrchu'r data hwn, penderfynom yn erbyn platfformau seiliedig ar fideo a delwedd, fel Instagram a Tik Tok. Mae cyfryngau cymdeithasol seiliedig ar destun, fel Twitter a Facebook, yn cael eu defnyddio'n amlach gan awdurdodau lleol Cymru ac mae eu data yn haws ei ddadansoddi a'i ddehongli.

¹⁰ Daw'r data yn **Tabl 1** o Online Audience Measurement Service Ipsos Iris. Mae ffigurau yn y tabl yn dangos nifer y defnyddwyr 15 oed a throsodd o fewn y cyfnod 1 Medi – 30 Medi 2021. Nid yw data am WhatsApp yn cael ei gynnwys yma oherwydd problemau technegol wrth y ffynhonnell.

Fodd bynnag, yn y pen draw, cyfyngom ni'r ymchwil gychwynnol hon i ddulliau o gyrchu **data Twitter** yn unig, a hynny oherwydd ei fod yn cael ei ddefnyddio'n eang gan gydweithwyr llywodraeth leol ac mae Twitter yn trefnu bod ei ddata ar gael yn gymharol hawdd. Roeddem yn hyderus, felly, y gallem ni gael set ddata a gwneud asesiad cychwynnol o ansawdd a defnyddioldeb. Adolygom ni bedwar dull o gyrchu data Twitter, fel a amlinellir isod.

Ffynhonnell data	Manteision	Anfanteision	Cost
<p>API Twitter</p> <p>Mae Twitter yn blatfform cyfryngau cymdeithasol Americanaidd sy'n gadael i ddefnyddwyr bostio negeseuon byr i fyrddau cyhoeddus. Mae'r postiadau hyn yn cael eu galw'n 'drydarau'. Hefyd mae'n bosibl creu rhwydwaith o gysylltiadau ar Twitter. Mae 'dilyn' cyfrif arall yn dynodi i Twitter eich bod am weld eu trydarau cyhoeddus.</p> <p>Mae defnyddwyr Twitter yn gallu anfon negeseuon at ei gilydd yn breifat hefyd. Mae Twitter yn casglu data am yr holl ryngweithiadau hyn ac yn trefnu bod data a metadata ar gael i ddatblygwyr ac ymchwilwyr cymeradwy drwy ei API.</p> <p>Dechreuom ni'r ymchwil hon drwy ganolbwyntio ar Twitter, gan wybod ei fod yn blatfform poblogaidd ymhlith cynghorau lleol yng Nghymru. Ar ben hynny, mae'r API yn cael ei hysbysebu fel bod yn yn hygyrch yn rhwydd drwy ffurflen gais ar-lein. Rhaid i'r cais hwn wneud yn glir dibenion y cais am ddata a sut maen nhw'n bodloni telerau defnydd Twitter. Yna mae Twitter yn asesu a yw'r nodau hyn yn gydnaws â'i bolisi defnydd.</p>	<p>API Twitter yw'r dull mwyaf uniongyrchol am gyrchu data Twitter. Mae'n pweru llawer o'r dulliau eraill yr ymchwiliwyd iddynt yn yr adran hon.</p>	<p>Anfantais bennaf cyrchu data Twitter trwy ei API ydy cost. Nid yw'r lefelau isaf o fynediad am dâl yn caniatáu am lawer o chwiliadau y mis ac mae cyfyngiadau isel ar lawrlwytho ganddyn nhw. Mae hyn yn golygu bod estyn casgliadau i raddfa ehangach yn gallu mynd yn gymharol gostus, gyda chostau misol sy'n amrywio o ryw \$150 i \$2,500, yn dibynnu ar faint o ddata mae ei angen.</p> <p>Mae rhwystr technegol i defnyddio API Twitter hefyd. Mae angen datblygwr neu ddadansoddwr sydd â sgiliau addas, yn ogystal â gallu o ran rheoli data a chodio.</p>	<p>Tua \$150 i \$2,500 y mis</p>

Ffynhonnell data	Manteision	Anfanteision	Cost
<p>Mae gan API Twitter dri math o fynediad:</p> <ol style="list-style-type: none"> 1 mynediad i drydarau presennol mewn amser real; 2 y gallu i chwilio am unrhyw drydar a wnaed yn y saith diwrnod diwethaf; a 3 gallu llawn i chwilio am unrhyw drydar ers 2006. <p>Yn nodweddiadol mae math 3 dim ond ar gael at ddibenion ymchwil academiaidd.</p>			
<p>Gwasanaethau gwrando cymdeithasol (Synthesio)</p> <p>Gwasanaeth gwrando cymdeithasol sydd ym mherchnogaeth Ipsos Mori yw Synthesio. Mae'r gwasanaeth hwn yn coladu data o amrediad eang o blatfformau cyfryngau cymdeithasol, gan gynnwys Twitter, Facebook, YouTube, a fforymau ar-lein. Mae gan wasanaethau gwrando cymdeithasol fel Synthesio fynediad i lawer iawn mwy o ddata nag y gallai sefydliad unigol ei lawrlwytho'n hawdd trwy API (fel API Twitter), a hynny am eu bod yn cronni data gyda thanysgrifiadau API cost-uchel dros gyfnod hir. Mae hyn yn y bôn yn gadael i'r gwasanaethau gwrando cymdeithasol gaffael bron yr holl ddata a</p>	<p>Mae mynediad premiwm Synthesio i API Twitter yn gadael iddyn nhw roi mynediad i gwsmeriaid i set ddata Twitter arbennig o fawr. Mae data o blatfformau cyfryngau cymdeithasol eraill ar gael hefyd.</p> <p>Mae Synthesio yn darparu rhyngwyneb defnyddwyr, gan osgoi rhai o'r rhwystrau technegol sydd ynghlwm</p>	<p>Mae gwasanaethau gwrando cymdeithasol yn ddrud, gyda thanysgrifiadau blynyddol sy'n costio degoedd o filoedd o bunnoedd. Gallai gwasanaethau gwrando cymdeithasol fod yn opsiwn priodol ar yr amod:</p> <ul style="list-style-type: none"> • bod yna achos gwerth-uchel am eu defnyddio a allai gyfiawnhau'r gost hon; neu 	<p>O £10,000oedd y flwyddyn</p>

