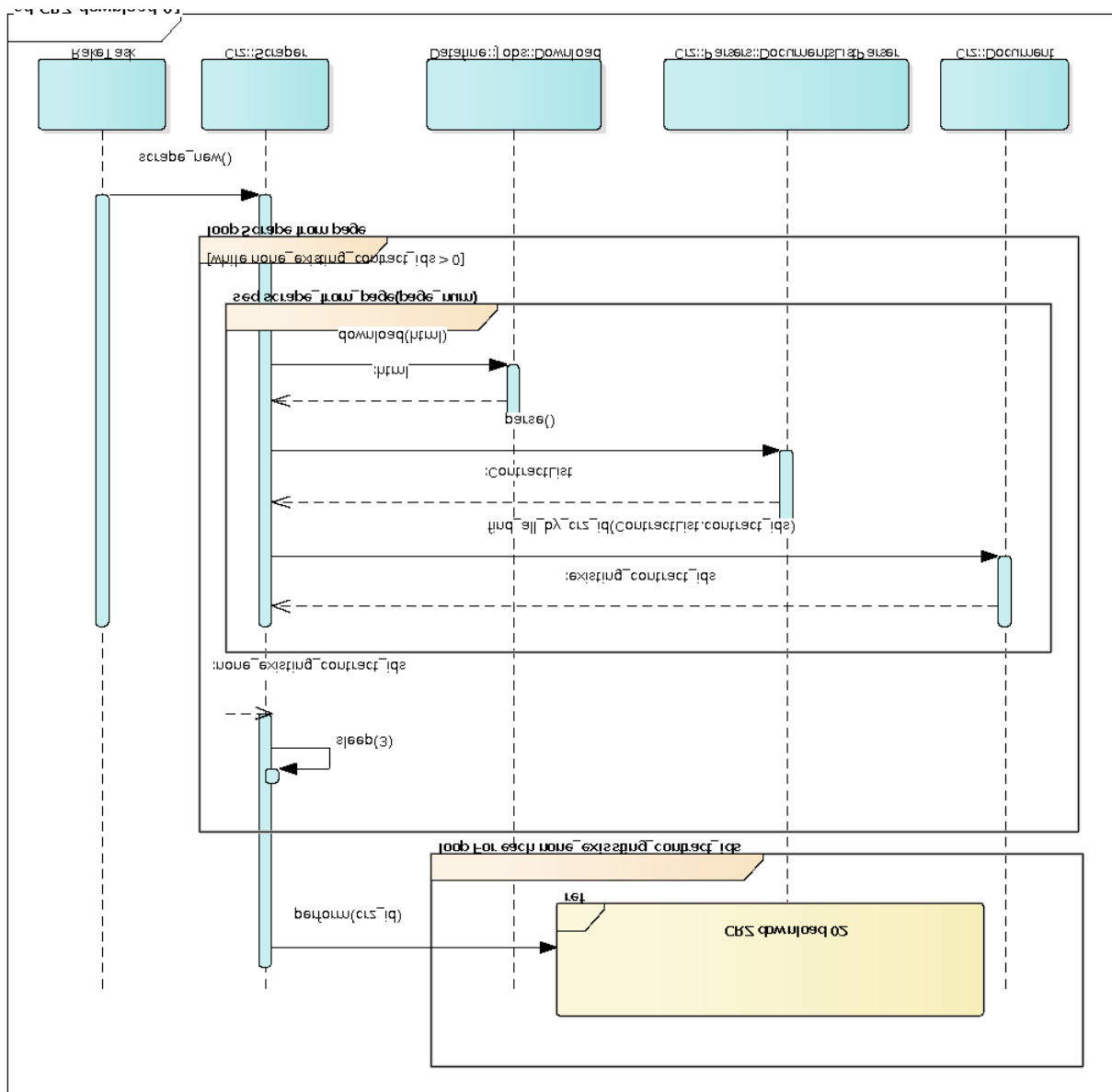


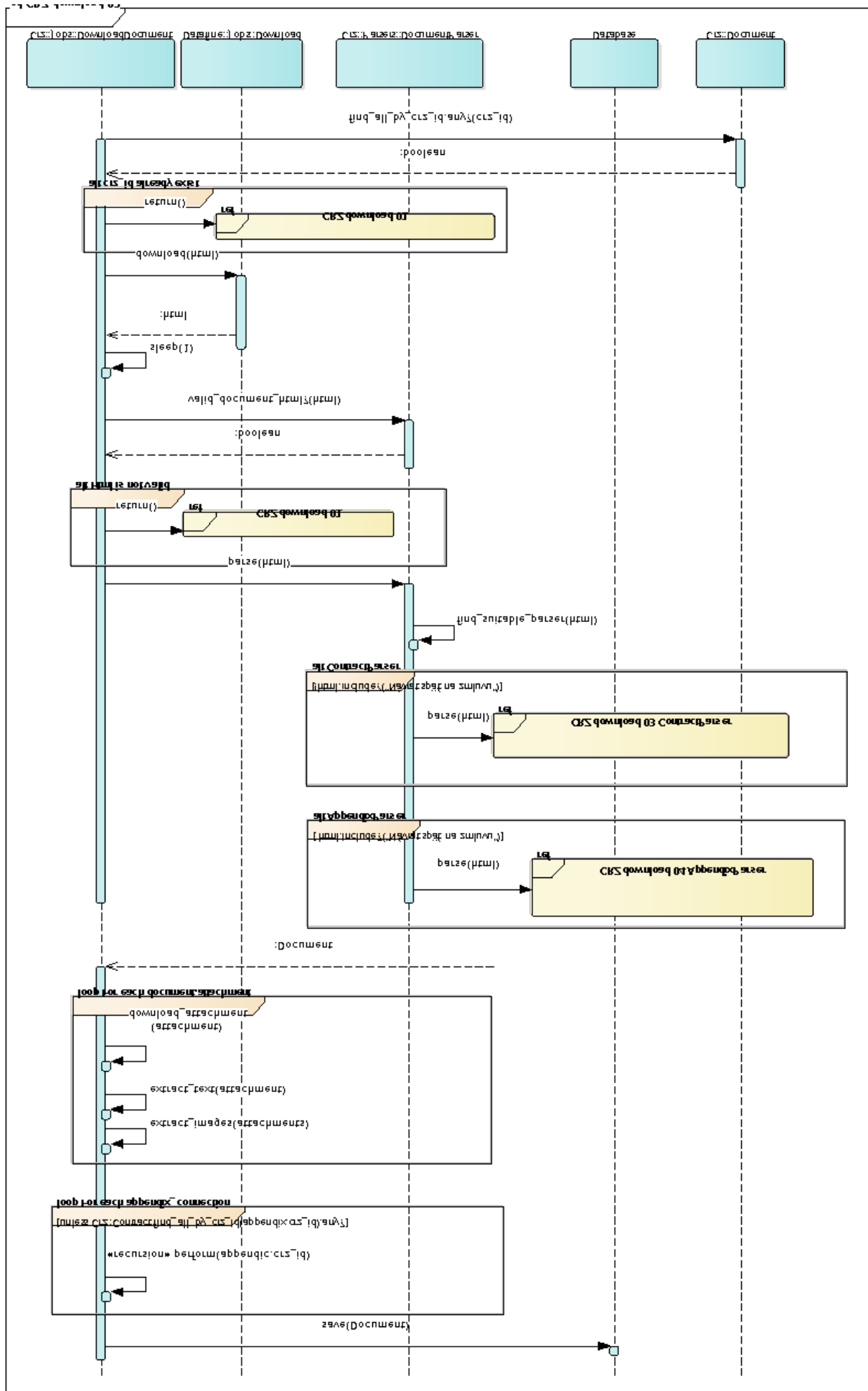
1. Sťahovanie zmlúv z porálu <http://www.crz.gov.sk/>

Niektoré procesy v tomto projekte sú spúšťané periodicky. To sa deje pomocou cron job (časový plánovač udalostí v systémoch založených na UNIXe). Tieto joby sú definované v súbore config/schedule.rb. Následne tieto cron joby vyvolajú tasky definované v súbore lib/tasks/crowdcloud.rake. Jeden z týchto taskov je sťahovanie nových (doteraz nestiahnutých) zmlúv z portálu <http://www.crz.gov.sk/>. Celé jeho vykonávanie je zobrazené na štyroch sekvenčných diagramoch uvedených nižšie.



