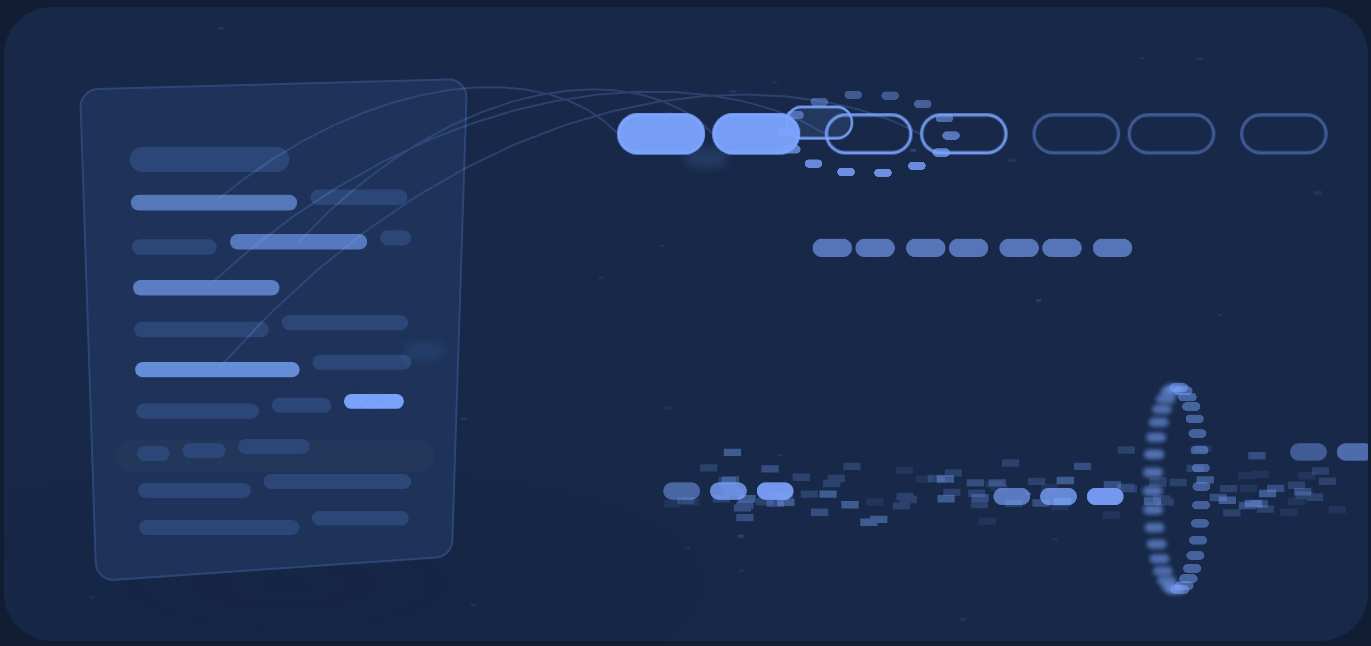
Rillor

METHODS NOTE

Building an obligations register from regulation text

A method for reading regulation text into a compliance obligations register: one row per obligation and party, with modality, binding status, exceptions, dates and the cited version.



8 November 2026 Document and regulation intelligence rillor.com/insights/obligations-register-from-regulation-text

Contents

01 Summary

02 Background

03 Method

04 Findings

05 Implications

06 Limits of this analysis

07 Sources

Summary

Published standards and drafting guides give each step of reading regulation text into a compliance obligations register a fixed field, so that a second reader can reproduce each row and a reviewer can check it.

- Each register row holds one obligation for one party. LegalRuleML, an OASIS standard, defines an obligation as an act or state a party is legally bound to achieve, whose failure is a violation, and sets it apart from prohibitions, permissions and rights [1]. In GDPR Article 33, the controller's duty to notify the supervisory authority and the processor's duty to notify the controller are two rows [2].
- Words that signal a duty need a stated convention. In English, binding EU acts generally use "shall" in their enacting terms [3], UK drafting policy avoids it [4], and U.S. House drafting treats "may not" as a prohibition although "may" alone grants a permission [5].
- Whether a provision binds is a field of its own. The UK Financial Conduct Authority's Reader's Guide, dated January 2019, gives each provision of its Handbook a status, among them rule, evidential provision, guidance and direction, each with a different legal effect [6], and EU recitals, the reasons stated before an act's articles, must not contain normative provisions [3].
- Exceptions are rules that override other rules. LegalRuleML models an exception as a more specific rule that defeats a general, defeasible one [1], and the SEC's rule at 17 CFR 248.30 sets a 30-day customer notice that Attorney General delays of up to 30, 30 and 60 days can push back [7].
- A row needs more than one date. The GDPR entered into force on 24 May 2016 (derived) and applies from 25 May 2018 [2], and the SEC's 2024 Regulation S-P amendments took effect on 2 August 2024, with compliance dates of 3 December 2025 and 3 June 2026 by entity size (derived) [8].
- Each row cites the version of the text it was read from. The Electronic Code of Federal Regulations (eCFR) lists four versions of 17 CFR 248.30 [9], and the Council of the EU's mandate of 21 October 2026 to set 96 hours for GDPR breaches likely to result in a high risk sits beside the row as a proposed change [10, 11].
- Rows extracted by a model need a completeness check. In one 2025 study, LLM formalizations of a consumer code labeled every element's modality correctly while recall of rules peaked at 82.69% [12], and a model-based completeness metric scored answers merely copied from source passages at 0.947, above readable generated answers at 0.639 [13].

