

# Все про машиночитаний формат у публічних закупівлях

Жовтневі [зміни](#) до Закону України «Про публічні закупівлі» принесли обов'язковість звітування саме в машинозчитувальному форматі під час закупівель робіт, поточного ремонту. Тож розібратись із цим та дізнатись, які формати можуть вважатись машиночитаними, нам допоможе експерт у сфері аналізу та обробки даних Ренат Насрідінов.

**Радник.** Закон України «Про публічні закупівлі» оперує таким поняттям, як [«машинозчитувальний формат»](#). При цьому положення про набори даних, які підлягають оприлюдненню у формі відкритих даних, що затверджене постановою КМУ від 21.10.2015 № 835, надає [визначення поняттю «машиночитаний формат»](#). Як ви вважаєте, чи є ці терміни тотожними та такими, що говорять про одне й те саме? Що таке машиночитаний формат?

**Експерт у сфері аналізу та обробки даних.** Насамперед тут слід зазначити, що Закон № 3988-ІХ від 19.09.2024, що вносить необхідні зміни для забезпечення прозорості будівництва, дійсно додає вимогу оприлюднення у машинозчитувальному форматі, однак сам по собі не містить визначення, що це таке. Поняття «машиночитана форма» міститься у [наказі](#) Держагентства електронного урядування (відомство-попередник Мінцифри) від 2018 року — це «форма **подання даних**, придатних до **автоматичного або автоматизованого** використання».

Водночас визначення машиночитаного формату з положення, затвердженого постановою Кабміну № 835, про яке ви згадали, каже нам, що це **формат даних**, структурований таким чином, що дає змогу ідентифікувати, перетворювати та отримувати конкретні дані, включаючи окремі факти та їх внутрішню структуру, **без участі людини**.

Безумовно, ці визначення тотожні: в обох випадках йдеться про автоматичне використання даних (в першому — це вказано прямо, в другому — нам про це каже відсутність людини в цьому процесі), і в обох випадках йдеться про подання та про формат даних.

Щоб зрозуміти, що таке, власне, машиночитаний формат, ми маємо спочатку окреслити важливий аспект щодо різниці між інформацією та даними.

Спрощено: будь-яка інформація може розглядатися як оповідь, до якої включено показники, наприклад, коли ми кажемо, що «1 жовтня курс гривні до євро був 46,15», це інформація, до якої включено такі показники: дату (1 жовтня 2024 року), валюту (євро), курс (46,15). Тут фраза про перше жовтня буде інформацією, а показники, які я щойно перелічив, — даними. І це працює в усіх випадках: наприклад, преамбула будь-якого

договору містить назви (або імена) сторін, їхні адреси, коди (ЄДРПОУ, РНОКПП) тощо. Тут преамбула в цілому — це інформація, а назва, адреса, код — показники.

Машиночитаність у нашому випадку не означає, що певний файл має просто відкриватись машиною. **Цього недостатньо.** Технічно будь-який файл є машиночитаним: PDF, MS Word, MS Excel, інші файли відкриваються і зчитуються комп'ютером. Для цілей роботи з даними **машиночитаність означає насамперед отримання даних без участі людини, тобто без необхідності попередньо відкривати файл і якимось чином змінювати його.**

Гарним прикладом даних, що містяться у немашиночитаній інформації, може бути звіт про договір про закупівлю із Prozorro, наприклад [цей](#). Це охайний на вигляд документ, що містить усю базову інформацію про закупівлю. Однак він не є машиночитаним. У ньому фактично п'ять окремих таблиць, і щоб отримати дані про замовника, постачальника, ціну, дату договору тощо, треба додатково вивчати структуру документа й напрацьовувати логіку отримання даних.

Тобто для машиночитаності навіть документа з таблицями не завжди достатньо. На щастя, Prozorro надає усі ці дані у машиночитаному форматі, їх можна [переглянути ось тут](#).

Це так званий формат JSON. Виглядає він доволі незвично, але саме з нього можна отримати дані одразу, до того ж структура цих даних чітко документована, і той, хто з ними працюватиме, може не боятися, що щось наплутає. Більшість читачів, гадаю, знайомі із системою моніторингу закупівель Clarity Project. Так от, вона саме ці машиночитані дані й використовує.

Отже, підсумовуючи, **машиночитаний формат** — це будь-який формат, який дозволяє користувачу одразу отримати дані, без необхідності додаткових перетворень. Безумовно, забезпечення лише самої машиночитаності мало, оскільки значна кількість показників та складна структура потребують належного документування, як ми бачимо на прикладі Prozorro. Деякі машиночитані дані, втім, документування не потребують, оскільки говорять самі за себе, наприклад [курс валют НБУ](#).

**Радник.** Спробуймо розібратись, чи може вважатись машиночитаним документ, який створений за допомогою звичної для усіх програми [Microsoft Word](#) із розширенням файлу DOC / DOCX.

