



СТЪПКА 3:

Доклад по направление



УПРАВЛЕНИЕ

КАЧЕСТВО НА ДАННИТЕ ПО РЕСОРИ И
ПОТЕНЦИАЛ ЗА ПРЕДОСТАВЯНЕ ПО
АВТОМАТИЗИРАН НАЧИН И В МАШИННО-ЧЕТИМ
ФОРМАТ

декември 2018
www.vizia.sofia.bg

РЕЗЮМЕ	2
ВЪВЕДЕНИЕ.....	4
КАЧЕСТВО НА ДАННИТЕ ПО РЕСОРИ И ПОТЕНЦИАЛ ЗА ПРЕДОСТАВЯНЕ ПО АВТОМАТИЗИРАН НАЧИН И В МАШИННО-ЧЕТИМ ФОРМАТ	6
1. МЕТОДОЛОГИЯ.....	6
2. РЕЗУЛТАТИ ОТ ИЗСЛЕДВАНЕТО	7
2.1 Уточнения:	7
2.2 Източници на данни.....	7
2.3 Съхранение на данните и работа с тях.....	8
2.4 Възможности за повторно използване и разширена обработка на данните	9
Първичен ключ	10
Формати.....	10
Достъп чрез програмен интерфейс.....	10
Цифрово подписване	11
Публичен достъп.....	11
2.5 Проблемни области.....	12
2.6. Анализ на сигурността на данните и методите за съхранение	12
Автентичност на източника	12
Коректност на входните данни.....	14
Съхранение на данните и достъп до тях	16
Заключение.....	17
2.7. Възможности за привеждане на данните в машинно-четим вид	17
3. Цели и мерки за подобряване на управлението на Столична община.....	18
Четири основни цели.....	18
Метод на работа при определяне на цели и мерки	19
ИЗВОДИ И ПРЕПОРЪКИ	21
Първите стъпки.....	21
Повишаване на капацитета на служителите за работа с данни	22

РЕЗЮМЕ

Данните са в основата на всички процеси, течащи в Столична община – както свързаните с директно обслужване на граждани и различни видове организации, така и процесите, подпомагащи общината в основните ѝ дейности.

Цели на изследването са каталогизиране на данните, събирани и обработвани от Столична община и оценка на тяхното състояние – сигурност, форми на употреба и възможности за машинна обработка. Последното е от изключителна важност за възможностите на Столична община да извършва анализи и да предприема мерки, базирани на реални данни.

Анализът е извършен на базата на два типа интервюта със служители от различните звена на Столична община – за идентифициране на наборите от данни и за тяхното подробно описание. Вторият тип интервюта следва предварително подготвен въпросник, съставен от 62 въпроса, отнасящи се към вида данни, формите на съхранение, достъп, сигурност, начините за събиране и употреба.

Идентифицирани са 248 набора данни, от които 106 подробно описани. Причините за липсата на подробно описание на данните са от различен характер: технически – блокиран достъп до онлайн въпросника; организационни – липса на време от страна на съответните служители.

Събраните данни се обработват и съхраняват в 18 специализирани информационни системи, автоматизираната информационна система на Столична община „Акстър офис“, файлове съхранявани на сървъри на общината, на потребителски компютри и на хартиен носител. Данните съхранявани в информационните системи и на сървъри на общината подлежат на автоматично архивиране и достъп до тях се осъществява след проверка за автентичност на потребителите. Около 23% от описаните набори се съхраняват по начин, изключващ автоматизираното архивиране и в известна степен регламентирания достъп – на потребителски компютри и на хартиен носител.

Деловодната система „Акстър офис“ съдържа в голяма степен както входните, така и изходните данни от различните процеси. Това са основно документи в сканиран вид – определено непригодни за надеждна машинна обработка – пълнотекстово търсене, справки и анализи. Достъпът до данните чрез програмен интерфейс е възможен в 13,9% от описаните набори. Малкият брой на такива интерфейси и като следствие от това невъзможността за системите да обменят информация, води до дублиране на едни и същи данни в различни системи.

Основните предложения и препоръки могат да бъдат обобщени по следния начин:

www.vizia.sofia.bg

- Реално дигитализиране на данните и процесите. Сканираните документи не дават възможно за автоматична машинна обработка и заемат голям обем. Съхраняването на документите в цифров вид ще даде възможност за обработването им с аналитична цел и ще улесни публикуването им в машинно-четими формати и повторната им употреба. Използването на цифрови подписи ще премахне необходимостта от сканираното копие на документа и ще катализира процеса по дигитализиране на данните.
- Координиране на заданията за проекти/поръчки на системи, работещи с данни с ИТ отделна на Столична община. Необходимо е всички следващи проекти да съхраняват данните по подходящ начин и да дават възможност за обмен с другите информационни системи на Столична община.



ВЪВЕДЕНИЕ

Настоящото изследване цели описването както на всички набори от данни, събирани и обработвани от Столична община, така и да оцени тяхното състояние през целия цикъл на употребата им.


В процеса на настоящото изследване бяха проведени два типа интервюта с представители на всички дирекции в Столична община – за идентифициране на наборите от данни и за тяхното подробно описание. Идентифицирането на наборите от данни беше извършено с интервюта, в които участваха директорите на дирекции в отделните направления, а подробното описание следваше предварително подготвен въпросник от 62 въпроса. Идентифицирани са 248 набора от данни, от които 106 са подробно описани.

Попълването на въпросниците от страна на служители на администрацията среща някои проблеми:

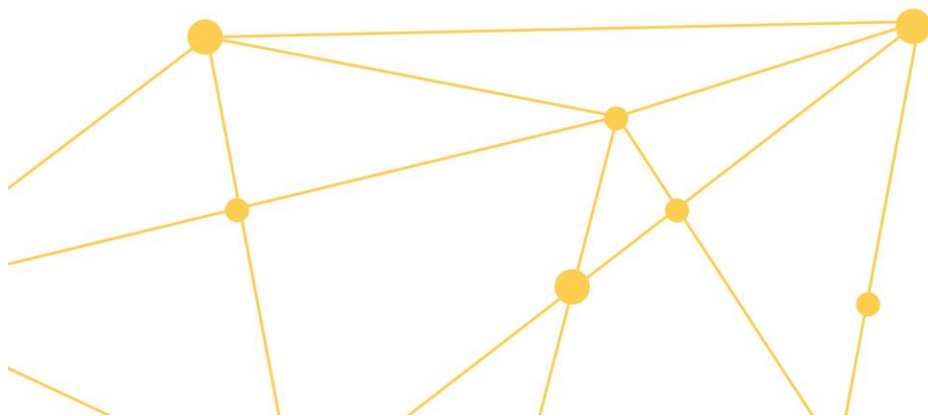
- Достъпът до въпросниците от офисите на Столична община беше блокиран от системата за достъп до Интернет;
- Служителите от администрацията на Столична община не във всички случаи можеха да отделят нужното време и внимание за да попълнят отделен въпросник за наборите данни, с които работят;
- Част от служителите на Столична община срещаха затруднения със спецификата на някои от въпросите, касаещи технически аспекти на обработката и съхранението на данните.