Background

What a register lists. LegalRuleML defines a legal norm as "a binding directive from a Legal Authority to addressees" [1]. An obligations register lists those directives one row per obligation: the party bound, what it must do or avoid, what sets the duty off, what excuses it and the dates that govern it.

What supervisors describe. The CFPB's examination manual reviews an institution's change management, with the objective to "respond promptly to changes in applicable Federal consumer financial laws", and expects policies and procedures to be "maintained and modified to remain current and complete" [14]. The FCA's Handbook "presents changes in a single, consolidated view" [15]. Neither source names an obligations register. Both describe the work one supports: knowing which rules apply and noticing when they change.

Why the reading needs rules. Careful readers can turn the same text into different rows. For three subsections of the Australian Telecommunications Consumer Protections Code, Horner and colleagues counted 69 terms and 52 rules where an earlier study had counted 65 and 36 [12]. Written reading rules let a second reader reproduce a row and a reviewer check it.

Who this note is for. Compliance and legal operations staff who build or audit a register, and engineers who build its extraction.

Method

What was examined. Standards for legal text [1, 16, 17], two IETF conventions used only as an analogy [18, 19], the FCA's status icons [6, 15], drafting guides from the EU, the UK and the U.S. [3, 4, 5, 20, 21] and one supervisory manual [14]. Two breach-notice duties were read through the method: GDPR Article 33 [2, 10, 11, 22] and the customer notice in 17 CFR 248.30 [7, 8, 9, 23, 24]. Nine research papers from 2020 to 2026 on finding, formalizing and checking obligations complete the set [12, 13, 25, 26, 27, 28, 29, 30, 31].

Sources and how. Thirty-one sources, each opened on 7 November 2026. Each finding gives a field's definition from its standard or guide and applies it to the two provisions. "Derived" marks a date or figure counted from a source's own numbers. Regulation text is quoted only in the short spans a row needs, model names in the papers are left out, and no product or vendor is named.

Where sources disagree. The EU guide says English generally uses "shall" in binding acts [3], and the UK's drafting office avoids it [4]. Horner and colleagues recompute an earlier study's reported recall of 90.78% as 75.38% from its own numbers [12].

Findings

1. Modality: obligation, permission, prohibition and right

Four modalities. LegalRuleML's glossary separates four [1]:

Modality	LegalRuleML definition, paraphrased	Test question for a row
Obligation	A state, act or course of action the bearer is legally bound to, whose failure is a violation	Is the party in breach if it does nothing?
Prohibition	The same form, with a violation if the state or act is achieved or performed	Is the party in breach if it acts?
Permission	The bearer has no obligation or prohibition to the contrary	Does the text free the party from a duty or a ban?
Right	A permission for the bearer that implies obligations or prohibitions on other parties	Does another party owe the bearer something?

Source: [1]. The test questions are this note's.

Weak and strong permission. A weak permission is the absence of a contrary duty. A strong permission is "an exception or derogation" of one, as a "U-turn permitted" sign derogates a general prohibition [1]. Strong permissions change what another row requires, so each gets a row of its own. GDPR Article 33(4) is one: information "may be provided in phases without undue further delay", which relaxes the content duty in Article 33(3) [2].

Achievement and maintenance. An achievement obligation is fulfilled if it holds at least once in its period, and a maintenance obligation "must be obeyed for all the instants when it holds" [1]. The notice in Article 33(1) is an achievement obligation [1, 2].

What follows a missed duty. In LegalRuleML "a violation does not result in an inconsistency", and a suborder list states what holds once a first obligation is violated, as with penalties and reparations [1]. The second sentence of Article 33(1) is a linked duty with its own trigger: any notification not made within 72 hours "shall be accompanied by reasons for the delay" [2]. It applies whether or not the delay breached the first sentence, since the 72 hours apply only "where feasible" [2], so it is no reparation. It is a row of its own, set off by the 72 hours passing.