Ffynhonnell data	Manteision	Anfanteision	Cost
<p>metadata a gyhoeddir ar y platfformau cyfryngau cymdeithasol.</p> <p>Mae gan Synthesio ymarferoldeb mewnol sy'n gadael i'r defnyddiwr chwilio yn ôl allweddair ac ar draws metadata i gulhau canlyniadau. Yn ogystal, mae'r offeryn yn cynnig swyddogaethau dadansoddi parod, fel dadansoddi testunau a theimladau.</p>	<p>wrth gyrchu a storio data o API Twitter yn uniongyrchol.</p>	<ul style="list-style-type: none"> roedd diddordeb gan nifer o gynghorau mewn rhannu'r gost. 	
<p>NCapture for NVivo</p> <p>Mae NVivo yn offeryn trwyddedig sy'n gallu cael ei ddefnyddio i drefnu a dadansoddi data ansoddol. Mae'n cael ei ddefnyddio'n bennaf i dagio a threfnu data testun. Fodd bynnag, mae gan fersiwn ddiweddaraf yr offeryn estyniad pwrpasol cipio data ar gyfer Google Chrome, o'r enw NCapture for NVivo.</p> <p>Mae'r estyniad porwr hwn yn caniatáu i gynnwys gwefan gael ei gipio a'i fewnforio i gymhwysiad NVivo lle mae'n gallu cael ei ddadansoddi. Mae NCapture for NVivo wedi cael ei optimeiddio'n benodol ar gyfer Twitter. Mae hyn yn golygu y bydd, ar gyfer Twitter yn unig, yn cipio cynnwys a metadata yn uniongyrchol o'r platfform cyfryngau cymdeithasol a throsi'r data i fformat taenlen neu</p>	<p>Mae'n defnyddiol i gasglu symiau bach o ddata Twitter at ddibenion cyfyngedig.</p>	<p>Mae'n ddarostyngedig i'r cyfyngiadau lawrlwytho sylfaenol sy'n cael eu gosod gan Twitter. Mae hyn yn golygu bod nifer gymharol fach o chwiliadau a lawrlwythiadau'n gallu cael eu cyflawni bob dydd. Ar ben hynny, nid yw'r offeryn yn hawdd ei ffurfweddu:</p> <ul style="list-style-type: none"> fel nad yw'r defnyddiwr yn gallu defnyddio metadata i bennu is-set o'r canlyniadau i'w lawrlwytho; 	<p>Amrywiol opsiynau o ran cost. Mae gwybodaeth bellach ar gael ar eu gwefan NVivo software</p>

Ffynhonnell data	Manteision	Anfanteision	Cost
<p>PDF. Yn achos pob gwefan arall, gall yr un broses gael ei defnyddio ond mae data'n cael ei grynhoi'n PDF yn unig.</p> <p>Trwy'r profion rhagarweiniol, roeddem yn gallu cipio data o Twitter gan ddefnyddio'r term chwilio 'covid'. Dychwelodd hyn 17,698 o ganlyniadau, a fewnforiom ni wedyn yn uniongyrchol i gymhwysiad NVivo a'u tagio i'w dadansoddi. Roedd y lawrlwythiad data yn cynnwys metadata ynghylch y defnyddiwr, gan gynnwys eu bywgraffiad¹¹, nifer eu dilynwyr, a'u lleoliad.</p> <p>Roeddem wedyn yn gallu allforio'r data hwn i daenlen Excel, a olygodd fod modd ei ddefnyddio mewn offerynnau di-rif, gan gynnwys meddalwedd ystadegol fel R a Python ac offerynnau lledaenu fel PowerBI.</p>		<ul style="list-style-type: none"> • nid yw nifer y trydarau a ddychwelir bob amser yn gyson¹²; • nid yw'r cyfnod dan sylw bob amser yn gyson – fel arfer y cynnig yw'r 7 diwrnod diwethaf¹³; ac • nid yw'r defnyddiwr yn gallu pennu dyddiadau i chwilio am drydarau rhyngddynt. 	

¹¹ Mae bywgraffiad Twitter yn grynodeb sydd ar gael yn gyhoeddus lle mae defnyddwyr yn ysgrifennu amdany'n nhw eu hunain.