Резултатите от формалното описание на данните показват, че потокът от данни е изключително голям. Ежедневно се въвеждат и обновяват голямо количество данни. Най-големият източник и най-големият потребител на тези данни е самата Столична община. В значителна степен източниците подават данните в структуриран вид, като най-голям е дялът на хартиения носител. По отношение на съхраняването на данните в 32% от случаите също се използва хартиен носител.

Потенциалът за подобряване на работата на Столична община във всички етапи на обработката и съхранението на данни е голям. Съществено е администрацията да промени отношението си към данните в посока пълната им дигитализация в структурирани машинно-четими формати и готовност за споделяне с други звена на общината и когато е възможно и публично.



КАЧЕСТВО НА ДАННИТЕ ПО РЕСОРИ И ПОТЕНЦИАЛ ЗА ПРЕДОСТАВЯНЕ ПО АВТОМАТИЗИРАН НАЧИН И В МАШИННО-ЧЕТИМ ФОРМАТ



КАЧЕСТВО НА ДАННИТЕ ПО РЕСОРИ И ПОТЕНЦИАЛ ЗА ПРЕДОСТАВЯНЕ ПО АВТОМАТИЗИРАН НАЧИН И В МАШИННО- ЧЕТИМ ФОРМАТ

1. МЕТОДОЛОГИЯ

Изследването се проведе в два отделни етапа:

- Интервюта с представители на всички звена на Столична община за идентифициране на наборите от данни, които съответното звено събира, обработва и поддържа. В хода на тези интервюта служителите от администрацията бяха поканени да разкажат за естеството на работата си по отношение на данни. Като набори от данни бяха определени както регистрите вменени на звеното по силата на закон или наредба, така и данните, които се използват с цел по-добра оперативност и удобство;
- Попълване на въпросник, описващ източниците на данни, техните потребители, начините на съхранение и обработка. Въпросникът отделя особено внимание на сигурността на данните, съхранението им и достъпът до тях. За всеки набор от данни се изисква попълването на отделен въпросник.

Попълването на въпросниците от страна на служители на администрацията среща проблеми от различен характер:

- Технически – Въпросникът беше реализиран в Google Forms, а системата на Столична община за достъп до Интернет възпрепятстваше попълването му;
- Организационни – Служителите от администрацията на Столична община в повечето случаи не можеха да отделят нужното време и внимание за да попълнят отделен въпросник за всеки набор от данни, с който работят. Попълването на отделен въпросник за всеки набор е необходимо условие за точното и подробно картиране на данните;
- Капацитет и комуникация – Част от служителите на Столична община срещаха затруднения със спецификата на някои от въпросите, касаещи технически аспекти на обработката и съхранението на данните. В допълнение голяма част от служителите нямаха информация за целите на Визия за София и конкретното изследване, като повечето от тях предполагаха, че имаме нужда от самите данни и препращаха към местата, където те са публикувани.



Екип на НПО Линкс посети всички звена, заявили готовност за попълване на въпросниците, но изпитващи затруднения от различен характер и с помощта на мобилен Интернет достъп и консултанти на място подпомогна попълването на въпросниците.

2. РЕЗУЛТАТИ ОТ ИЗСЛЕДВАНЕТО

В изследването работата с данни на Столична община е разделена на три основни етапа – получаване, обработка и съхранение и потребление. Във всички етапи участват различни видове субекти от обществения живот – физически лица, частни юридически лица, структури на държавата и структури на общината.

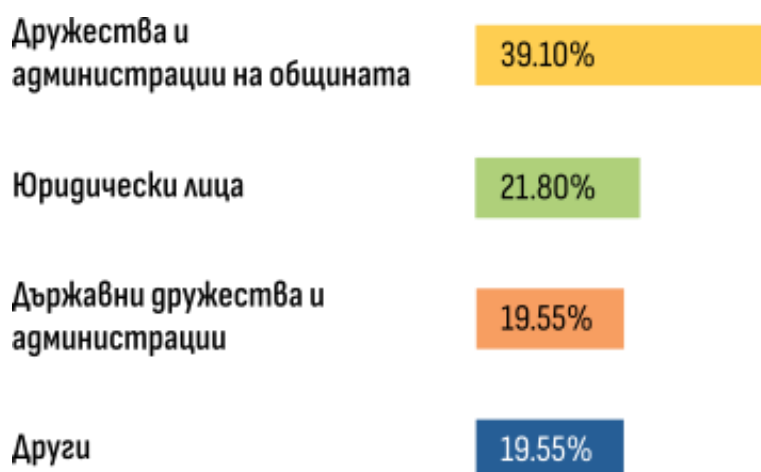
2.1 Уточнения:

- Данните в проценти/съотношения в резултатите от изследването се отнасят за набори от данни, а не за количествата данни, които те съдържат;
- Данните в проценти/съотношения в резултатите от изследването се отнасят само за наборите данни за които има попълнени въпросници, т.е. 100% = 106 набора от данни;
- Данните в изследването са формирани само от отговорите на интервюираните служители на Столична община, а не в резултат на проверка на работата с данни.

2.2 Източници на данни

Столична община ежедневно приема за обработване огромно количество данни. Съществено е да се прави разлика между видове източници и обемите от данни, които всеки източник предоставя.

Анкетиранияте са определили като най-често използвани източниците на данни както е посочено на графика 1.



Графика 1 – източници на данни

Според интервюираните служители най-големият обем данни, обаче идва от физическите и частните юридически лица във вид на сигнали, заявления и искания за услуги.