Modality belongs to a party. LegalRuleML's deontic operators "have parties (e.g. Bearer)" [1], and the LEXDEMOD corpus annotates contracts with modality "expressed with respect to a contracting party or agent" [30]. One sentence can oblige one party and permit another, so the register keeps one row per

obligation per party.

2. Words that signal modality, and where they mislead

An engineering analogy. RFC 2119 defines MUST and SHALL as "an absolute requirement of the specification", and MAY as "truly optional" [18]. RFC 8174 added that "only UPPERCASE usage of the key words have the defined special meanings" and that "a lot of normative text does not use them and is still normative" [19]. These conventions govern protocol documents and carry no legal authority. Two lessons carry over: a register states its convention, and a provision with no key word can still bind.

Legislatures differ. EU enacting terms generally use "shall" [3]. The UK's Office of the Parliamentary Counsel has a policy "to avoid the use of the legislative 'shall'" [4]. The U.S. House Legislative Counsel reads "shall" as required and "may" as permitted, and calls "may not" mandatory and "the preferred language for denying a right, power, or privilege" [5]. Garner and Kimble read must as "is required to" and "may ... only" as "conditional permission", and advise replacing shall [20].

Phrase	Usual reading	Where it misleads	Basis
shall	Obligation in EU enacting terms	UK bills avoid it, so its absence says nothing	[3, 4]
must	Obligation	Text with no such word can still be normative	[19, 20]
may	Permission	"may not" is a prohibition	[5, 20]
may ... only	Conditional permission	The same rule reads as a prohibition with an exception	[20]
should	Directory provision	In an EU recital it has no normative force	[3, 20]

Plural subjects. The House guide rewrites the ambiguous "Drivers may not run red lights." as "A driver may not run a red light." [5], so a row names its bearer in the singular.

Recital against article. Recital 85 says the controller "should notify" within 72 hours "unless the controller is able to demonstrate" that the breach is unlikely to result in a risk [2]. Article 33(1) says "shall ... notify", with an exception that carries no burden phrase [2]. Recitals "shall not contain normative provisions" [3]. A reader keyed on modal words would add a second, weaker duty from the recital. The register reads the article and links the recital as explanation.

3. Binding status: rules, evidential provisions, guidance and directions

The FCA's status icons. The FCA's Reader's Guide, dated January 2019, gives each Handbook provision a status icon. The four a register uses most are [6]:

Icon	Status	What the guide says it does
R	Rule	Creates binding obligations on firms in most cases, and contravention can bring enforcement action and action for damages
E	Evidential provision	Not binding in its own right and tied to another binding rule. Compliance with it can tend to establish compliance with that rule, and breach can tend to establish contravention
G	Guidance	Not binding, and need not be followed. Acting in line with general guidance in the circumstances it contemplates is treated as compliance with the rule it relates to
D	Direction	Binding on the persons or categories of person it is addressed to

Source: [6]. Notes "provide informative, but non-legislative, material" [6].

EU acts. The GDPR is "binding in its entirety and directly applicable in all Member States" [2].

Guidance as a linked source. The EDPB's Guidelines 9/2022, version 2.0 of 28 March 2023, read the trigger word "aware" as "a reasonable degree of certainty that a security incident has occurred that has led to personal data being compromised" [22]. In their first example, a lost USB key holding unencrypted data has to be notified, and the controller becomes aware "when it realised the USB key had been lost" [22]. The register links this passage to the trigger field as a separate source with its own status and version.

4. Bearer, trigger and required action

Bearer and auxiliary party. LegalRuleML's bearer is the role "to which the Deontic Specification is primarily directed", and an auxiliary party a role to which it is related [1]. In Article 33(1) the bearer is the controller and the auxiliary party the supervisory authority competent under Article 55. Article 33(2) gives the processor a duty to notify the controller "without undue delay after becoming aware of a personal data breach" [2].

Trigger. The trigger is the condition that sets the duty off, with each defined term linked. Article 33(1) applies "in the case of a personal data breach", which Article 4(12) defines as a breach of security leading to accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data [2]. The clock starts on "having become aware of it" [2]. Under 17 CFR 248.30(a)(4)(iii) it starts on becoming aware that unauthorized access to or use of customer information "has occurred or is reasonably likely to have occurred" [7].

Action and content. In Article 33(1) the action is to notify the breach to the supervisory authority. Article 33(3) says the notification "shall at least" describe the breach, give a contact point, and describe the likely consequences and the measures taken or proposed [2]. Each content duty is a linked row.

