Polisi tynnu data oddi ar y we

Ein polisi ar gyfer tynnu data oddi ar y we ar food.gov

Cefndir

Mae defnyddio ffynonellau data amgen yn hanfodol er mwyn i’r Asiantaeth Safonau Bwyd (ASB) gyflawni ei nod, sef diogelu iechyd y cyhoedd rhag risgiau a allai godi mewn perthynas â bwyta bwyd (gan gynnwys y risgiau sy’n deillio o’r ffordd y caiff ei gynhyrchu neu ei gyflenwi) a diogelu buddiannau defnyddwyr mewn perthynas â bwyd. Tynnu data oddi ar y we (web scraping) yw un o'r ffynonellau data amgen hyn a ddefnyddir yn aml i ategu'r ffynonellau data presennol. Felly, pwrpas y polisi hwn yw sicrhau bod gweithgareddau tynnu data oddi ar y we yn cael eu cynnal yn dryloyw, yn foesegol ac yn unol â'r holl ddeddfwriaeth berthnasol.

Mae'r polisi hwn yn nodi'r arferion a'r gweithdrefnau y bydd yr ASB yn eu dilyn wrth dynnu data oddi ar y we, fel arfer o HTML. Diffinnir tynnu data oddi ar y we, neu ‘web scraping’, fel casglu data yn awtomatig o'r we gan ddefnyddio darn o feddalwedd neu raglen.

Cwmpas

Mae'r polisi hwn yn nodi'r arferion a'r gweithdrefnau y mae'n rhaid i weithwyr yr ASB, staff dros dro a staff achlysurol, ac unrhyw gontractiwr, ymchwilydd neu bartner academaidd sy'n gwneud gwaith i'r Asiantaeth neu'n gweithredu ar ein rhan, eu dilyn wrth dynnu data oddi ar y we.

Wrth asesu sefydliadau allanol a allai dynnu data oddi ar wefannau ar ran yr ASB, naill ai trwy weithgareddau busnes fel arfer neu ar sail ad hoc, bydd y polisi hwn yn cael ei ddefnyddio fel canllaw i lywio gweithgareddau sicrwydd. Mae hyn er mwyn sicrhau bod strwythurau llywodraethu addas ar waith wrth gasglu a phrosesu data a gesglir trwy weithgareddau tynnu data.

Sylwch nad yw'r ddogfen hon yn cynnwys defnyddio Rhyngwynebau Rhaglennu Cymwysiadau (APIs) ac mae'n berthnasol i'r ASB yn unig.

Adolygiad Nesaf

Ein nod yw sicrhau bod y canllawiau yn gyfredol a chynnal adolygiadau rheolaidd i sicrhau bod y canllawiau hyn yn parhau i fod yn berthnasol. Y dyddiad adolygu nesaf sydd wedi’i drefnu ar gyfer y canllawiau hyn yw 15 Ebrill 2022.

Cydymffurfiaeth

Rhaid i holl staff yr ASB gydymffurfio â'r polisi hwn mewn unrhyw brosiect sy'n cynnwys tynnu data oddi ar y we. Gall methu â chydymffurfio arwain at gamau disgyblu yn unol â pholisi disgyblaeth y sefydliad.

Datganiad polisi

Mae'r ASB yn cynnal gweithgarwch tynnu data oddi ar y we at ddibenion ei swyddogaethau a nodir yn Neddf Safonau Bwyd 1999.

Dyma sail gyfreithiol yr ASB ar gyfer gweithgarwch tynnu data o'r fath:

Mae adrannau 8, 10, 11 a 21(1) Deddf Safonau Bwyd 1999 yn rhoi pwerau i'r ASB i gael a chasglu gwybodaeth, trwy gynnal arsylwadau, am faterion sy'n ymwneud â diogelwch bwyd neu fuddiannau defnyddwyr mewn perthynas â bwyd, gan gynnwys, ond heb fod yn gyfyngedig i, ffynonellau bwyd, safleoedd bwyd, busnesau bwyd neu weithrediadau masnachol sy'n cael eu cynnal mewn perthynas â bwyd.

Wrth gyflawni'r gweithgarwch tynnu data oddi ar y we a ddisgrifir yma, mae'r ASB yn arfer ei phwerau statudol ac yn gweithredu o fewn ei hawdurdod statudol. Mae adran 19 Deddf Safonau Bwyd 1999 hefyd yn caniatáu i'r ASB rannu gwybodaeth ble fyddai gwneud hynny er budd y cyhoedd.