Данните се предоставят най-често в структуриран формат – обикновено това са образци на документите, необходими за предоставяне на услуги от общината, сигнали и заявления. Контактният център на Столична община приема информацията от гражданите и я въвежда, доколкото е възможно, в структуриран вид. На подсайт на Столична община (<https://usl.sofia.bg>) са публикувани образци на заявления за услуги. От 95 услуги, заявления за 45 от тях могат да бъдат подадени онлайн в структуриран формат в цифров вид. Останалите 50 услуги изискват попълнените бланки да бъдат отпечатани, подписани и физически представени на гише в Столична община. Хартиеният носител е посочен от интервюираните като най-често използван носител на данни, включително и за обслужване на процеси между отделни звена на Столична община.

Данните се проверяват за автентичност на източника им при постъпването им за обработка в даденото звено в 54,5% от случаите. В 11,9% служителите считат, че автентичността на източника е гарантирана от определени атрибути на данните. В останалите 33,7% не се прави проверка за автентичност.

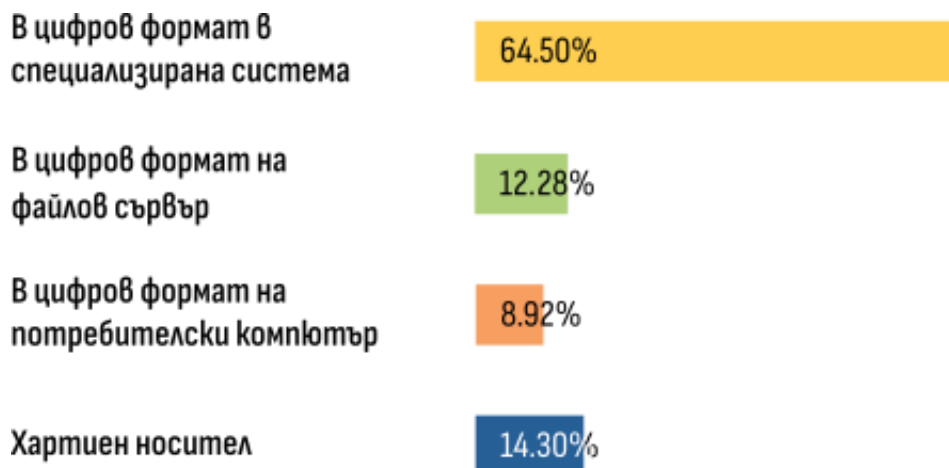


Графика 2 – проверка за автентичност на източника на данни

Коректността на данните е от изключителна важност за нормалното функциониране на общината. Проверки за коректност на данните се извършват в 69,3% от процесите, обработващи данни, в 5,9% данните се считат за коректни поради вида на източника или определени атрибути в тях, а в 20,8% не се прави проверка за коректност на данните.

2.3 Съхранение на данните и работа с тях

Съхранението и обработката на данни в Столична община се извършва чрез използването на 18 специализирани системи, автоматизираната информационна система на Столична община „Акстър офис“, файлове съхранявани на сървър на общината или на потребителски компютър и на хартиен носител. Най-често използваните носители на данни са показани на графика 3.



Графика 3 – съхранение на данните

В повечето случаи дейностите, извършвани от звената по конкретни преписки са придружени от документация в деловодната система „Акстър офис“. Файловете, прикачени към всяка преписка най-често са сканирани документи, придружени от кратка информация в свободен текст.

Файловете сървъри са собственост на Столична община и са разположени в специализирания ѝ център. Повечето информационните системи са разположени на системи, собственост на общината, а останалите са в процес на преместване или планирани за преместване.

Всички системи, разположени върху хардуер на Столична община са обект на регулярно архивиране, като готовите архиви се съхраняват физически на различно място. **На автоматизирано архивиране не подлежат потребителските компютри, както и данните на хартиен носител, т.е. около 23% от наличната информация.** Решение на потребителите е кога и как ще правят резервни копия на съдържанието на ползваните от тях компютри. В повечето случаи това се случва рядко и нерегулярно.

Достъпът до данните във файловете сървъри и специализираните системи се осъществява след проверка за автентичност на потребителите. За всеки потребител се създава достъп в специализираните системи и файловете сървъри в съответствие с длъжността му и по надлежния рег. Повечето от тези системи, както и файловете сървъри имат възможност да предоставят информация за дейностите извършвани от потребителите – *audit trail*. Информация за действията на отделните потребители не може да се получи за данните съхранявани на потребителски компютри и на хартиен носител. Според интервюираните служители на Столична община в 16% от процесите за обработване на данни, достъпът до тях не е регламентиран.

2.4 Възможности за повторно използване и разширена обработка на данните

Събраните веднъж данни могат да бъдат използвани повторно за най-различни цели – анализи, прогнози, журналистически разследвания, системи подобряващи ежедневието



на групи хора и организации и др. Ползватели на данните могат да бъдат най-различни субекти от обществения живот – други отдели на общинската администрация, държавна администрация, граждани и организации. За да може да се осъществи повторна употреба на данните е необходимо данните да бъдат представени и поддържани подходящо. – това включва минимум поддръжка на първични ключове, подходящи формати и достъп през програмен интерфейс (API)

Първичен ключ

Данните, поддържани в специализирани системи имат формиран първичен ключ. Доста често това е автоматично увеличаващ се номер, без претенции за уникалност извън съответната система и е основно предназначен за ползване изключително в рамките на съответната система. Все още има набори от данни, които не поддържат първичен ключ: 35.6% според данните от интервюта, направени със служители на администрацията. Това са основно тези набори, които се поддържат във вид на потребителски файлове и не са обект на специализирана система. На тези места служителите емпирично различават отделните записи в набора по комбинация от полета, но този подход не е приложим за машинна обработка и връзка между системи. Наборите от данни, които се поддържат на хартиен носител, най-често се отнасят за данни подадени от граждани или други звена на администрацията на Столична община чрез деловодната система „Акстър офис“ и използват входящия номер, генериран от тази система за първичен ключ и реферирание на конкретен запис.