¹² <https://support.alfasoftware.com/hc/en-us/articles/360001218318-How-many-tweets-can-you-capture-for-NVivo-using-NCapture->

¹³ Mae swm y data sy'n cael ei gipio fel set ddata gan NCapture yn cael ei gyfyngu gan ei ddefnydd o API Twitter. Fel arfer mae trydarau dim ond yn cael eu darparu o'r 7 diwrnod diwethaf. Mae'n bosibl na fydd tydarau hyn yn cael eu cipio oherwydd cyfyngiadau API Twitter, hyd yn oed os ydynt yn weladwy ar wefan Twitter.

Ffynhonnell data	Manteision	Anfanteision	Cost
<p>COSMOS a Phartneriaeth Academaidd</p> <p>Ar sail ein hymchwil gychwynnol, gwnaethom fireinio'n chwiliad, gan nodi bod arnom ni eisiau offeryn a oedd yn ddi-dâl neu'n rhad i'w gyrchu ac a oedd yn gadael i ni ffurfweddu ein chwiliadau. Arweiniodd hyn ni at COSMOS.</p> <p>Mae COSMOS yn cael ei ddatblygu gan y Labordy Gwyddor Data ym Mhrifysgol Caerdydd a'i ariannu gan y Cyngor Ymchwil Economaidd a Chymdeithasol. Datblygan nhw feddalwedd COSMOS i ganiatáu mynediad rhadd a di-dâl i ddata cyfryngau cymdeithasol ar gyfer ymchwilwyr a sefydliadau nid-er-elw cymeradwy.¹⁴</p> <p>Mae meddalwedd COSMOS yn gadael i ddefnyddwyr gyrchu data Twitter drwy offeryn pen blaen. Mae hyn yn cyfuno casglu data, gan gynnwys y gallu i ffurfweddu chwiliad hyd lefel sylfaenol, â rhywfaint o allu dadansoddol. Edrychir ar hyn yn fanwl yn hwyrach yn yr adroddiad. Mae modd allforio'r holl ddata sy'n cael ei gasglu trwy COSMOS yn hawdd trwy daenlen a'i ddefnyddio</p>	<p>Yn darparu rhyngwyneb defnyddiwr, gan osgoi rhai o'r rhwystrau technegol sydd ynghlwm wrth gaffael a storio data o API Twitter yn uniongyrchol.</p> <p>Yn cynnig lefelau tebyg o fynediad â defnyddio API Twitter yn uniongyrchol ond heb unrhyw gost.</p>	<p>Nid yw'n rhoi rheolaeth lawn i ddefnyddwyr dros ei fynediad i API Twitter.</p> <p>Mae casgliadau'n cael eu cyfyngu i nifer benodedig o drydarau yr eiliad. Mae hyn yn codi cwestiynau ystadegol am ba drydarau sy'n cael eu hepgor a bydd efallai'n gwneud COSMOS yn anaddas am rai mathau o ymchwil. Mae hyn oherwydd lefel y mynediad i API Twitter y mae tîm COSMOS wedi'i drefnu gyda Twitter. Yn wahanol i ddefnyddio API Twitter neu wasanaeth gwrando cymdeithasol, nid oes modd i ni dalu mwy er mwyn codi'r terfyn hwn.</p>	<p>Yn ddi-dâl am sefydliadau academaidd ac nid-er-elw</p>

¹⁴ COSMOS is licensed to academic or not-for-profit organisations. Each organisation wishing to use the tool should consult its terms and conditions first.

Ffynhonnell data	Manteision	Anfanteision	Cost
mewn offerynnau eraill fel R, Python neu PowerBI i'w ddadansoddi ymhellach neu ei ledaenu. Bydd yn rhaid i sefydliadau sydd am ddefnyddio'r offeryn adolygu'r telerau ac amodau ac ymgysylltu â Phrifysgol Caerdydd i bennu eu cymhwyster.			

O ganlyniad i'r dadansoddiad uchod, dewisom brofi'r feddalwedd COSMOS yn y peilot hwn gan ei bod yn bodloni ein hanghenion pennaf sef bod yn hygyrch, yn ddi-dâl i'w defnyddio, ac yn cynnig rhywfaint o allu dadansoddol mewnol. Hefyd mae'n dileu'r baich technegol sydd ynghlwm wrth gyrchu a storio data o API Twitter.

Canfyddiadau cyffredinol

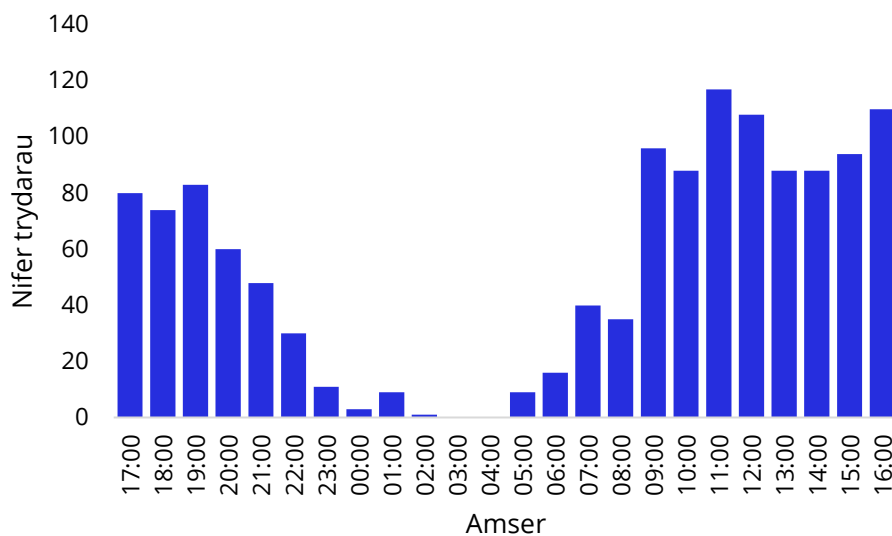
Mae casgliad data COSMOS yn casglu data am set o drydarau gan gynnwys:

- testun trydar ac ID trydar;
- enw defnyddiwr ac ID defnyddiwr awdur trydar;
- amser a dyddiad ei drydar;
- a oedd yn "aildrydar"; a
- metadata am iaith a lleoliad daearyddol y trydar.

Mae COSMOS hefyd yn cynnwys algorithmau sy'n amcangyfrif teimlad testun trydar (cadarnhaol neu negyddol) ac enw a rhyw'r awdur. Fodd bynnag, nid oedd angen y data hwn arnom, ac ni wnaethom ofyn amdano na'i ddefnyddio.

Dengys **Ffigur 1** nifer y trydarau a gasglwyd ym mhob awr casgliad prawf COSMOS¹⁵. Dechreuodd y prawf ychydig cyn 17:00 a pharhaodd am 24 awr. Yn unol â'r disgwyl, mae llai o drydarau yn oriau mân y bore, gyda brig tua chanol y dydd.

Ffigur 1: Nifer trydarau a gasglwyd yr awr trwy gydol y cyfnod prawf 24-awr

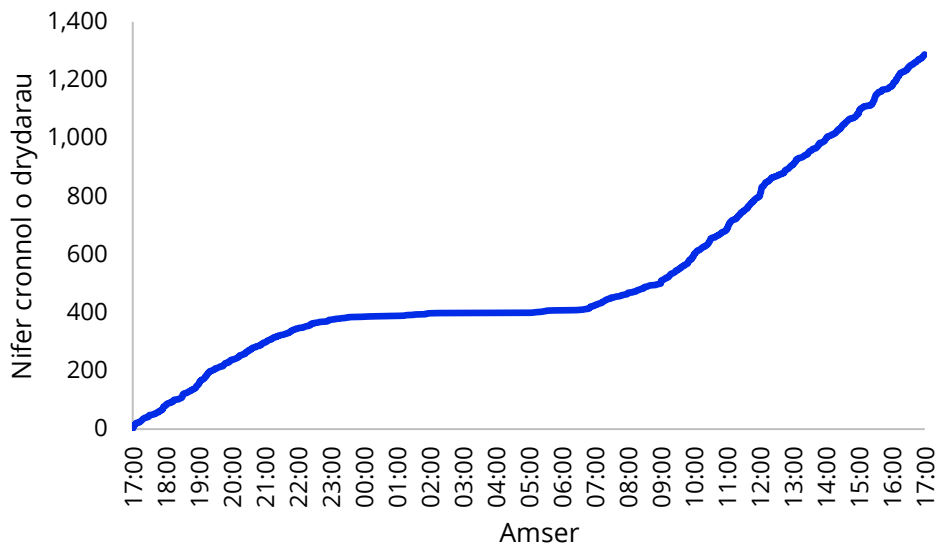


Ffynhonnell: Data Cymru gan ddefnyddio data a gasglwyd drwy COSMOS

Dengys **Ffigur 2** gyfanswm y trydarau a gasglwyd ers dechrau casglu'r data. O fewn y cyfnod prawf 24-awr, cafodd mwy na 1,200 o drydarau eu casglu.

¹⁵ Casglwyd y data yma yn ystod Prawf 2a, sy'n cael ei ddisgrifio yn yr adran 'Profion penodol a gynhaliom ar COSMOS'.

Ffigur 2: Nifer cronol trydarau a gasglwyd trwy gydol y cyfnod prawf 24-awrd



Ffynhonnell: Data Cymru gan ddefnyddio data a gasglwyd drwy COSMOS

Dengys **Ffigur 1** a **Ffigur 2** fod COSMOS yn casglu trydarau mewn amser real. Dechreuodd y casgliad COSMOS hwn ychydig cyn 17:00. Ni chafodd unrhyw drydarau a gasglwyd eu creu cyn 17:00.

Gall COSMOS greu cymylau geiriau, sy'n gynrychiolaeth weledol o'r geiriau sy'n cael eu defnyddio yn y trydar. Maen nhw'n cael eu defnyddio i amlygu geiriau poblogaidd ar sail amlder a pherthnasedd. Dengys **Ffigur 3** enghraifft o ddelweddiad cwmwl geiriau o'n profion COSMOS.

