

Изх. № 42/ 20.12.2020

ДО: АТАНАС ТЕМЕЛКОВ

ПРЕДСЕДАТЕЛ

ДЪРЖАВНА АГЕНЦИЯ „ЕЛЕКТРОННО УПРАВЛЕНИЕ“

КОПИЕ: СВЕТЛОГОР КИРОВ

ОТДЕЛ „ДАННИ“

ДЪРЖАВНА АГЕНЦИЯ „ЕЛЕКТРОННО УПРАВЛЕНИЕ“

КОПИЕ: РЕНИ БОРИСОВА

ОТДЕЛ „ДАННИ“

ДЪРЖАВНА АГЕНЦИЯ „ЕЛЕКТРОННО УПРАВЛЕНИЕ“

ДО: КРАСИМИР БОЖАНОВ

ДИРЕКТОР

ДИРЕКЦИЯ „МОДЕРНИЗАЦИЯ НА АДМИНИСТРАЦИЯТА“

МИНИСТЕРСКИ СЪВЕТ

КОПИЕ: РАЛИЦА ВЕЛИЧКОВА

ИНИЦИАТИВА „ПАРТНЬОРСТВО ЗА ОТКРИТО УПРАВЛЕНИЕ“

ОТ: ТЕОДОРА ГАНДОВА

ПРЕДСЕДАТЕЛ НА УС

НПО ЛИНКС

АТАНАС КАРАШЕНСКИ

СОФТУЕРЕН АРХИТЕКТ

НПО ЛИНКС

Относно: политики и процес на отваряне на данни в машинно-четим формат

Уважаеми г-н Темелков,

Екипът на НПО Линкс има дългогодишен опит за популяризирането на инициативата за отворени данни в България. Във връзка с изпълнението на проект „Повишаване на гражданското участие в процесите на формулиране, изпълнение и мониторинг на политики и законодателство“ BG05SF OP001—2.009 г. и с проведената конференцията „Отворен диалог за отворени данни“ бихме искали да предоставим нашите препоръки относно процеса на отваряне на данни в машинночетим формат, които ние обобщихме в три посоки.

1) Повишаване на капацитета на администрацията за работа с данни

В последните години усилията на ДАЕУ бяха насочени към ясно разграничаване на понятията „обществена информация“ и „информация от общественния сектор“ (отворени структурирани данни в отворен машинночетим формат) и отражението им в ЗДОИ, Наредбата към него и други нормативни и ненормативни документи. Това позволи процесът да бъде поставен на ясна регламентирана основа, но е време този процес да бъде организационно и административно осигурен, като активно бъдат включени и другите администрации, собственици на данни.

От нашия опит в последните години е ясно, че визията и политическата воля са основополагащи за успеха в процеса на отваряне на данни в машинно-четим формат. Това беше в основата на успеха в периода 2014-2016 г., а липсата им за последващото забавяне на процеса.

Опитът ни и работата с институциите по темата "Отворени данни" в последните години показва, че в администрациите има остра нужда от структури, които от една страна да са пряко отговорни за данните, които ведомството произвежда и публикува, а от друга да имат санкция да изискват данни в рамките на собствената си администрация. Успешен пример на НПО Линкс в тази посока на местно ниво беше препоръката ни към Столична община, в резултат на действията по инициативата "Визия за София", която се реализира под формата на заместник-кмет по дигитализацията.

В тази връзка предлагаме на ниво изпълнителна власт обособяването на административна структура за данни, която освен да концентрира технологичен опит, да разполага и с нужната санкция да изисква от другите ведомства данни с необходимото качество. Тази структура, по примера на повечето европейски страни трябва има съответните правомощия за структуриране, контрол, санкции и мониторинг на отворени данни, обезпечена с човешки ресурси за провеждане на политиките, а не както е в момента да бъде само отдел „Данни“ към ДАЕУ.

Независимо дали отворените данни ще се използват за еднократен анализ, изследване или за продължаващ във времето проект, като мониторинг или услуга, необходимо е данните да бъдат качествено събирани и адекватно представени. Тъй като администрацията е източник на данни, необходимо е повишаване на капацитета по отношение на събиране, структуриране и публикуване на данни. Един от най-преките начини за подобряване на качеството на отворените данни е ролята на администрацията, като се преосмисли нейната роля в процеса и тя освен източник, да бъде и ползвател на данни.

НПО Линкс предлага всяка държавна структура или община да обособи *Chief Data Officer* с ранг заместник-министър, заместник-кмет или заместник-директор, който да отговаря за данните на отделната административна структура за тяхното структуриране, качество, надеждност и за обучението на персонала, с оглед конкретните специфики на отделните ведомства. Въвеждането на подобна фигура във всяка административна структура ще помогне за решаването на проблема с метаданните в Портала за отворени данни и ще гарантира оперативна съвместимост за обмен на метаданни с различни каталози.

По този начин ще се повиши капацитета на администрацията по отношение на отворените данни, защото ще се идентифицират дъщерни институции, които са в отношения „източник на данни – ползвател на данни“, които да участват в съвместно изработване на проект/задача свързана с тези данни. Така двете страни ще могат да осъзнаят изискванията към качеството на публикуваните данни.