Segmenting the text. 1 U.S.C. 104, quoted in the House guide, asks that a section contain "as nearly as may be, a single proposition of enactment" [5]. Regulation text often packs several. Horner and colleagues split long articles into "law snippets" and told the model to split enumerations of more than two elements [12].

Field lists from research. Sleimi and colleagues define six metadata types for legal statements and 18 for their phrases, among them agent, action, condition, exception, modality, sanction and time, with precision of 97.2% and 82.4% and recall of 94.9% and 92.4% on two case studies over Luxembourgish legislation [26]. Singhal and Breaux turned 13 U.S. state data breach notification laws into representations "encodable and executable as Python code" that passed about 89.4% of test cases, with precision of 82.2 and recall of 88.7 [27].

5. Exceptions and overrides as defeasible rules

Defeasible rules. LegalRuleML states that normative conditionals "are generally defeasible", and that "one application of defeasible reasoning is the ability to model exceptions in a simple and natural way" [1]. An exception is a more specific rule with the opposite conclusion, which blocks the general one. Rules are strict, defeasible or defeaters, and an override is "an indication that a Legal Rule takes precedence over another Legal Rule" [1].

One exception. Article 33(1) has one: "unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons" [2]. The register stores it as an exception row that overrides the notice row, with its phrase linked.

A chain of overrides. 17 CFR 248.30(a)(4)(iii) reads into five rows [7]:

Row	Modality	Condition	Limit
Customer notice	Obligation	Awareness of unauthorized access to or use of customer information, actual or reasonably likely	As soon as practicable, not later than 30 days
First delay	Permission, overrides the notice deadline	The Attorney General determines that notice poses a substantial risk to national security or public safety and notifies the Commission in writing	A period the Attorney General specifies, up to 30 days
Further delay	Permission	The Attorney General determines the risk continues and notifies the Commission in writing	Up to 30 days
Final delay	Permission	Extraordinary circumstances, and the Attorney General determines a continuing risk to national security and notifies the Commission in writing	A final additional period of up to 60 days

Row	Modality	Condition	Limit
Commission delay	Permission	The Attorney General indicates further delay is necessary, and the Commission grants it by exemptive order or other action	None stated

Source: [7]. The three Attorney General delays add up to 120 days (derived), after which the Commission may grant more [7]. Each is a strong permission in LegalRuleML's sense, a derogation from the deadline [1]. A register that keeps only the first row misstates the deadline when the conditions hold, and one that keeps only the delays has lost the duty they qualify.

Missing parent rules. Horner and colleagues check completeness first. In their example, a permission to close a complaint with consent or under clauses 8.2.1(c) to (e) came out as four permissions and missed `complaint(X) => [O] -closeComplaint(X)`, an obligation not to close a complaint [12]. Read as a prohibition, it is the rule the four permissions are exceptions to. A register runs the same test: every permission and exception row names the row it overrides.

6. Dates in force, efficacy and applicability

Three axes. LegalRuleML says norms are valid "with respect to three main legal axes": entry into force, efficacy and applicability [1]. Those dates belong to the norm. Times inside its content, such as when a person is to present a tax application, are content [1], so the 72-hour deadline goes in the deadline field.

EU dates. The EU guide separates "the date of entry into force and the date from which provisions are to have effect", and notes that the date of application may differ from the date of entry into force [3]. Without a stated date, an act enters into force "on the twentieth day following that of their publication" [3]. GDPR Article 99 uses that rule and adds "It shall apply from 25 May 2018" [2]. Published on 4 May 2016, the GDPR entered into force on 24 May 2016 (derived) [2].

U.S. dates. The Office of the Federal Register's drafting handbook ties the effective date to the amendment of the CFR and says "effective dates cannot be retroactive" [21]. A compliance or applicability date, where a rule has one, "is the date that the affected classes must comply with the rule" [21]. The 2024 Regulation S-P amendments were effective on 2 August 2024, with compliance periods of 18 months after publication for larger entities and 24 months for smaller ones [8].

Date field	GDPR Article 33	17 CFR 248.30(a)(4)(iii)
Published	4 May 2016, OJ L 119 [2]	3 June 2024, 89 FR 47688 [8]
In force or effective	24 May 2016 (derived) [2]	2 August 2024 [8]

Date field	GDPR Article 33	17 CFR 248.30(a)(4)(iii)
Applies or compliance	25 May 2018 [2]	3 December 2025 for larger entities, 3 June 2026 for smaller ones (derived) [8]

A row that did not exist. The eCFR shows 17 CFR 248.30 as of 1 August 2024 under its old title, with no 30-day notice paragraph [23]. A register read that day had no row for the duty, and one read on 2 August had a row whose first compliance date was about 16 months away (derived) [8].