Формати

Специализираните системи, в общия случай, съдържат данните в подходящи формати. Това не винаги означава, че тези данни се публикуват в машинно-четими и използвани за последваща обработка формати. Не се отчита спецификата на употребата на публикувани данни, които съдържат дати и адреси, което ги прави на практика невъзможни за ефективна употреба от страна на други звена на Столична община, граждани и фирми. 25% от подлежащите на публикуване набори или под 10% от

обсъжданите набори от данни се публикуват във вид удобен за машинна обработка. 62% от публикуваните набори са във формати абсолютно непригодни за машинна обработка. Деловодната система „Акстър офис“ съдържа в голяма степен както входните, така и изходните данни от различните процеси. Това са основно документи в сканиран вид – определено непригодни за надеждна машинна обработка – пълнотекстово търсене, справки и анализи.

Достъп чрез програмен интерфейс

Достъпът до данните чрез програмен интерфейс не е особено популярен при наборите от данни, обработвани от Столична община – едва 13,9% от наборите поддържат такъв. Липсата на такива интерфейси и като следствие от това невъзможността за системите да обменят информация, води до дублиране на едни и същи данни в различни системи. Всяка система ги поддържа по различен начин и с различни характеристики. Пример за това са договорите, които Столична община сключва с изпълнители на

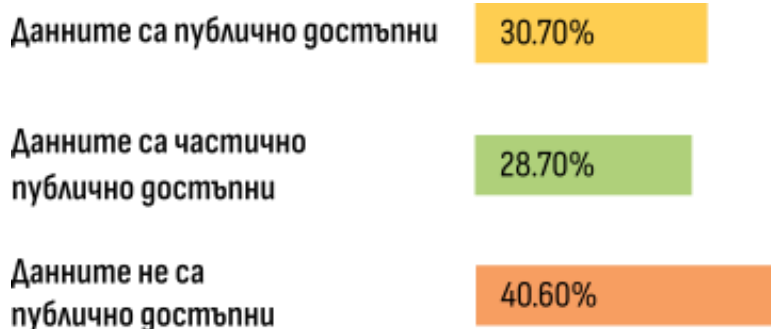
различни дейности – присъстват сканирани в деловодната система „Акстър офис“ и с различни мета данни в системите на звената, контактуващи с изпълнителите. Отделно данни за договорите се съдържат и в счетоводната система „Комси 54“, където описанието им е едно от най-пълните според служителите от направление „Финанси“.

Цифрово подписване

Всеки изведен от Столична община документ съдържа поне един подпис. Немалка част от дейността на общината се състои в издаване на различни документи, голяма част от които трябва да бъдат съгласувани между две или повече звена. Физическото полагане на подписи, когато става дума за две или повече звена е процес, който може да отнеме повече време и човешки ресурс, отколкото изготвянето на съответния документ по същество. Различните звена са разпръснати в различни офиси на територията на гр. София и понякога се налага един документ да измине буквално километри преди да бъде завършен..

Публичен достъп

Според интервюираните служители 40,6% от наборите данни, обработвани от Столична община не са публично достъпни, а 28,7% са частично достъпни поради наличие в тях на лични данни – 15,2%, професионална тайна – 1,7% и други видове тайна. В около 2,8% от наборите става дума за предварителна, в процес на подготовка информация. В 30,7% от случаите данните са с право на пълен публичен достъп.



Графика 4 – подлежаат ли данните на публичен достъп

Публикуваната информация, най-често – 75%, е във вид неподходящ за надеждна машинна обработка – сканирани документи, .pdf файлове, таблици на уебсайт, документи в свободен текст. **25% от публикуваните набори или 7,6% от всички идентифицирани набори е в машинно-четим вид и подходящ за повторна обработка.**

Столична община спазва формално ангажимента си за публикуване на данни за повторна употреба, но те не са приведени в необходимия вид. Причините за това са:

- поддържането на оригиналната информация в неподходящи файлови формати. Най-често това са формати ориентирани към отпечатване на документи – Ms Word, PDF;



- невъзможността на част от системите за автоматизирано експортиране на съответната информация;
- необходимостта от подписи в публикуваните документи, съчетана с неизползването на цифрови подписи от страна на администрацията на Столична община

2.5 Проблемни области

Голямо количество сканирани документи, които не дават възможност за търсене и автоматизирано обработване. Сканираните документи, поради по-големия си обем спрямо файловете с данни, водят и до необходимостта от многократно по-големи масиви за поддържането им.

Използването на подписи на хартиен носител, вместо цифрови такива върху съответен файл е сериозна пречка за поддържането на данните в изцяло цифров и удобен за обработка вид. В допълнение неизползването на цифрови подписи е част от причините за публикуването в неподходящи формати – напр. всеки график за различните видове дейности трябва да бъде одобрен, респ. подписан и се публикува като сканиран документ.

2.6. Анализ на сигурността на данните и методите за съхранение

За нормалното ежедневно опериране на различните звена на общината е изключително важно данните с които работят да са надеждни, т.е. да идват от автентичен източник, да са коректни към момента на подаване, актуални в дадения момент, да се съхраняват по начин който запазва цялостта им независимо от хардуерни проблеми, обработката им да е надеждна и да се достъпват само от оторизирани субекти – потребители и достъп през програмно-приложни интерфейси.

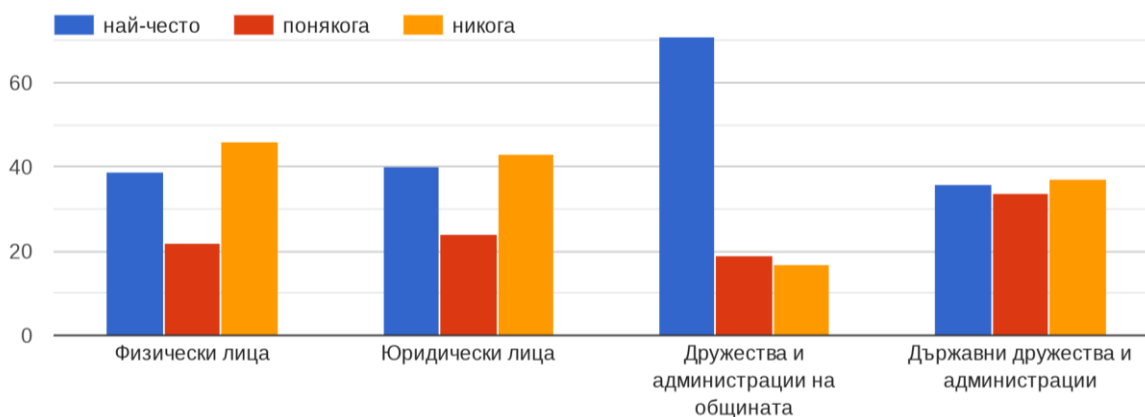
Важно е да се отбележи, че графиките в този анализ отразяват набори от данни, а не данните, които те съдържат.