I'r graddau bod yr wybodaeth sy'n cael ei thynnu yn nodi unigolion byw, mae'r gweithgarwch tynnu data oddi ar y we yn cynnwys prosesu rhywfaint o ddata personol. Bydd yr ASB yn prosesu data o'r fath yn unol â darpariaethau Rheoliad Cyffredinol ar Ddiogelu Data (GDPR) yr Undeb Ewropeaidd (UE) 2016/679 a Deddf Diogelu Data'r Deyrnas Unedig (DU) 2018 (DPA).

Mae'r ASB yn credu bod y prosesu yn angenrheidiol o fewn swyddogaeth tasg gyhoeddus yr ASB (Erthygl 6.1(e) GDPR) ac, i'r graddau y mae data troseddol yn berthnasol, yn angenrheidiol i'r ASB fel awdurdod cymwys at ddibenion gorfodi'r gyfraith er budd y cyhoedd a/neu ar gyfer honiadau cyfreithiol (GDPR Erthygl 9.2(f) a (g) ac Atodlen 2 Rhan 1 y DPA ac Atodlen 2 Rhan 1, adran 5 ac Atodlen 7) neu mae'n debygol ei fod wedi'i roi yn y parth cyhoeddus gan yr unigolion dan sylw (Erthygl GDPR 9.2(e)).

Bydd yr ASB yn mabwysiadu'r egwyddorion canlynol i arwain ein gweithgareddau tynnu data oddi ar y we:

- Ceisio lleihau'r baich ar gwefannau lle mae'r data'n cael ei dynnu
- Cadw at yr holl ddeddfwriaeth berthnasol a monitro'r sefyllfa gyfreithiol a moesegol sy'n esblygu
- Diogelu'r holl ddata personol

Manylion Polisi

Ceisio lleihau'r baich ar y tudalennau gwe lle mae'r data'n cael ei dynnu

Gall defnydd gormodol o dynnu data oddi ar y we fod yn faich, a gall gorddefnyddio ar un dudalen we gael effaith niweidiol ar weithrediad y wefan. Er mwyn osgoi sefyllfaoedd o'r fath, byddwn ni'n dilyn yr arferion hyn, lle bo hynny'n berthnasol, a gweithredu o fewn terfynau ein swyddogaethau statudol bob amser:

- Asesu ddulliau casglu data eraill, fel API's, cyn ystyried tynnu data oddi ar y we
- Parchu cyfyngiadau tynnu data a orfodir trwy ddogfennau fel y Protocol Eithrio Robotiaid (Robots.txt) a Thelerau ac Amodau'r Safle os a phan fo hynny'n bosibl
- Oedi tynnu data oddi ar dudalennau ar yr un parth
- Cynnwys amser segur rhwng ceisiadau
- Cyfyngu dyfnder y chwilio (crawl) o fewn y parth penodol
- Wrth dynnu data oddi ar sawl parth, gweithredu chwilio cyfochrog (parallelised crawling) i leihau ceisiadau olynol i'r un parth
- Tynnu data ar adeg o'r dydd pan nad yw'r wefan yn debygol o fod yn profi traffig trwm – er enghraifft, yn gynnar yn y bore neu gyda'r nos
- Optimeiddio'r strategaeth tynnu data oddi ar y we er mwyn lleihau nifer y ceisiadau i barthau
- Casglu rhannau o dudalennau sy'n ofynnol at y diben cychwynnol yn unig

- Cynnwys Llinyn Asiant Defnyddiwr i wahaniaethu’r ASB oddi wrth ddefnyddwyr eraill.

Nid yw'r rhestr hon o arferion gorau yn gynhwysfawr a bydd yr ASB yn ymdrechu i gynnal safon uchel wrth dynnu data oddi ar y we trwy fonitro a gweithredu arferion gorau.

Cadw at ddeddfwriaeth berthnasol a monitro'r sefyllfa gyfreithiol a moesegol sy'n esblygu

Mae'r ASB yn deall bod y sefyllfa gyfreithiol sy'n ymwneud â thynnu data oddi ar y we yn gymhleth ac yn esblygu'n barhaus; a chymharol ychydig o gynseiliau cyfreithiol perthnasol sy’n bodoli ar gyfer cynnal gweithgareddau o'r fath. Bydd yr ASB yn parhau i fonitro'r sefyllfa gyfreithiol wrth iddi esblygu ac yn diwygio ein dull yn unol â hynny. Yn yr un modd, mae'r ASB yn cydnabod y gall tynnu data oddi ar y we a phrosesu data a gesglir yn y modd hwn arwain at amrywiaeth o faterion a heriau moesegol. Mae ymdrechion yn mynd rhagddynt i ddatblygu fframwaith ymarferol ar gyfer ystyried a lliniaru risgiau moesegol posibl a briodolir yn uniongyrchol ac yn anuniongyrchol i dynnu data oddi ar y we.