7. Stable citations and versions of the provision

Work and expression. Akoma Ntoso, an OASIS standard of 2018, is "meant to make the structure and meaning of legal documents machine readable", which supports point-in-time consolidation of legislation [16]. It separates the WORK, "the abstract concept of the legal resource", the EXPRESSION, a version with specified content, the MANIFESTATION, a format such as XML or PDF, and the ITEM, a stored copy [16]. A row cites the expression: the act, the provision and the dated version read.

Identifiers that survive renumbering. Each life-cycle event carries a date and the document that caused it [16]. The original attribute "stores the first wId, that never evolves", and the current attribute stores the identifier valid in a given period [16]. A row keeps both.

ELI. The European Legislation Identifier gives legislation across Europe a unique identifier and a common data model for its metadata [17]. The GDPR's is data.europa.eu/eli/reg/2016/679/oj [2].

U.S. versions. The eCFR page for 17 CFR 248.30 reads "up to date as of 11/05/2026" and says the eCFR "is not an official legal edition of the CFR" [7]. Its versioner lists four versions: 31 December 2016, the same date reissued on 30 March 2023 with no substantive change, 3 June 2024 and 2 August 2024, when the section was renamed [9]. The point-in-time system shows changes since 2017 and compares content on two dates [24]. A new version is the signal to re-read the rows tied to the section.

Text and rules stay tied. LegalRuleML asks for "a one-to-one correspondence" between rules and the text units that express them, because "the legal text is the only legally binding element" and rules "need to be updated according to the textual changes" [1].

Proposed changes. The Commission's Digital Omnibus proposal, COM(2025) 837, was published on 19 November 2025 [11]. The Council's negotiating mandate, agreed by the Permanent Representatives Committee on 21 October 2026, would limit Article 33(1) to "a personal data breach that is likely to result in a high risk to the rights and freedoms of natural persons", notified to the supervisory authority "not later than 96 hours after having become aware of it" "with the support of the national entry point" set up under Directive (EU) 2022/2555, and directly to the supervisory authority until that entry point exists [10]. On 7

November 2026 the Parliament's procedure file read "Awaiting committee decision" [11]. The mandate's wording changes the trigger and the deadline, and adds the national entry point's support [2, 10]. The row carries it as a proposed change under 2025/0360(COD), and the 2016 text stays the rule in force.

8. Checking LLM-extracted obligations for completeness and contradiction

A benchmark built on obligations. ObliQA holds "27,869 questions" drawn from 40 Abu Dhabi Global Markets (ADGM) financial regulation documents totaling 640,000 words [25]. Its metric, RePASs, "tests whether generated answers accurately capture all relevant obligations while avoiding contradictions": a classifier finds obligation sentences, and an inference model counts an obligation as covered at an entailment score above 0.7 [25]. Gold answers scored 0.8101 with full coverage. A baseline of lexical retrieval and a commercial LLM covered 0.214 to 0.222 and scored 0.466 to 0.470 [25].

A metric can be gamed. In the RIRAG 2025 shared task, answers copied from retrieved passages reached "a dubiously high score (0.947)" by exploiting the metric's obligation-extraction step, while readable generated answers scored 0.639 [13]. A coverage score computed by a model needs a check against a labeled list of obligations.

Modality right, rules missed. In Horner and colleagues' study every element had the right modality label, against 47 of 49 (95.92%) in the earlier study, while precision on rules peaked at 87.18% and recall at 82.69% across model families and settings [12]. With seed and temperature fixed, they still "encountered non-deterministic behavior" [12].

Sentence-level extraction. EURO-5K holds sentence-level reporting obligations and challenging negative examples from 136 EU legislative acts. Fine-tuned encoders reached 0.89 F1, fine-tuned LLMs matched them, and "all approaches converge around 3K samples" [28].

Atomic requirements. ARCCS "decomposes raw regulatory text into atomic, traceable requirements" and reached 98.8% accuracy in violation detection on more than 1,200 EU public-procurement rule checks [29]. Its GDPR figure, consistency in up to 96.67% of cases, was graded by LLM judges [29].

Reasoning over rules. On DeonticBench's 6,232 tasks in U.S. federal taxes, airline baggage policies, immigration administration and state housing law, "best hard-subset performance reaches only 44.4% on SARA Numeric and 46.6 macro-F1 on Housing" [31].

Checks a register can run.

1. Coverage: compare extracted rows with a labeled list of obligations for a sample of provisions, and report recall [12, 13, 28].
2. Parents: every permission and exception row names the row it overrides [1, 12].
3. Phrase in passage: each field's linked phrase appears in the cited version of the text [1].
4. Contradiction: rows from one provision are compared for conflicting modality or deadlines [25].