Автентичност на източника

Данните започват “живота” си в Столична община от техния източник. Източниците могат да бъдат най-различни – физически лица, юридически лица, дружества и администрации на общината и държавни дружества и администрации.



От къде идват данните при Вас?



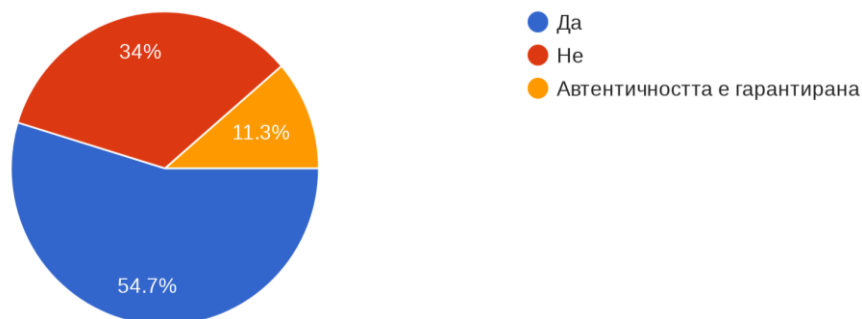
Фигура 1

Както се вижда от *Фигура 1* общината е източник на данни за най-много набори. Това включва и районните администрации като източник на данни. Останалите видове източници най-често се използват почти по равно.

В големи организации – структура и брой хора, както и организации, контактуващи с голям брой външни за нея субекти рискът от т.нар. social hacking нараства значително. За намаляването му е необходимо да се проверява автентичността на източниците на данни, независимо от техния вид. Желателно е броят на звената през които могат да постъпят данни, независимо от източника да е относително малък, а служителите добре обучени да проверяват за автентичност източниците. Проверката за автентичност за служителите в звената, явяващи се входни точки за данни е рутинна операция и като такава е с по-малки възможности за човешка грешка¹ и пропуски.

Извършва ли се проверка за автентичност на източника на данни?

106 responses



Фигура 2

1 Smith, Dr David J., "Reliability and Maintainability and Risk", Extracts from Appendix 6, 7th Edition, Elsevier, 2005

Според отговорите дадени от служителите, в 34% от наборите данни не се прави проверка за автентичност на източника на данни. При някои видове набори от множеството на наборите при които не се прави проверка за автентичност на източника такава проверка е неприложима – напр. “Списък на използваните в дейността на отдел Транспортно обслужване образци от документи”, “Телефонна централа”, докато при други набори, напр. “Разрешение за ползване на тротоарно право, оценки за отстраняване на гървесна растителност” е необходимо да се въведе.

В случаите, когато служителите смятат, че автентичността е гарантирана, се позовават на типа на източника – държавни органи и звена на Столична община, както и на реквизити от типа на подпис и печат. Точно такъв тип ситуации са удобни за social hacking – когато запознати с начина на работа и структурата на съответните организации могат да злоупотребят за да постигнат желан от тях резултат.

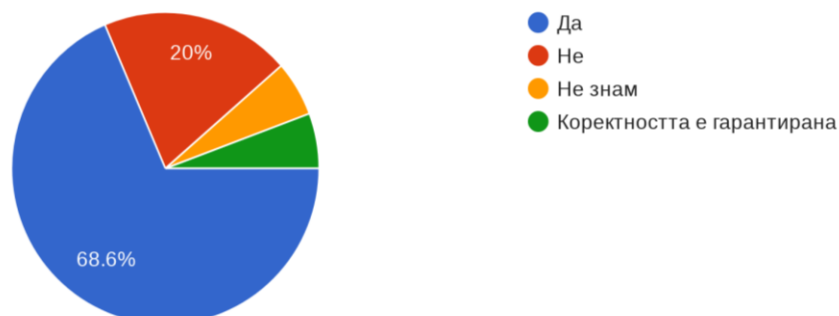
Коректност на входните данни

Некоректните данни всъщност не са данни, те могат, в зависимост от вида си, да окажат по-малко или по-голямо влияние върху вземаните решения.

Фигура 3 показва доколко се извършва проверка за коректност на входните данни. В 20% от обсъжданите набори не се прави проверка за коректност, а в още 5.7% не е ясно дали се извършва такава проверка. В други 5.7% от обсъжданите набори служителите считат, че коректността на данните е гарантирана, като основна причина за това посочват източника на данни. Тук, аналогично на проверката за автентичност на източника на данни, има набори при които проверка за коректност не е приложима – източникът носи отговорност за коректността на подадените данни, напр. “Регистър на постъпилите в дирекция “Финанси” одитни доклади”, докато за други е необходимо да се извършва. В определени случаи проверката за коректност на подадените данни не е възможно да се осъществява непрекъснато, а се извършва под формата на проверки на място или по сигнали, напр. “Регистър на детските градини”.

Извършва ли се проверка за коректността на данните?

105 responses

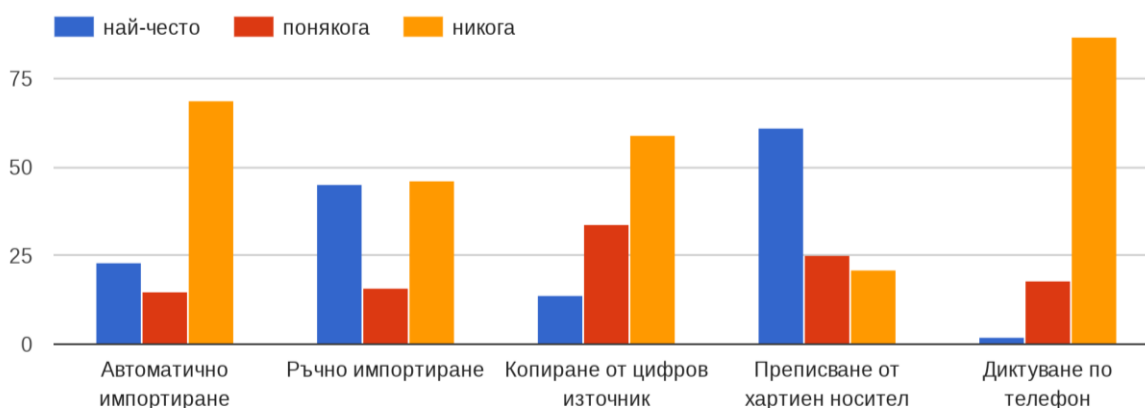


Фигура 3

Съществено е по какъв начин данните достигат от източника до системата, в която ще се съхраняват. Колкото повече ръчна обработка има, толкова повече нараства вероятността за човешка грешка или умишлено недобронамерено модифициране.