Ffigur 3: Cwmwl geiriau o eiriau poblogaidd sy'n cael eu defnyddio mewn trydarau sy'n ymgysylltu â chyfrifon Twitter llywodraeth leol (Saesneg yn unig)

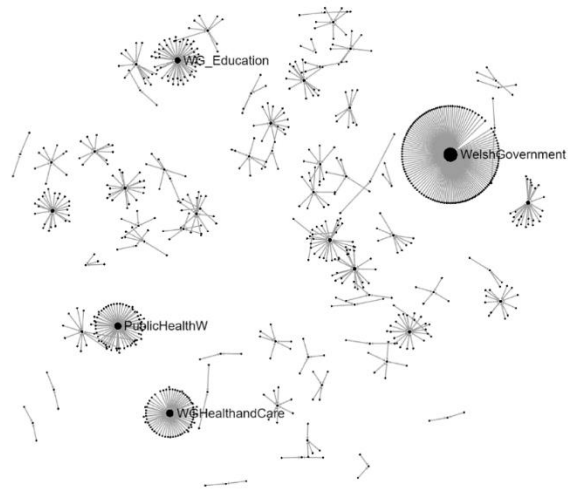


Ffynhonnell: COSMOS

Mae COSMOS yn gallu cynnig dadansoddiad rhwydwaith syml hefyd o drydarau a sydd wedi'u casglu. Dengys **Ffigur 4** enghraifft o'r pwyntiau (defnyddwyr Twitter â'u henwau wedi'u tynnu) sy'n cael eu cysylltu gan linellau (aildrydarau). Mae aildrydar yn cael ei ddefnyddio'n gyffredin i ddod â thrydar i sylw'ch dilynwyr. Er eu bod yn brin am gasgliad

data byr 24-awr, mae'r ddelwedd hon yn dangos y potensial i gael cipolwg ar bwy sy'n ymgysylltu â llywodraeth leol a pha gysylltiadau sy'n bodoli.

Ffigur 4: Diagram rhwydwaith o aildrydaru cynnwys gan gyfrifon



Ffynhonnell: COSMOS

Profion penodol a gynhaliom ar COSMOS

Wrth weld bod gan COSMOS y potensial i gynnig mewnwelediadau gwerthfawr, penderfynom ni ymchwilio i'r offeryn yn fanylach er mwyn deall:

1. Pa olwg syd ar gasgliad data COSMOS;
2. Pa fewnwelediad a gwerth gallai'r data hwn eu creu ar gyfer ein partneriaid; a
3. Phryd y gallai offerynnau data cyfryngau cymdeithasol eraill fod yn fwy priodol na COSMOS.

Er mwyn sicrhau ein bod yn deall perfformiad COSMOS, cynhaliom ni **bum** casgliad prawf. Roedd y rhain wedi'u dylunio i bennu a allai ddarparu gwybodaeth am y daearyddiaethau, y cyfnodau, a'r lefelau manylion roedd eu hangen arnom ni.

Beth oeddem ni am ddarganfod?	Beth wnaethom ni?	Canfyddiadau allweddol
<p>Prawf 1: Oes modd i ni gasglu'r holl drydarau o Gymru?</p>	<p>Defnyddiom ni opsiynau hidlo daearyddol COSMOS i ddewis dim ond trydarau â chanddynt "geotag" yn deillio o Gymru¹⁶.</p>	<p>Anfanteision:</p> <ul style="list-style-type: none"> • Roedd y prawf hwn yn aflwyddiannus. Y brif broblem yw bod geotagio trydarau yn nodwedd ddewisol, a lleiafrif bach o ddefnyddwyr Twitter sy'n optio i mewn i ganiatáu i'w trydarau ddwyn geotag. • Hefyd mae defnyddwyr COSMOS yn cael eu cyfyngu o ran y modd gallan nhw gyfuno hidlo daearyddol â hidlyddion defnyddiol eraill, fel allweddair a hidlyddion ar sail cyfrifon. <p>Er enghraifft, bydd hidlo yn ôl allweddair ac ardal ddaearyddol yn dychwelyd yr holl drydarau am y hidlydd allweddair ar holl drydarau o fewn yr ardal ddaearyddol. Nid yw defnyddiwr yn gallu hidlo am ddim ond y trydarau a oedd yn cynnwys yr allweddair yn ogystal â bod o fewn yr ardal ddaearyddol.</p>
<p>Prawf 2: Oes modd i ni gasglu dim ond trydarau sy'n ymgysylltu â sefydliadau llywodraeth (leol) Cymru?</p>	<p>Ar gyfer y prawf hwn, dewisom ni restr o 46 cyfrif Twitter a oedd yn berthnasol i lywodraeth leol yng Nghymru.</p> <p>Yna profom ni allu COSMOS i gasglu trydarau a oedd yn</p>	<p>Manteision:</p> <ul style="list-style-type: none"> • Cynhyrchodd y prawf hwn y set ddata ddisgwyliedig; cynhyrchodd Prawf 2a 1,351o drydarau a chynhyrchodd Prawf 2b 1,701 o drydarau yn y cyfnod prawf 24 awr. Mae hyn yn dangos bod casglu trydarau yn effeithiol o'u hidlo yn ôl cyfrifon neu achosion o grybwyll cyfrifon.

¹⁶ Gall defnyddwyr Twitter ychwanegu data daearyddol at eu trydariadau. Mae'r nodwedd hon yn ddewisol ac yn analluogi yn ddiodyn.