**Експерт у сфері аналізу та обробки даних.** Точно кажучи, ні. Якщо вдаватися у деталі, то формат DOC взагалі двійковий (binary), що унеможлиблює роботу з ним як зі структурованим форматом. **Щодо DOCX, то він, безумовно, виглядає структурованим, але не машиночитаним, тому що з нього отримується інформація, а не дані.**

Повертаючись до прикладу з курсом валют, навіть якби ми мали [ось цю таблицю](#) курсів валют в документі Word, потрібні були би додаткові зусилля, щоб отримати дані про курс валют з цього документа (принагідно зверну увагу, що Нацбанк не лише на тій самій

сторінці дає змогу завантажити максимально наближений до машиночитаного файл MS Excel з курсом валют, але й оприлюднює дані щодо курсу в машиночитаному форматі).

З форматом DOCX пов'язаний один неочевидний момент. MS Word дозволяє зберегти документ у форматі XML (Extensible Markup Language), це надзвичайно потужний формат для опису, передачі й контролю даних.

Окрім широкого застосування в IT-сфері він використовується і в інших: до прикладу, бухгалтерська та податкова звітність подається в цьому форматі, він широко використовується у фінансовій та банківській сферах.

Розпорядники, які оприлюднюють дані відповідно до постанови № 835, іноді бувають введені цим в оману, оскільки XML цією постановою визначений як один із форматів для оприлюднення даних, **вони просто зберігають документ DOCX як XML, але це не дані у форматі XML, це сам документ Word, описаний у форматі XML** — з купою службової інформації про відступи, форматування, перевірку орфографії тощо.

Ви можете у цьому впевнитися, завантаживши такий документ з порталу відкритих даних і відкривши його за допомогою Блокнота Windows або будь-якого іншого текстового редактора. Але на порталі можна знайти й дані, оприлюднені у форматі XML, і якщо порівняти ці два приклади, то стає зрозумілою відмінність.

**Радник.** *Добре, а якщо говорити про найрозповсюдженіший сканований документ, що має формат PDF, чи може вважатись такий документ машиночитаним? Кажуть, що до машиночитаного формату належить PDF, що був перетворений з Word. Яка Ваша думка щодо цього?*

**Експерт у сфері аналізу та обробки даних.** Це, на жаль, ще одна поширена помилка, і коротка відповідь на обидва запитання — однозначно ні, навіть якщо цей PDF-документ має текстовий шар, з якого можна скопіювати текст. Такий скопійований текст матиме усі недоліки, які має і документ Word, з якого його створили, плюс зламане форматування рядків і таблиць, тобто буде ще гірше, ніж було.

І якщо ми вже торкнулися цього питання, цілком аналогічна ситуація і з таблицями Microsoft Excel, які збережені в PDF. Безумовно, способи видобути таблицю з такого документа існують, але отримана таблиця, найімовірніше, буде брудними даними, насамперед через переноси рядків у PDF.

Тому якщо у вас є дані в табличному вигляді з якогось табличного процесора (це може бути не лише Microsoft Excel, а й OpenOffice, LibreOffice, Apple Keynote — будь-яка програма для створення таблиць), будь ласка, не треба робити з них PDF, це дуже ускладнює роботу з ними й може навіть призвести до ситуації «сміття на вході — сміття на виході».

**Радник.** *Можете навести перелік форматів, що можуть вважатись машиночитаними? Які, на Вашу думку, формати будуть найбільш доцільними для договору з додатками, які містять великі обсяги табличної інформації (кошторис, інші документи, документи, що містять інформацію про ціни на матеріальні ресурси)?*

**Експерт у сфері аналізу та обробки даних.** Оскільки перелік форматів для оприлюднення у нас вже затверджено урядовою постановою № 835, звернімося до неї, це пункт 9 Порядку. Одразу скажу, що з незрозумілих причин до переліку форматів включено неструктуровані формати (відео, аудіо, графічні дані), тому ці формати нас не цікавлять, **нас цікавлять формати структурованих даних, саме вони є машиночитаними.**

В нашому ж випадку (оприлюднення документів) підійдуть не всі там перелічені, а лише табличні формати: XLSX, ODS і, можливо, CSV. Чому я кажу можливо, бо CSV має ряд обмежень, які не дозволяють зберігати складні таблиці, наприклад кошториси на будівельні роботи.

Чому нас цікавлять не всі формати? В ідеалі для оприлюднення складної документації – договору з багатьма розділами, додатків до нього, якими є складні, часто ієрархічні таблиці, підійшов би формат XML, однак для цього потрібне спеціальне програмне забезпечення, на кшталт того, в якому бухгалтери готують звітність для подання у фонди та податкову (вони ж не набирають XML вручну), а до цього програмного забезпечення — шаблони форм і документів, схеми їх контролю тощо. На жаль, наразі нічого такого не існує, і до електронних договорів у будівництві, на мою думку, нам ще кілька років.

Тому давайте говорити про реалії. Навіть оприлюднення даних у структурованому форматі само по собі не дає гарантії машиночитаності. Існує кілька простих правил, дотримання яких гарантує, що ваші дані будуть машиночитані. Перш ніж перейти до них, нагадаю, що ми говоримо про табличні формати даних, а отже оперуємо такими поняттями, як рядок, стовпчик і комірка.