На *Фигура 4* са представени начините за въвеждане на данните. Най-често, според служителите това се случва чрез преписване от хартиен носител, което е подход, който позволява грешки и недобронамерена промяна на данните.

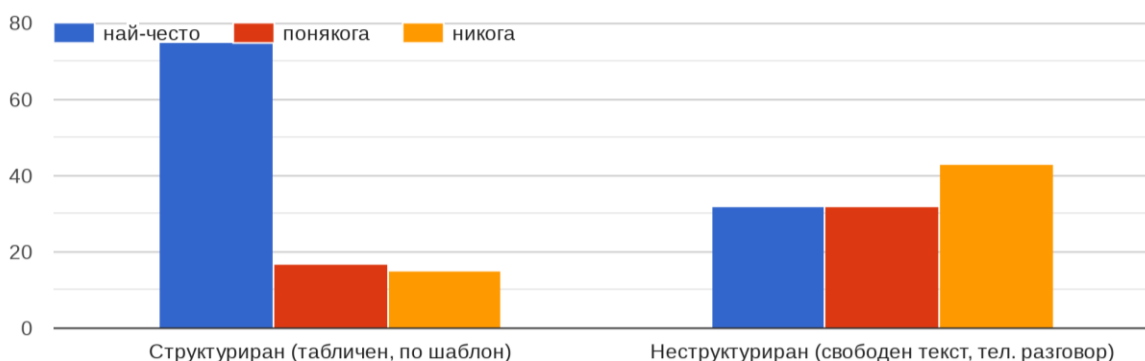
Как се въвеждат данните?



Фигура 4

Друг съществен фактор, който може да окаже влияние на коректността на данните е когнитивното натоварване на служителите при въвеждането им в системата за съхранение/обработка. На *Фигура 5* са представени използваните формати за предоставяне на данни в две основни групи – структурирани и неструктурирани. Неструктурираните формати, използвани приблизително в 1/3 от обсъжданите набори изискват по-голямо когнитивно усилие респ. са по-вероятен източник на грешки.

Какъв е форматът на предоставените данни?



Фигура 5



Съхранение на данните и достъп до тях

Данните на Столична община се съхраняват в 19 информационни системи, във вид на файлове на сървъри на Столична община, във вид на файлове на потребителски компютри на Столична община и на хартиен носител.

- Комси 54
- Инфо ТМ
- Информационна система за местни данъци и такси (МАТЕУС)
- Информационна система на Контактния център на Столична община
- Локална база данни "Население"
- Информационна система за обслужване на детските градини (ИСОДЗ)
- ИС "Промис"
- Регистър на сдруженията на собствениците в режим на етажна собственост
- ИС за управление на отпадоците
- Система за контрол и управление на съдебните дела (СКУД)
- Административна информационна система за общински електронни услуги
- Регистър на даренията
- ИС на човешките ресурси Human
- DocuWare
- Информационна система за Столична програма "Култура"
- Система за поддържане профила на купувача на Столична община
- Информационна система на НАГ
- Информационна система за проследяване разхода на горива
- Система за съхранение на данни на ГИС

В 43.8% от случаите служителите посочват че данните не се съхраняват в никоя от посочените системи, а по следните начини:

- На хартиен носител – 45.75%
- В цифров формат на файлов сървър – 31.5%
- В цифров формат на потребителски компютър – 22.75%

Начинът на съхранение на данните има сериозно отношение към възможностите за умишленото или неволното им компрометиране.

Големият брой информационни системи, съчетан с липсата на единна система (напр. директорийна услуга) за проверка на автентичността на потребителите и авторизирането им не позволява да се наложи единна политика за идентификация – потребителско име и парола, двуфакторна идентификация. В допълнение се дава възможност за съществуването на активни профили на напуснали служители.

Данните, съхранявани в специализираните системи и като файлове на сървъри на Столична община са обект на редовно архивиране, докато тези, които се съхраняват на потребителски компютър или на хартиен носител – приблизително ¼ от всички набори от данни – не се архивират, респ. могат да бъдат увредени или загубени при различни обстоятелства.



Заклучение

Информационната сигурност не е продукт, а процес, който изисква непрекъснато внимание. За постигане на висока сигурност е необходимо да се идентифицират местата на които е възможно неволно или умишлено компрометиране на данни, което следва да е обект на дейност на специализирано звено, и да се вземат съответните мерки. От данните събрани по време на картирането на данни могат да се направят следните препоръки:

- изграждане на единна система за проверка за автентичност на потребителите и оторизацията им, напр. Active Directory, LDAP и интегриране на съществуващите информационни системи с нея;
- осигуряване на процес за архивирането на всички данни, които в момента не подлежат на архивиране – съхраняваните на потребителски компютри;
- одит на сигурността на съществуващите информационни системи;
- одит на процесите на събиране, съхраняване и обработване на данни.

2.7. Възможности за привеждане на данните в машинно-четим вид

Столична община поддържа данните си в специализирани системи, във файлови формати от офис пакет и на хартиен носител. От своя страна специализираните системи използват реляционни бази данни или проблемно ориентирани такива, напр. предназначени за работа с географски координати, полигони и др.

За да може данните да бъдат приведени в машинно-четим вид е необходимо:

- да бъдат в цифров вид;
- да имат установена и постоянна във времето структура с добро ниво на гранулиране;
- всеки запис/обект да има първичен ключ;

Всички специализирани системи в голяма степен отговарят на тези изисквания. Тези условия са абсолютно задължителни за функционирането на въпросните системи. Важно е да се уточни, че данни в цифров вид изключва сканирани документи, тъй като те не могат да бъдат надеждно обработени машинно. Това означава, че една много голяма част от документите в информационната система “Акстър” не могат, на този етап, напълно да бъдат приведени в машинно-четим вид дори и да се използва OCR (Optical Character Recognition) – тяхната структура е загубена. Аналогично на “Акстър” Комси 54 също съдържа информация за договорите, сключени със Столична община, но и там те са във вид на сканирани документи.