Beth oeddem ni am ddarganfod?	Beth wnaethom ni?	Canfyddiadau allweddol
	<p>gysylltiedig â'r rhestr ddethol yma yn unig.</p> <p>Roedd 2 ran i Brawf 2, sef:</p> <p>Prawf 2a a geisiodd gasglu'r holl drydarau a <u>gyhoeddwyd</u> gan gyfrif yn y rhestr ddethol.</p> <p>Prawf 2b a geisiodd gasglu'r holl drydarau a <u>soniodd</u> am gyfrifon yn y rhestr ddethol neu yn eu <u>haildrydar</u>¹⁷.</p>	<ul style="list-style-type: none"> Gallai casgliadau o'r math hwn fod yn ddefnyddiol wrth gasglu gwybodaeth am ymgysylltiad y cyhoedd â sefydliadau llywodraeth leol. Mae cryn amrywiaeth yn nhôn a diben trydarau yn y set ddata hon, sy'n amrywio o gefnogaeth i brosiectau neu wasanaethau lleol, canmol neu feirniadu polisi lleol a Llywodraeth Cymru, rhwydweithio busnes a sbam. Rhaid i'r safbwyntiau amrywiol hyn gael eu hystyried mewn unrhyw ddefnydd data cyfryngau cymdeithasol yn y dyfodol. Profom ni hefyd a oedd COSMOS yn casglu pob trydar a oedd ar gael neu sampl ohonyn nhw. Ar ôl y broses gasglu, defnyddiom ni Twitter â llaw i gasglu trydarau gyda'r un meini prawf chwilio. Casglodd y casgliad â llaw hwn un union yr un set o drydarau â set ddata COSMOS, gan gadarnhau bod COSMOS wedi casglu pob trydar a oedd ar gael yn yr achos hwn. <p>Anfanteision:</p> <ul style="list-style-type: none"> Mae testun aildrydarau yn cael ei newid yn ystod casgliad data COSMOS. Yn nodweddiadol mae ond yn cynnwys testun y trydar gwreiddiol, gan gwtogi peth neu'r cyfan o'r testun a ychwanegwyd gan y defnyddiwr a'i aildrydarodd. Dylai hyn

¹⁷ A 'retweet' is a feature of Twitter where one user effectively reposts another user's tweet.

Beth oeddem ni am ddarganfod?	Beth wnaethom ni?	Canfyddiadau allweddol
		<p>gael ei gymryd i ystyriaeth wrth gynllunio unrhyw dddansoddiad.</p> <ul style="list-style-type: none"> • Derbyniod y mwyafrif o'r cyfrifon Twitter a oedd o fewn ein hidlydd gymharol ychydig o drydarau. Mae hyn yn awgrymu y bydd angen casgliad data parhaus hirdymor i gasglu set ddata a oedd yn ddigon mawr i fod yn ystadegol ddefnyddiol.
<p>Prawf 3: "A gawn ni gasglu'r holl drydarau sy'n cynnwys allweddeiriau sy'n ymwneud â phwnc llywodraeth leol Cymru?"</p>	<p>Profom ni effeithiolrwydd hidlo allweddeiriau COSMOS drwy gasglu trydarau a ddefnyddiodd unrhyw rai o restr o 77 allweddair sy'n berthnasol i awdurdodau lleol Cymru, ond nid i hynny'n unig.</p> <p>Roedd yr allweddeiriau hyn yn cynnwys pethau fel ACEs (Adverse Childhood Experiences), Dementia, Deddf Cenedlaethau'r Dyfodol, ac yn y blaen.</p> <p>Gallai hyn o bosibl gynnig amrediad ehangach o safbwyntiau ac, yn holl-bwysig, gasglu barn pobl tra nad ydyn</p>	<p>Manteision:</p> <ul style="list-style-type: none"> • Roedd y prawf yn llwyddiannus wrth gasglu 38,875 o drydarau a oedd yn cynnwys un neu fwy o'r allweddeiriau a ddewisom yn y cyfnod prawf 24-awr. <p>Anfanteision:</p> <ul style="list-style-type: none"> • Er i'r allweddeiriau gael eu dewis i fod yn berthnasol i'r sector cyhoeddus, nid oedd sawl un o'r trydarau a gasglwyd. Er enghraifft, roedd y mwyafrif o'r canlyniadau a oedd yn cynnwys yr allweddair 'ACEs' yn ymwneud â chanwyr pop Ffilipinaidd o'r enw BGYO o ganlyniad i'w hymgyrch gyhoeddusrwydd 'Aces and Bloom' 2021. • Mewn casgliad data cyfryngau cymdeithasol, mae'n hanfodol bod: <ul style="list-style-type: none"> ○ yr allweddeiriau'n cael eu dewis yn ofalus i leihau casglu data amherthnasol; neu