На Портала за отворени данни, поддържан от ДАЕУ, има голямо количество данни от множество източници, като една немалка част от тях страдат от следните недостатъци:

Липса на метаданни. В много случаи данните не се нуждаят от метаданни, защото файловете и форматите, в които са представени ги описват достатъчно точно. В други случаи, обаче липсата на метаданни е осезаема. Последното важи с особена сила в наборите от данни, в които се използват абривиатури и номенклатури. За всеки набор е необходимо е да се посочва уникалният идентификатор, както и в какъв обхват е уникален. Когато данните в един набор имат връзка към един или повече други набори, желателно е да се описват тези връзки.

Липса на идентификатори. Идентификаторите са от определящо значение за надеждната обработка на данните и без тях е невъзможно свързването на данни от различни набори. Желателно е, когато е възможно да се използват идентификатори с максимално широк обхват.

В практиката на НПО Линкс често се налага да използваме данни, публикувани от различни институции на собствените им сайтове. Важно е да се разбира, че Excel не е еквивалентно на отворени данни в машинно-четим формат, особено когато има разлики във структурирането и форматирането между отделните файлове от един и същи тип данни, сляти клетки и др.

Дати на валидност. Голяма част от публикуваните набори показват състояние на дадени данни към определен момент. При някои набори този момент е упоменат в името, а при други не е, и тъй като липсва отделно поле, датата към която е актуален даденият набор е обект на догадки.

Гореописаните недостатъци могат да бъдат преодолени чрез две стъпки:

Повишаване на капацитета на администрацията за работа с данни;

Автоматизиране на процедурите по извличане на данните и публикуването им

Последното гарантира консистентност в публикуваните данни и дава възможност за надеждната им обработка.

2) Модел за мониторинг на публични политики, базиран на отворени данни

Общественият живот е сложна система от участници, всеки с различна степен на влияние. За да може обективно да се проследява развитието на една или друга част от него е необходимо да има ясни индикатори, които се формират на базата на надеждно събрани и обработени данни.

В тази връзка мониторингът на политиките за отворени данни може да се осъществява както от обособената агенция за отворени данни, посредством оценки за въздействието, така и от неправителствения сектор. В България има мрежа от ентузиастични за отворени данни, НПО Линкс заедно с ДАЕУ може да организира класация на административните структури за отворени данни и по този начин да ги стимулира да работят в тази посока.

Пример за мониторинг на публични политики, базиран на отворени данни е проектът по оперативна програма „Добро управление“ Граждански контрол върху реформата в съдебната система“ /BG05SFOP001-3.003/. Проектът използва данни от различни източници – Министерство на вътрешните работи, Висш съдебен съвет, Министерство на финансите (СЕБРА), анализи, направени върху отговорите на запитвания по Закона за достъп до обществена информация и др. Важно е да се отбележи, че част от данните, използвани в този проект не са публикувани в машинно-четим формат, което наложи ръчното им извличане от PDF файлове.

Системите за мониторинг не биха могли да функционират без данни – ключово е наличието на консистентно публикувани данни. В противен случай, всяка промяна във вида на данните ще изисква пренастройване на системата за мониторинг.

Препоръчителна е еднократната инвестиция в добавка към съществуващите регистри и системи за съхранение и обработка на данни, която да извършва автоматичното извличане и публикуване на данните на определен период от време. По този начин, в дългосрочен план ще се осигури регулярно публикуване, консистентен вид на данните, ще се намалят човешките грешки и ще се оптимизират необходимите ресурси.

3) Приоритизиране на набори от данни за отваряне в машинно-четим вид

По отношение на приоритизирането на видовете данни, подлежащи на отваряне може да се подходи по два начина:

Извършване на проучване на нуждите от публикуване на един или друг вид данни и последващото им публикуване;

Публикуване на данните, поддържани в дигитален вид.

Първият подход изисква ресурси за определяне на необходимите набори от данни и рискува да попадне на набор от данни, който не се съхранява в подходящ вид или има други пречки за публикуването му.

НПО Линкс смята, че приоритизирането трябва да се случва с оглед актуалните и конкретните нужди на обществото, какъвто и примерът за COVID пандемията в момента и в съответствие с актовете за изпълнение към Директива 2019/1024. Приоритизирането не трябва да се случва чрез списък, а трябва да се случва веднага, когато данните бъдат поискани и предоставени в съответния машинно-четим вид, което позволява сравнително бързо публикуване на данни и отваря възможности за широк кръг анализи и приложения, включително и такива, които използват данни от различни набори, публикувани от различни

институции. При този подход е важно да се гарантира, че наборите от данни не се отварят еднократно и „ръчно“, а се публикуват редовно и консистентно.

Заклучение

Поддържането на данните изцяло в цифров вид е подход, който залага първо на данните, а след това на изграждането на тяхна база на системи, платформи, услуги, продукти, екосистеми и др. НПО Линкс предлага на препоръките да се даде възможно най-висок приоритет, тъй като те са отключващи за процеса на реална и ефективна дигитализация, като ние допълнително ще изпратим нашите анализи и презентации по двата проекта, по които сме работили. Нашите препоръки са насочени към постигане оптимални резултати в политиките за отворени данни на всички нива и власти, като се подобри взаимодействието между държава, бизнес и граждани.

20.12.2020 г.

С уважение,
Теодора Гангова
Председател на УС
НПО Линкс