Rhaid i'r rhai sy'n dymuno tynnu data oddi ar y we gofnodi a chyfiawnhau'r rhesymeg, y rhesymu cyfreithiol a moesegol a buddion tynnu’r data. Defnyddir y rhesymeg hon fel sail i geisio cymeradwyaeth gan y Bwrdd Llywodraethu Gwybodaeth a chaiff ei graffu yn ystod y cam asesu i sicrhau aliniad parhaus â deddfwriaeth berthnasol. Gweler adran 9 am ragor o fanylion am hyn.

Diogelu’r holl ddata personol

Mae'r ASB wedi ymrwymo'n llawn i gydymffurfio â Deddf Diogelu Data 2018 a'r Rheoliad Cyffredinol ar Ddiogelu Data a sicrhau bod data personol yn cael ei ddefnyddio'n gyfreithlon, yn deg ac yn dryloyw. Mae diogelu data yn bwysig iawn i'r ASB, nid yn unig am ei fod yn hanfodol i waith ein sefydliad, ond hefyd oherwydd ei fod yn sicrhau ein bod ni’n diogelu preifatrwydd unigol ac yn cynnal hyder y cyhoedd yn yr Asiantaeth.

Swyddogaethau a Chyfrifoldebau

Swyddogaeth	Cyfrifoldebau
Staff yr ASB sy’n tynnu data oddi ar y we	• Cydymffurfio â'r polisi tynnu data oddi ar y we • Cadw at y broses tynnu data y cytunwyd arni a sefydlwyd ar draws yr Asiantaeth
Rheoli Gwybodaeth a Diogelwch (KIMS)	• Rhannu’r polisi a gwneud staff ym mhob adran yn ymwybodol ohono fel rhan o Fframwaith Llywodraethu Gwybodaeth yr ASB yn gyffredinol

Bwrdd Llywodraethu Gwybodaeth	• Cael ei hysbysu am weithgareddau tynnu data oddi ar y we arfaethedig gan yr ASB a’u cymeradwyo • Cyfiawnhau gweithgareddau tynnu data oddi ar y we na chaiff eu cymeradwyo
Gwasanaethau Cyfreithiol	• Rhoi cyngor ar faterion cyfreithiol cyfredol a thirwedd sy'n esblygu os oes angen

Llywodraethu

Perchennog Polisi	Rheoli Gwybodaeth a Diogelwch (KIMS)
Cymeradwyo Polisi	Bwrdd Llywodraethu Gwybodaeth
Monitro Cydymffurfiaeth	Rheoli Gwybodaeth a Diogelwch a Phennaeth Adrannau
Adolygu a Gwella	Timau Rheoli a Diogelwch, a Data a Gwybodaeth

Canllawiau cysylltiedig

Camau Tynnu Data oddi ar y We

Cais i Dynnu Data oddi ar y We

- Dylai unedau busnes yr ASB sy'n dymuno tynnu data oddi ar y we gysylltu â'r Penaethiaid Data a KIMS cyn gynted â phosibl ar ôl nodi'r gofyniad
- Rhaid hefyd atodi cais tynnu data oddi ar y we wedi'i gwblhau i'w adolygu
- Nodwch y gallai’r ASB dynnu data oddi ar eich gwefan er mwyn pennu dichonoldeb. Ni fydd yr ASB yn storio unrhyw ddata yn ystod y cyfnod hwn. Bydd yr ASB yn sicrhau bod y broses hon yn parhau i fod yn unol ag adran 6 y ddogfen hon; fodd bynnag, ni fydd yn destun gwerthusiad gan y Bwrdd Llywodraethu Gwybodaeth

Gwerthuso

- Bydd y Pennaeth Data a thimau KIMS yn rhoi gwybod i’r Bwrdd Llywodraethu Gwybodaeth am y cais newydd a darparu'r ddogfen gais
- Bydd y Bwrdd Llywodraethu Gwybodaeth yn gwerthuso'r ddogfen gais ac yn gwirio a oes modd tynnu data oddi ar y we

Log Tynnu Data Oddi ar y We

Bydd y Pennaeth Data a thimau KIMS yn diweddaru'r log tynnu data canolog gyda'r gweithgareddau cymeradwy ac yn rhoi gwybod i'r uned fusnes eu bod yn gallu dechrau tynnu’r data oddi ar y we. Mae hyn er mwyn sicrhau bod y lefel gywir o ddiwydrwydd dyladwy (due diligence) yn cael ei neilltuo i bob prosiect a bod yr holl weithgareddau'n cael eu cofnodi at ddibenion archwilio ac atebolrwydd yn y Log.