Beth oeddem ni am ddarganfod?	Beth wnaethom ni?	Canfyddiadau allweddol
	<p>nhw'n ymgysylltu'n uniongyrchol â chyfrifon llywodraeth leol.</p> <p>Nid oeddd unrhyw gyfyngiad daearyddol i'r prawf hwn, felly roedd yn cynnwys trydarau o lawer o wledydd.</p>	<ul style="list-style-type: none"> ○ bod canlyniadau amherthnasol yn cael eu hidlo allan ar ôl i'r data gael ei gasglu¹⁸. • Yn yr un modd â Phrawf 2, roeddem am brofi hefyd a oedd COSMOS yn casglu pob trydar a oedd ar gael a oedd yn gysylltiedig â'r term chwilio. Ar ôl y broses gasglu, defnyddiom ni Twitter â llaw i gasglu is-set o'r trydarau gan ddefnyddio un o'n hallweddeiriau dethol. Yn wahanol i Brawf 2, ni chipiodd COSMOS bob un o'r trydarau a oedd ar gael. Awgrymodd ymchwil bellach fod nodwedd fwriadol API Twitter yn gyfrifol am hyn. Yn dibynnu ar yr haen dalu mae defnyddiwr API Twitter arni, mae terfyn yn bodoli ar nifer y trydarau sy'n gallu cael eu casglu yr eiliad cyn bod cap ar y casgliad data hwnnw. Tu hwnt i'r pwynt hwn, nid yw pob trydar sy'n cyd-fynd â'r meini prawf chwilio yn cael ei ddychwelyd; Mae hyn yn cael ei alw'n 'cyfyngu cyfradd gasglu'. Gan hynny, bydd COSMOS – ac unrhyw offeryn arall sy'n defnyddio API Twitter – dim ond yn casglu sampl o'r holl drydarau posibl os yw'r hidlyddion ar y casgliad hwnnw'n cipio trydarau'n rhy gyflym.

¹⁸ Mae dulliau o'r fath yn cynnwys dadansoddi metadata lleoliad ac iaith, chwilio am restrï o allweddeiriau neu enwau cyfrif (fel ym Mhrawf 2) a dulliau ystadegol sy'n pennu pwnc trydar. Mae cryn gymhlethdod ynghlwm wrth y dulliau hyn o ran dadansoddi data a diogelu data.

Beth oeddem ni am ddarganfod?	Beth wnaethom ni?	Canfyddiadau allweddol
<p>Prawf 4: Oes modd i ni gasglu'r holl drydarau Cymraeg?</p>	<p>Ymchwiliom i a oedd hi'n bosibl caffael trydarau ar sail iaith eu cyhoeddi yn unig. Yn benodol, ceisiodd gasglu trydarau yr oedd ganddyn nhw'r tag metadata Cymraeg.</p> <p>Nod y prawf hwn oedd asesu gallu'r hidlydd iaith i ddychwelyd set o drydarau Cymraeg.</p>	<p>Anfanteision:</p> <ul style="list-style-type: none"> Nid yw'n bosibl defnyddio COSMOS i gasglu trydarau ar sail iaith yn unig; mae'n rhaid defnyddio hidlo iaith ar y cyd â hidlydd allweddair neu gyfrif.
<p>Prawf 5: A oes modd atgynhyrchu'r profion hyn?</p>	<p>Roedd y prawf hwn yn cynnwys dau ailadroddiadau byr (dwy awr) o Brawf 2a. Hynny yw, cafodd trydarau eu casglu a bostiwyd gan un o'r rhestr ddethol o gyfrifon.</p> <p>Y bwriad oedd sefydlu a allai trydarau gael eu casglu ar ôl y ffaith neu eu casglu mewn 'amser real' yn unig, a fyddai felly'n gwneud casgliad yn anodd ei ailadrodd.</p>	<p>Manteision:</p> <ul style="list-style-type: none"> Cynhaliodd y ddau brawf ar yr un diwrnod ac roedd y data a gynhyrchwyd gan bob casgliad yn unigryw. Mae hyn yn awgrymu bod API Twitter yn gweithio mewn 'amser real' ac nad yw'n tynnu data o "gronfa ddyddiol o drydarau". Yng ngoleuni casglu trydarau mewn 'amser real', mae dylunio casgliad data gan ddefnyddio COSMOS yn gymharol syml; rhaid i gasgliadau gael eu cynllunio ymlaen llaw a rhaid i COSMOS redeg yn barhaus. <p>Anfanteision:</p> <ul style="list-style-type: none"> Cadarnhaodd y prawf fod trydarau ond yn cael eu casglu am y cyfnod mae casgliad COSMOS wrthi ar waith. Ni chafodd unrhyw drydarau eu casglu tu allan i'r priod gyfnodau casglu yn y naill brawf ailadrodd na'r llall. Mae hyn yn golygu bod

Beth oeddem ni am ddarganfod?	Beth wnaethom ni?	Canfyddiadau allweddol
		<p>COSMOS dim ond yn gallu cipio data Twitter mewn amser real ac na all gael ei ddefnyddio i gasglu data hanesyddol.</p> <p>Mae casglu data mewn 'amser real' yn fantais ac yn anfantais, yn dibynnu ar y defnydd a fwriedir am y data.</p>

Casgliad

Nod yr ymchwil hon oedd deall a oedd data ar gael, yn ddefnyddiadwy ac o gymorth i'r sawl sy'n gwneud penderfyniadau mewn llywodraeth leol yng Nghymru. Dechreuom ein hymchwiliad drwy adolygu **pedwar** opsiwn am gyrchu data cyfryngau cymdeithasol, sef:

- Opsiwn 1 – API Twitter;
- Opsiwn 2 – Sythesio, gwasanaeth gwrando cymdeithasol masnachol;
- Opsiwn 3 – NCapture, ychwanegyn i offeryn dadansoddi data sy'n bodoli'n barod; ac
- Opsiwn 4 – COSMOS, offeryn a gynhyrchwyd yn academiaidd i gyrchu API Twitter.