Имайки предвид, че тенденцията е от една страна да се отварят най-различни системи, а от друга за да са истински полезни е необходимо данните да са свързани, е важно да се отбележи как следва да се реализира първичния ключ. В повечето системи първичният ключ към даден обект/запис е уникален в рамките на таблицата, която го съдържа, т.е. той не е уникален дори в рамките на базата данни, което е предпоставка за затруднения при изграждането на връзки между данните. Препоръчително е да се

премине към глобално уникални първични ключове, напр. UUID2, което значително ще улесни реферирването на данни между различните системи.

В зависимост от потока данни и нуждите от тези данни в машинно-четим вид те могат да се представят като:

Снимка на състоянието в даден момент (експорт) – подходящо за данни, които не обновяват често. Този подход е подходящ и когато обемът е голям и интересът към данните е специализиран, напр. различните географски информационни системи.

Програмно-приложен интерфейс (API) – подходящ за данни, които се обновяват често и към които има голям интерес.

Съществуващите системи могат да реализират и двата подхода – въпрос на преценка на начина на употреба на данните и собствеността им.

Силно препоръчително е когато се използва експортиране на данни в машинно-четим вид то да се извършва автоматично на определен период – практиката показва, че когато има човешко участие данните не се експортират редовно.

Всички съществуващи системи с помощта на програмни средства (привеждане на съществуващите данни в нужния вид – най-вече генериране на уникални първични ключове и подмяна на референциите) и модификация на кода (да поддържа направените промени в данните) биха могли да доставят данните, които съдържат в машинно-четим вид.

Данните, които се съдържат във потребителски файлове и особено тези на хартиен носител не могат да бъдат приведени в машинно-четим вид без ръчна работа. Компютърните файлове могат да бъдат експортирани в машинно-четим вид, но този подход не е надежден – не може да се разчита, че ще се случва редовно, в допълнение е възможно волно или неволно разместване на колони или промяна на смисъла им, което да компрометира последващата им обработка.

За да може Столична община да разчита на данни в машинно-четим вид е необходимо всички данни на хартиен носител да преминат на цифров носител при максимално добра структура. Поддържането на данните в структуриран вид дава възможности за всякакви обработки, т.е. ще осигури възможности за трансформации с програмни средства за дълъг период напред (future proof).

3. Цели и мерки за подобряване на управлението на Столична община

Четири основни цели

Съвета на Европа дефинира дванадесете принципа на доброто управление³, които залягат в основата на Стратегията за иновации и добро управление на местно ниво, одобрена през март 2008г.

От тези дванадесет принципа, определихме четири цели, които имат силно отношение към това как Столична община използва, съхранява и произвежда данни, както за обслужване на гражданите и бизнеса, така и за целите на управлението, а именно:

1. Ефективност и ефикасност – да се правят правилните неща и нещата да се правят по правилните начини. Рационално използване на наличния ресурс –

2 https://en.wikipedia.org/wiki/Universally_unique_identifier

3 <https://www.coe.int/en/web/good-governance/12-principles-and-elope>

човешки, времеви и финансов. Сътрудничество между всички участници в процесите – служители на Столична община, дружества на общината, експлоатационни дружества, държавни структури, граждани и бизнес.

2. Отвореност и достъпност – лесен достъп до услуги за всички групи граждани.
3. Прозрачност и отчетност – публичен достъп до цялата некласифицирана информация по начин и ред, определен от закона. Възможност за ефективно проследяване на постигнатите резултати.
4. Гъвкавост и иновативност – нови и ефективни решения на проблемите и да се приоритизират съвременните методи за предоставяне на услуги. Готовност за въвеждане и тестване на нови програми, както и за използване на натрупания чужд опит.

Реализацията на мерки, водещи до постигане на тези цели и до подобряване на показателите, свързани с тях е начинът да се подобри управлението на Столична община.

Метод на работа при определяне на цели и мерки

Методът на работа е съставен от следните стъпки:

1. Анализ на актуалното състояние на боравенето на Столична община с данни

Анализът е извършен на базата на два типа интервюта със служители от различните звена на Столична община – за идентифициране на наборите от данни и за тяхното подробно описание. Вторият тип интервюта следва предварително подготвен въпросник, съставен от 62 въпроса, отнасящи се към вида данни, формите на съхранение, достъп, сигурност, начините за събиране и употреба.

По време на проучването бяха идентифицирани, най-общо, три проблемни области:

- Една част от данните все още се поддържа на хартиен носител, а в друга данните не са реално дигитализирани – поддържат се във вид на сканирани документи. Като резултат от това обработката им с аналитична цел е много трудна и публикуването им за повторна употреба в машинно-четими формати е на практика невъзможно.
- Комуникацията между съществуващите системи за съхранение и обработка на данни е много слаба.
- Недостатъчен капацитет в отделните звена по отношение на организацията на работа с данни.

2. Определяне на специфични цели и индикатори за изпълнението им

Определянето на специфичните цели се извърши на три етапа:

- Предложение за цели, мерки и индикатори от всеки отделен екип;
- Две сесии от дискусии в мултидисциплинарни екипи;
- Итеративно обобщение и изравняване на нивото на детайлност на отделните цели и мерки.