5. Stability: extraction runs twice, and differences go to review [12].
6. Review: a person accepts, corrects or rejects each row, since the hard-subset scores above stay under 50% [31].

9. A worked provision, read into one row

GDPR Article 33(1), first sentence:

Field	Value	Basis
Citation	Regulation (EU) 2016/679, Article 33(1), first sentence	[2, 17]
Source version	As published in OJ L 119, 4 May 2016	[2]
Binding status	Enacting term of a regulation, binding and directly applicable. Recital 85 is linked as explanation	[2, 3]
Bearer	Controller	[2]
Auxiliary party	Supervisory authority competent under Article 55	[2]
Modality	Obligation ("shall"), achievement	[1, 2, 3]
Trigger	A personal data breach (Article 4(12)) and the controller "having become aware of it"	[2]
Action	Notify the breach to the supervisory authority	[2]
Deadline	Without undue delay and, where feasible, not later than 72 hours after becoming aware	[2]
Exception	The breach is unlikely to result in a risk to the rights and freedoms of natural persons	[2]
Linked rows	33(1) second sentence, 33(2), 33(3), 33(4), 33(5)	[2]
In force, applies	24 May 2016 (derived), 25 May 2018	[2, 3]
Guidance	EDPB Guidelines 9/2022 v2.0 on "aware"	[22]

Field	Value	Basis
Proposed change	Council mandate of 21 October 2026: high-risk breaches, 96 hours. Parliament: awaiting committee decision on 7 November 2026	[10, 11]
Review	Reviewer, time and source version	This note

Paragraph 2 binds the processor, paragraph 3 sets the content, paragraph 4 permits information in phases, and paragraph 5 obliges the controller to document any breach, with the facts, its effects and the remedial action taken [2]. Read the same way, 17 CFR 248.30(a)(4)(iii) gives one obligation row and four permission rows, with one effective date and two compliance dates [7, 8].

10. Register fields, in order

Order	Field	What it holds	Basis
1	Row ID	An identifier that stays with the row across versions	[16]
2	Citation	Act, provision and sentence, with a standard identifier	[16, 17]
3	Source version	The dated version read, and its identifier	[9, 16]
4	Source phrase	The exact words each field came from	[1]
5	Binding status	Rule, evidential provision, guidance, direction, recital or note	[3, 6]
6	Bearer	The party the provision is directed at, in the singular	[1, 5]
7	Auxiliary party	Any other party the provision involves	[1]
8	Modality	Obligation, prohibition, permission or right, and achievement or maintenance	[1]
9	Modality phrase	The words read as the modality, and the convention applied	[18, 19, 20]
10	Trigger	The condition, with each defined term linked to its definition	[2, 26]

Order	Field	What it holds	Basis
11	Action	What the bearer must do or avoid	[26]
12	Deadline	Time inside the provision's content	[1]
13	Exceptions and overrides	Rows that defeat or qualify this one, with priority	[1]
14	Linked rows	Other rows of the same provision, including what follows a missed duty	[1, 2]
15	Dates	In force and application, or effective and compliance	[1, 3, 21]
16	Guidance	Interpretive sources, each with its own status and version	[6, 22]
17	Proposed changes	Text under negotiation, its procedure reference and the date its status was read	[10, 11]
18	Review	Reviewer, time and source version of the decision	[12]

Implications

For compliance and legal operations teams.

1. Write down the reading convention for each rulebook [3, 4, 5, 20].
2. Keep one row per obligation per party [1, 30].
3. Fill binding status apart from modality, and link guidance and recitals as sources [3, 6, 22].
4. Give every permission and exception row a parent row [1, 12].
5. Keep entry into force, application, effective and compliance dates apart [1, 3, 21].
6. Cite the dated version read, and re-read rows when a new version appears [9, 16, 24].
7. Carry proposed changes with their procedure reference, apart from the rule in force [10, 11].

For engineers building the extraction. Store each row with its passage, version identifier, model, prompt and run date, since runs can differ with fixed settings [12]. Measure recall against a labeled set, and treat a coverage score computed by a model as one signal among several [13].

For reviewers. Review new rows, rows whose passage changed, permissions with no parent row and rows that rest on guidance.

Rillor reads obligations registers from the regulation text, one row per obligation, and every row links to the passage it came from. When a rule changes, rows tied to a changed passage go back to review, and each row keeps its source version and every change since.