Yn dilyn yr adolygiad hwn, dewisom ni COSMOS fel yr opsiwn byddem yn ymchwilio iddo ymhellach. Mae COSMOS yn cynnig mynediad rhwydd i gasgliadau bach o **ddata Twitter yn unig**. Nid yw'n **gofyn am unrhyw sgil technegol** i'w weithredu ac, i sefydliadau academiaidd ac nid-er-elw, mae'n **ddi-dâl i'w ddefnyddio**.

Mae casgliad data COSMOS yn cipio trydarau mewn **amser real**; nid yw casgliadau COSMOS yn gallu cyrchu trydarau hanesyddol ac nid yw casgliadau'n gallu cael eu hatgynhyrchu ar ôl y ffaith. Rhaid i gasgliad data COSMOS gael ei adael ar waith am y cyfnod gofynnol.

Mae testun trydar yn cael ei storio gydag **amrediad o fetadata** gan gynnwys amser casglu, enw defnyddiwr Twitter, iaith, geodata ac ai aildryder oedd y trydar. Mae COSMOS hefyd yn amcangyfrif enw go iawn a rhywedd awdur y trydar yn ogystal ag a oes gan y trydar deimlad cadarnhaol neu negyddol. **Mae ansawdd y data yn gyffredinol yn uchel**.

Ar gyfer casgliadau bach, dengys ein profion fod yr offeryn yn cipio pob trydar sy'n cael ei bostio. Ond pan fydd casgliad yn cwmpasu nifer fawr o drydarau, nid yw COSMOS yn cipio pob trydar sydd ar gael, a hynny am fod Twitter yn gosod cap ar y raddfa mae data'n gallu cael ei lawrlwytho drwy gyfrwng ei API.

Os y nod yw cynnal casgliadau mawr, neu gasgliadau ar ddata cyfryngau cymdeithasol hanesyddol (h.y. nid byw) yna byddem ni'n awgrymu defnyddio porthiannau data cwmniau cyfryngau cymdeithasol eu hun neu ymgysylltu â sefydliadau academiaidd neu sefydliadau gwrando cymdeithasol masnachol. Fodd bynnag, mae cost ariannol ynghlwm wrth yr opsiynau hyn.

Pa blatfform bynnag sy'n cael ei ddefnyddio, **nid yw'n bosibl casglu'r holl drydarau mewn ardal ddaearyddol benodol** mewn ffordd bwrpasol. Er enghraifft, nid oes ffordd hawdd o ddefnyddio COSMOS, API Twitter, na gwasanaeth gwrando cymdeithasol i gasglu "pob trydar sy'n deillio o Gymru" neu "pob trydar sy'n crybwyll llywodraeth leol yng Nghymru", a hynny am nad yw trydarau, yn ddi-ofyn, yn cael eu

geotagio gan yr awdur. Gan hynny, os na fydd defnyddiwr yn dweud o ble mae'n postio, nid oes gan Twitter unrhyw ffordd o wybod o ba ardal y deilliodd trydar.

Er gwaetha'r cyfyngiadau hyn, mae **COSMOS yn ffynhonnell data briodol** am fonitro'r ymgysylltiad â Twitter o amgylch nifer fach o allweddeiriau neu enwau cyfrif. Byddai'r dull hwn yn ddefnyddiol ar gyfer tasgau fel monitro ymgysylltiad â chyfrifon Twitter llywodraeth leol yng Nghymru. Er enghraifft, gellid cael dealltwriaeth bellach drwy gynnal gwaith i nodi trydarau sy'n cyfeirio at bynciau, meysydd gwasanaeth neu'n sôn am ardaloedd daearyddol penodol. O fewn pob un o'r rhain, byddai modd cael cipolwg ar nifer y trydarau, pa mor gyffredin mae pynciau penodol yn cael eu crybwyll ac a yw pynciau'n gysylltiedig fel arfer â theimladau cadarnhaol neu negyddol. Drwy gynnal casgliad parhaus COSMOS, gallai data gael ei gymharu am fisoedd neu flynyddoedd gwahanol. Gallai hyn gael ei ddefnyddio i ddeall newid ym mlaenoriaethau'r cyhoedd ac o bosibl yn waelodlin am gofnodi tystiolaeth o ymateb y cyhoedd i newidiadau mewn gwasanaethau neu bolisiau.

Os bydd achos a llywodraethu priodol gan awdurdod lleol am ei ddefnyddio, mae COSMOS yn offeryn grymus ar gyfer deall sut mae llywodraeth leol yn cysylltu â dinasyddion yng Nghymru, a'u barn nhw amdani.

Adnoddau pellach

- [‘Using Social Media for Social Research: An Introduction’](#), Ymchwil Gymdeithasol y Llywodraeth (2016) (Saesneg yn unig).
- [‘Guidelines for the Use of Social Media Data in Research’](#), Prifysgol Caerfrog (Saesneg yn unig).
- [‘Using Twitter as a data source: an overview of social media research tools’](#), Ysgol Economeg Llundain (2019) (Saesneg yn unig).
- [‘Social Media Research: A Guide to Ethics’](#), Prifysgol Aberdeen (2016) (Saesneg yn unig).

Cysylltwch â ni am fwy o wybodaeth:

 www.data.cymru

 ymholiadau@data.cymru

 029 2090 9500

 DataCymru2018

 Data
Cymru