3. Определяне на мерки, за постигане на специфичните цели

При определянето на мерките за подобряване на работата с данни на Столична община, основните акценти са:

КАЧЕСТВО НА ДАННИТЕ ПО РЕСОРИ И ПОТЕНЦИАЛ ЗА ПРЕДОСТАВЯНЕ ПО АВТОМАТИЗИРАН НАЧИН И В МАШИННО-ЧЕТИМ ФОРМАТ

- Максимална дигитализация на данните и процесите в Столична община, за да може управленските решения да се взимат на базата на реални и актуални данни;
- Поставянето на данните на централно място в управлението на Столична община, както по отношение на политики, базирани на данни, така и като звено в структурата на общината. Това включва и повишаване на капацитета на служителите на Столична община за работа с данни;
- Промяна на подхода на изграждане и обновяване на софтуерни системи за работа с данни от waterfall в посока Agile⁴:

“Ние откриваме по-добри начини за разработване на софтуер чрез постоянна практика и помагайки на другите. В процеса на работа започнахме да ценим:

- **Хората и взаимоотношенията** повече от процесите и инструментите
- **Работещия софтуер** повече от изчерпателната документация
- **Сътрудничеството с клиента** повече от формализирането на договора
- **Реагирането на промяната** повече от следването на плана
- Ние разбираме, че нещата отгясно са важни, но повече ценим нещата отляво.”
- В България масовият подход при изготвяне на заданията и впоследствие договорите в обществените поръчки за софтуерни системи е да се използва waterfall подход при изграждането на съответната система. Практиката, обаче показва, че той не е достатъчно ефективен и гъвкав. Agile е ефективен не само за бизнес приложения, но и за нуждите на администрациите.^{5 6}
- Централизирано съхранение на данните.

4 <http://agilemanifesto.org/iso/bg/manifesto.html>

5 https://www2.deloitte.com/content/dam/insights/us/articles/3897_Agile-in-government/DUP_Agile-in-Government-series.pdf

6 <https://techfarhub.cio.gov/handbook/>

ИЗВОДИ И ПРЕПОРЪКИ

Първите стъпки

По време на неформалните интервюта със служителите на Столична община ни направи впечатление, че всички документи, които общината изготвя се подготвят в цифров вид (MS Office, MS Excell), но се съхраняват във вид на сканиран документ. Това е огромен проблем пред последващата им обработка за анализ и увеличава в пъти нуждата от пространство за съхранението им. Причината за този подход е, че документите трябва да бъдат подписани и вместо с цифрови подписи се извършва подписването им на хартиен носител. Столична община използва информационна система "Акстър" за управление на потока от документи, но когато е необходимо документите да се подпишат от повече от един служител на общината, а това се случва често, се налага те да бъдат физически пренасяни, което е крайно неефективно.

Поддържането на данните изцяло в цифров и максимално структуриран вид, без сканирани документи е ключово за подобряване на работата на общината. Абсолютно необходимият минимум е данните и документите да са така маркирани, че чувствителните данни да могат автоматично да бъдат заличени/анонимизирани при обмена на данни с други системи, в това число и при публикуване. Поддържането в цифров вид ще позволи драстично намаляване на съхраняваните обеми, необходимия ресурс за архивиране и възстановяване от аварии. В допълнение ще се даде възможност за пълнотекстово търсене, различни видове справки, автоматично изчисляване на индикатори и анализи включително и такива, използващи *machine learning*.

Задължително е да се въведе единен подход за генериране на уникални идентификатори за всеки набор от данни, напр. *UUID*.

Необходимо е да се доизградят и подобрят дигиталните канали за получаване на данни, в това число и в комуникацията с районните администрации, физически лица и фирми. Подаването на данни към Столична община в цифров вид ще минимизира ресурса, необходим за дигитализацията им и ще ускори процеса на работа.

Изграждането на програмни интерфейси към съществуващите системи ще даде възможност за по-добра координация на отделните звена, ще намали дублирането на данни и ще даде възможност за автоматизирано публикуване на данните в машинно-четими формати.

Предстоящите за изграждане системи следва да се координират с отдела по Информационни технологии за да се подсигури качеството им и необходимата функционалност за взаимодействие с останалите информационни системи на Столична община.

Препоръчително е изграждането на единна база данни за проверка за идентичност и авторизация на служителите. В момента всеки служител има различни акаунти в поне 3 различни системи на Столична община.

Първата стъпка към реалната дигитализация следва да е поддържане на документите, които общината издава в цифров и структуриран вид. Това може да стане с минимални разходи и за много кратко време като:

- Изготвяне на шаблони в MS Office за различните видове документи. Шаблоните изглеждат като формуляр и имат полета за попълване. Важно е да се стандартизират названията на едни и същи полета в различните видове документи, напр. име, фамилия и др. подобни. Това ще позволи разграничаването на отделните части на документа, което е ключово за последващите обработки – търсене, сортиране, агрегиране и т.н. Доброто маркиране ще позволи автоматично публикуване на данни в машинно-четим формат с възможности за премахване или анонимизиране на лични данни в публикуваните данни;
- Използване на цифрови подписи за подписване на изготвените документи. По този начин всичко остава в цифров вид, не се губи ресурс за физически пренос на документи, съкращава се времето за обработване на документи от повече от едно звено;
- Използване на софтуерни продукти, реализиращи функционалността “календар”, за дейности свързани с планиране на време, напр. графици за ремонти, график на културни и спортни събития и др. подобни. Изискване от експлоатационните дружества да поддържат и предоставят достъп до актуални календари, свързани с техните дейности на територията на Столична община.

Втората стъпка е интеграция на информационната система “Акстър” със система за пълнотекстово търсене, като целта е всеки завършен документ да бъде автоматично индексирани. Освен възможностите за анализ, системата реализира “виртуални регистри” на всички видове документи, които издава общината – всеки “виртуален регистър” е резултат от търсене на определена ключова дума в полето за тип на документ.

Третата стъпка е изготвянето на софтуерен компонент – регистър с общо предназначение. Освен документите, общината трябва да поддържа информация за различни типове обекти – собствени сгради, паркове, училища, детски градини и др. подобни. Описанието на повечето типове обекти може да се реализира в плоска структура – ред от таблица. Необходимо е да създаде софтуерен инструмент, който за всеки тип обект се конфигурира отделно, реализира делегиран достъп до отделните ресурси до ниво ред, предоставя API за достъп до данните.

Реалната дигитализация на Столична община ще подобри значително ефективността ѝ, а оттам и степента на удовлетвореност и доверието на гражданите в нея.

Повишаване на капацитета на служителите за работа с данни

Необходимо е да се повиши капацитета на служителите по отношение работата с данни в следните направления:

- Файлови формати – предимства, недостатъци и начините им на използване;

- Използване и предимства на цифровите подписи;
- Повторна употреба на данните;
- Следване на утвърдени от Столична община политики за работа с данни.

Препоръчително е обученията да се правят по звена като отразяват спецификата на работа с данни в конкретната област.