Отже, правила:

кожна комірка має містити одне і тільки одне значення. Тобто в комірці не має бути значення «1234,56 грн», має бути комірка, яка містить число 1234,56, та стовпчик з одиницями виміру;

об'єднаних комірок, рядків та/або стовпчиків в таблиці існувати не повинно (це стосується також і заголовків таблиці).

В ідеалі також таблиця не має містити підсумкових рядків, проміжних підсумків, будь-якої інформації до таблиці та після неї. Прикладом машиночитаних даних є [індекси цін у будівництві](#), що їх оприлюднює Держстат.

**Радник.** Просимо надати поради щодо того, як уповноваженій особі, так би мовити, оцифрувати в машиночитаний формат звичайний підписаний паперовий договір з усіма його невід'ємними додатками (в т. ч., наприклад, кошторисами й іншими відомостями про ціну на матеріалі ресурси). Чи потрібно використовувати онлайн-ресурси для конвертації документа, чи є інші шляхи?

**Експерт у сфері аналізу та обробки даних.** Стосовно будівельної документації, у тому числі договорів та додатків до них, то, враховуючи їхню складність, забезпечення дотримання навіть цих простих правил є доволі працемістким, і дуже ймовірно потребує розробки уніфікованої структури таблиць. Наразі важко сказати, чи візьме це на себе якийсь з ЦОВВ або регуляторних органів, тож маємо працювати з тим, що у нас є. Вбачаючи, що у нас є інформація (договір) та табличні дані до договору (додатки, кошториси, специфікації), я би пропонував такий підхід: **розділити інформацію та дані** і оприлюднювати інформацію у вигляді текстових документів (у форматах MS Word, OpenOffice), **а таблиці оприлюднювати як таблиці — у форматах MS Excel або OpenOffice**, тим більше що все ПЗ для роботи з кошторисами будівництва, яке є де-факто стандартом індустрії в Україні («АВК-5», «Кошторис XXI» тощо), здатне експортувати кошториси у формат MS Excel. Приклад того, як виглядатиме кошторис у форматі MS Excel:

№ Ч.ч	Номери кошторисів і кошторисних розрахунків	Найменування глав, будівель, споруд, лінійних об'єктів інженерно-транспортної інфраструктури, робіт і витрат	Кошторисна вартість, тис.грн.			
			будівельних робіт	устаткування, меблів та інвентарю	інших витрат	загальна вартість
1	2	3	4	5	6	7
<b>Глава 2. Об'єкти основного призначення</b>						
1	02-01	Реконструкція лабораторії	10020,398	13524,720	-	23545,118
2	02-02	Заміна ліфта	340,611	1443,151	-	1783,762
<b>Разом по главі 2:</b>			10361,009	14967,871	-	25328,880
<b>Разом по главах 1-7:</b>			10361,009	14967,871	-	25328,880
<b>Разом по главах 1-8:</b>			10361,009	14967,871	-	25328,880
<b>Разом по главах 1-9:</b>			10361,009	14967,871	-	25328,880
<b>Глава 10. Утримання служби замовника та інжинірингові послуги</b>						

Звісно, сам договір не буде в цьому випадку машиночитаним (і це наразі неможливо забезпечити в розумні терміни співставними зусиллями), однак це забезпечить доступність інформації. При цьому щодо договору про закупівлю та додатків до нього є також окрема вимога щодо розміщення, а саме в пункті 15 Порядку розміщення

інформації про публічні закупівлі, затвердженого наказом Мінекономіки від 11.06.2020 № 1082, встановлено, що договір про закупівлю, додатки до договору про закупівлю, зміни до договору про закупівлю розміщуються замовником в електронній системі закупівель у форматі PDF. Отже, допоки не будуть внесені відповідні зміни в закупівельне законодавство, радимо оприлюднювати документи в декількох форматах.

І останнє, але не менш важливе: разом з документами у форматах MS Word, MS Excel або OpenOffice доцільно також оприлюднювати файли кошторисів у форматі обміну кошторисною документацією (IBS, IMS та ін.), раз вже ми говоримо про прозорість та підзвітність. Оприлюднення цих файлів дасть змогу швидко та всебічно дослідити кошторис будівництва.

### ***Радник. Щиро вдячні за детальні роз'яснення, які наразі дуже цінні та актуальні у сфері публічних закупівель!***

Статті та зразки документів за тематикою інтерв'ю на порталі «Радник у сфері публічних закупівель»:

*«Замовників зобов'язали публікувати інформацію про ціни на матеріальні ресурси: зміни в закупівлі послуг з поточного ремонту та робіт з будівництва»;*

*Шпаргалка щодо реєстрації зобов'язань за договорами з урахуванням вимог щодо оприлюднення інформації про ціни на матеріальні ресурси;*

*«Інформація про ціни на матеріальні ресурси: зразок для закупівель з 23.10.2024»;*

*«Три питання-відповіді про прозоре будівництво».*