Tynnu Data oddi ar y We

Gall gweithgareddau tynnu data oddi ar y we gychwyn a dim ond yn unol â'r Polisi Tynnu Data oddi ar y We a'r arferion gorau cyfatebol y byddant yn cael eu cynnal.

Manylion Cysylltu

Rheoli Gwybodaeth a Diogelwch

Swyddfa'r Comisiynydd Gwybodaeth

Rhestr o dermau

Rhyngwyneb Rhaglennu Cymwysiadau (API)

O ran tynnu data oddi ar y we, gall perchennog gwefan adeiladu API i ganiatáu mynediad hawdd at ddata a gedwir ar y wefan heb yr angen i adeiladu system tynnu data oddi ar y we.

Dyfnder y Chwilio

Mae dyfnder y chwilio yn cyfeirio at ba mor bell i mewn i hierarchaeth tudalen ar y wefan y mae’r person sy’n chwilio yn cael mynediad ato. Hafan gwefannau yw brig yr hierarchaeth hon (lefel 0), mae tudalennau sydd wedi'u cysylltu o'r hafan ar lefel 1 tra bod tudalennau sydd wedi'u cysylltu o dudalennau lefel 1 yn lefel 2 ac ati. Felly mae cyfyngu dyfnder y chwilio yn golygu cyfyngu ar ba lefel y bydd y person sy’n chwilio yn gallu cael mynediad ati.

Iaith Marcio Hyperdestun

Yr iaith farcio safonol ar gyfer diffinio strwythur tudalennau gwe. Defnyddir elfennau HTML i ddweud wrth y porwr sut i arddangos cynnwys tudalen we.

Amser segur

Amser segur, neu amser cysgu, yw'r saib rhwng pob cais a wneir i'r wefan gan y person sy’n tynnu’r data.

Chwilio cyfochrog (Parallelising the Crawling)

Mae’r system sy’n chwilio’n gyfochrog yn cynnal sawl proses ochr yn ochr gyda’r nod i gynyddu cyfraddau lawrlwytho i'r eithaf a lleihau gorbenion (overheads) tynnu data oddi ar y we. Er

enghraifft, os yw gwefannau A a B yn cynnwys nifer o dudalennau gwe, yna gallai chwilio cyfochrog ddal tudalen o wefan A, ac yna tudalen o wefan B ac ati. Os nad yw chwilio'n gyfochrog, yna mae'r gorbenion ar un wefan yn debygol o gynyddu'n sylweddol.

Protocol Eithrio Robotiaid

Gelwir hefyd yn ffeil Robots.txt. Mae hon yn safon a ddefnyddir gan wefannau i gyfathrebu â'r rheiny sy'n chwilio'r we a robotiaid gwe eraill. Yn y bôn, mae'n darparu dull i berchnogion gwefannau gyfyngu rhan o'u gwefannau rhag tynnu data oddi ar y we neu gyfyngu ar dynnu data'n gyfan gwbl i ddim ond tynwyr data ardystiedig fel peiriannau chwilio. Mae rhagor o wybodaeth am y pwnc hwn ar gael ar wefan robots.txt

Llinyn Asiant Defnyddiwr (User Agent String)

Mae Llinyn Asiant Defnyddiwr yn ddarn o destun wedi'i deilwra y gellir ei ychwanegu at broses tynnu data sy'n caniatáu i berchennog y wefan adnabod gweithredwr a/neu bwrpas y gweithgarwch tynnu data oddi ar y we. Gellir addasu hyn wrth greu'r system tynnu data oddi ar y we ac fe'i defnyddir yn aml er mwyn sicrhau tryloywder.

Chwiliwr y We (Web Crawler)

Mae chwiliwr y we, y cyfeirir ato weithiau fel pry cop, yn bot rhyngrwyd sy'n pori'r We Fyd-Eang yn systematig at ddibenion casglu gwybodaeth. Yr enghraifft fwyaf cyffredin o hyn yw peiriannau chwilio fel Google sy'n defnyddio chwilwyr gwe i gadw eu canlyniadau chwilio yn gyfredol.

Tynnu Data oddi ar y We

Diffinnir tynnu data oddi ar y we, neu 'web scraping', fel casglu data yn awtomatig o'r we gan ddefnyddio darn o feddalwedd neu raglen.