Limits of this analysis

- This note describes how to read provisions into fields. It does not say what any law requires of any firm, it is not legal advice, and no source reviewed requires a register by that name.
- Both worked provisions are breach notices. Provisions on capital, conduct or reporting may stress other fields.
- The FCA's Reader's Guide is dated January 2019 [6], and a new FCA Handbook site launched on 8 September 2025 [15]. The guide is used for the status icons only.
- RFC 2119 and RFC 8174 are engineering conventions, used here as an analogy only [18, 19].
- The Article 33 amendment is a negotiating mandate as of 7 November 2026 [10, 11]. In its marks, the high-risk trigger and the 96 hours are the Commission's wording kept by the Council, and the national entry point's support is the Council's [10].
- Research figures belong to their own corpora of regulation, codes, statutes and contracts [12, 13, 25, 26, 27, 28, 29, 30, 31]. None measures a full register or the people who review one.

Sources

1. LegalRuleML Core Specification Version 1.0. OASIS Standard, OASIS, 30 August 2021, viewed 7 November 2026. https://docs.oasis-open.org/legalruleml/legalruleml-core-spec/v1.0/legalruleml-core-spec-v1.0.html

2. Regulation (EU) 2016/679 (General Data Protection Regulation). European Parliament and Council of the European Union, 27 April 2016, Official Journal L 119, 4 May 2016, viewed 7 November 2026. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

3. Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation, second edition. Publications Office of the European Union, 2015, viewed 7 November 2026. https://op.europa.eu/en/publication-detail/-/publication/3879747d-7a3c-411b-a3a0-55c14e2ba732

4. Drafting guidance, 19 March 2024. Office of the Parliamentary Counsel, GOV.UK, publication page updated 4 April 2024, viewed 7 November 2026. https://www.gov.uk/government/publications/drafting-bills-for-parliament/2024-03-19-drafting-guidance

5. HOLC Guide to Legislative Drafting. Office of the Legislative Counsel, U.S. House of Representatives, viewed 7 November 2026. https://legcounsel.house.gov/holc-guide-legislative-drafting

6. Reader's Guide: An introduction to the Handbook. Financial Conduct Authority, January 2019, viewed 7 November 2026. https://www.fca.org.uk/publication/handbook/readers-guide.pdf

7. 17 CFR 248.30, Electronic Code of Federal Regulations. Office of the Federal Register and Government Publishing Office, up to date as of 5 November 2026, viewed 7 November 2026. https://www.ecfr.gov/current/title-17/section-248.30

8. Regulation S-P: Privacy of Consumer Financial Information and Safeguarding Customer Information, final rule, 89 FR 47688, FR Doc. 2024-11116. Securities and Exchange Commission, Federal Register, 3 June 2024, viewed 7 November 2026. https://www.federalregister.gov/documents/2024/06/03/2024-11116/regulation-s-p-privacy-of-consumer-financial-information-and-safeguarding-customer-information

9. eCFR versioner API, versions of 17 CFR 248.30. Office of the Federal Register, viewed 7 November 2026. https://www.ecfr.gov/api/versioner/v1/versions/title-17.json?section=248.30

10. Digital Omnibus: mandate for negotiations with the European Parliament, Council document 14756/26, interinstitutional file 2025/0360 (COD). Council of the European Union, General Secretariat, 22 October 2026, viewed 7 November 2026. https://data.consilium.europa.eu/doc/document/ST-14756-2026-INIT/en/pdf

11. Procedure file 2025/0360(COD), Legislative Observatory. European Parliament, viewed 7 November 2026. https://oeil.secure.europarl.europa.eu/oeil/en/procedure-file?reference=2025/0360(COD)

12. Toward Robust Legal Text Formalization into Defeasible Deontic Logic using LLMs. Elias Horner, Cristinel Mateis, Guido Governatori and Agata Ciabattoni, arXiv 2506.08899, version 3, 31 December 2025, viewed 7 November 2026. https://arxiv.org/abs/2506.08899

13. AUEB-Archimedes at RIRAG-2025: Is obligation concatenation really all you need? Ioannis Chasandras, Odysseas S. Chlapanis and Ion Androutsopoulos, RIRAG 2025 shared task, RegNLP workshop at COLING 2025, arXiv 2412.11567, 16 December 2024, viewed 7 November 2026. https://arxiv.org/abs/2412.11567

14. Compliance Management Review, CFPB Supervision and Examination Manual. Consumer Financial Protection Bureau, August 2017, viewed 7 November 2026. https://files.consumerfinance.gov/f/documents/201708_cfpb_compliance-management-review_supervision-and-examination-manual.pdf

15. Handbook of rules and guidance. Financial Conduct Authority, first published 20 April 2016, last updated 21 September 2026, viewed 7 November 2026. https://www.fca.org.uk/about/how-we-regulate/handbook

16. Akoma Ntoso Version 1.0 Part 1: XML Vocabulary. OASIS Standard, OASIS, 29 August 2018, viewed 7 November 2026. https://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0-part1-vocabulary.html

17. European Legislation Identifier (ELI), EU Vocabularies. Publications Office of the European Union, viewed 7 November 2026. https://op.europa.eu/en/web/eu-vocabularies/eli

18. RFC 2119: Key words for use in RFCs to Indicate Requirement Levels. S. Bradner, Internet Engineering Task Force, BCP 14, March 1997, viewed 7 November 2026. https://www.rfc-editor.org/rfc/rfc2119

19. RFC 8174: Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words. B. Leiba, Internet Engineering Task Force, BCP 14, May 2017, viewed 7 November 2026. https://www.rfc-editor.org/rfc/rfc8174

20. Essentials for Drafting Clear Legal Rules. Bryan A. Garner and Joseph Kimble, hosted by the United States Courts, 2024, viewed 7 November 2026. https://www.uscourts.gov/sites/default/files/essentials_for_drafting_clear_legal_rules_2024.pdf

21. Document Drafting Handbook, August 2018 Edition, Revision 2.2. Office of the Federal Register, National Archives, June 2025, viewed 7 November 2026. https://www.archives.gov/files/federal-register/write/handbook/ddh.pdf

22. Guidelines 9/2022 on personal data breach notification under GDPR, Version 2.0. European Data Protection Board, adopted 28 March 2023, viewed 7 November 2026. https://www.edpb.europa.eu/system/files/2023-04/edpb_guidelines_202209_personal_data_breach_notification_v2.0_en.pdf

23. 17 CFR 248.30 as of 1 August 2024, eCFR point-in-time view. Office of the Federal Register and Government Publishing Office, viewed 7 November 2026. https://www.ecfr.gov/on/2024-08-01/title-17/section-248.30

24. Using the eCFR Point-in-Time System, eCFR Reader Aids. Office of the Federal Register, viewed 7 November 2026. https://www.ecfr.gov/reader-aids/using-ecfr/ecfr-changes-through-time

25. RIRAG: Regulatory Information Retrieval and Answer Generation. Tuba Gokhan, Kexin Wang, Iryna Gurevych and Ted Briscoe, arXiv 2409.05677, version 2, 2 December 2024, viewed 7 November 2026. https://arxiv.org/abs/2409.05677

26. An Automated Framework for the Extraction of Semantic Legal Metadata from Legal Texts. Amin Sleimi, Nicolas Sannier, Mehrdad Sabetzadeh, Lionel Briand, Marcello Ceci and John Dann, arXiv 2001.11245, 30 January 2020, viewed 7 November 2026. https://arxiv.org/abs/2001.11245

27. Legal Requirements Translation from Law. Anmol Singhal and Travis Breaux, IEEE RE 2025, arXiv 2507.02846, 3 July 2025, viewed 7 November 2026. https://arxiv.org/abs/2507.02846

28. EURO-5K: When Does Domain Pretraining Matter? Benchmarking Transformers for EU Reporting Obligation Extraction. Marios Koniaris, Vasileios Kotronis, Eugenia Giannini and Panayiotis Tsanakas, arXiv 2606.02971, 2 June 2026, viewed 7 November 2026. https://arxiv.org/abs/2606.02971

29. ARCCS: An Automated Regulatory Compliance Checking System. Giorgos Filandrianos, José Menezes, Chrysoula Zerva and Alessandro Gianola, extended version of an EMNLP 2026 system demonstration, arXiv 2610.01345, 1 October 2026, viewed 7 November 2026. https://arxiv.org/abs/2610.01345

30. Agent-Specific Deontic Modality Detection in Legal Language. Abhilasha Sancheti, Aparna Garimella, Balaji Vasan Srinivasan and Rachel Rudinger, EMNLP 2022, arXiv 2211.12752, 23 November 2022, viewed 7 November 2026. https://arxiv.org/abs/2211.12752

31. DeonticBench: A Benchmark for Reasoning over Rules. Guangyao Dou and others, arXiv 2604.04443, 6 April 2026, viewed 7 November 2026. https://arxiv.org/abs/2604.04443

About this report

Published by Rillor on 8 November 2026. Superintelligence (SI) tools assist research and drafting, and every figure cites its source. Rillor builds and secures SI systems, supplies the data they learn from and delivers the infrastructure they run on. We apply SI to financial and compute markets.

Cite as: Rillor. Building an obligations register from regulation text. 8 November 2026. https://rillor.com/insights/obligations-register-from-regulation-text

Notices

Superintelligence (SI) outputs can be wrong. Important decisions should include human review.

Work with Rillor

Send a sample of the rulebooks, regulations, contracts or filings you work from and the fields you need from them. Contact us at rillor.com/contact?type=document-intelligence or contact@rillor.com. More on document and regulation intelligence: rillor.com/si/document-intelligence