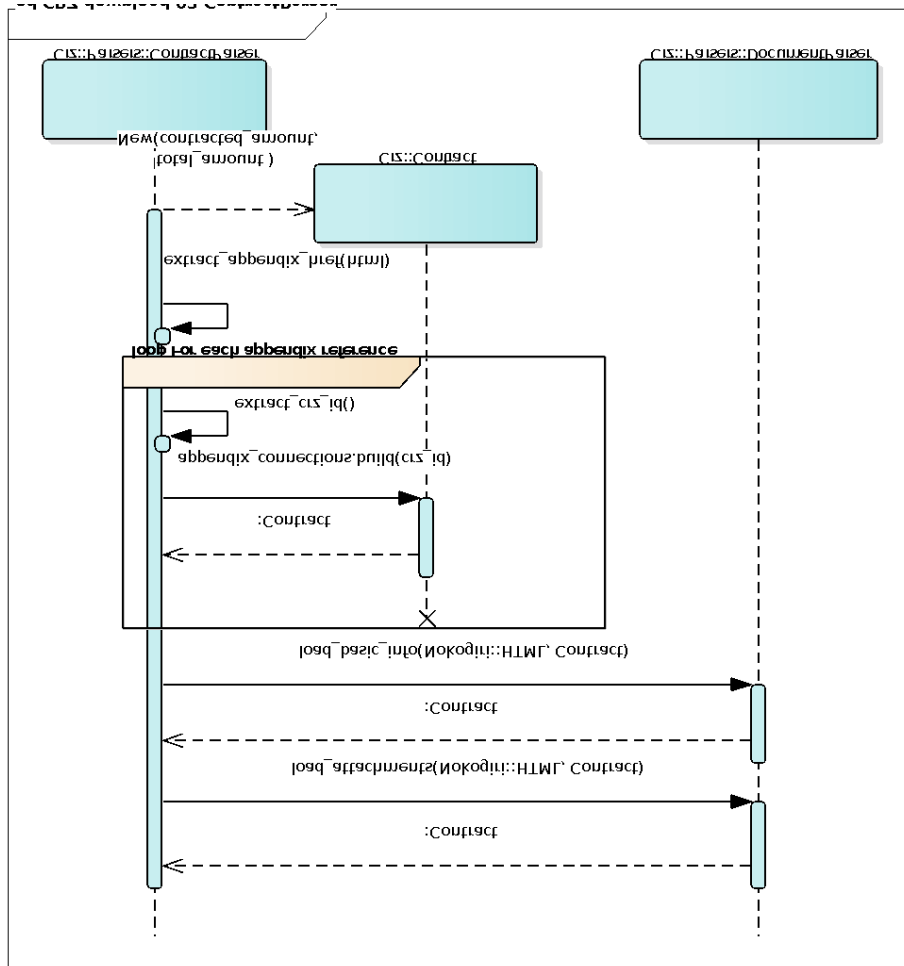
Popis k sd CRZ download 01

Zo stránky, kde sa nachádzajú zmluvy zoradené podľa dátumu najnovších zmlúv sa stiahne zoznam z 20 zmluvami. Vyparsujú sa id daných zmlúv. Tento proces pokračuje pokiaľ sa všetky zmluvy na danej stránke (teda 20), už nachádzajú v databáze.



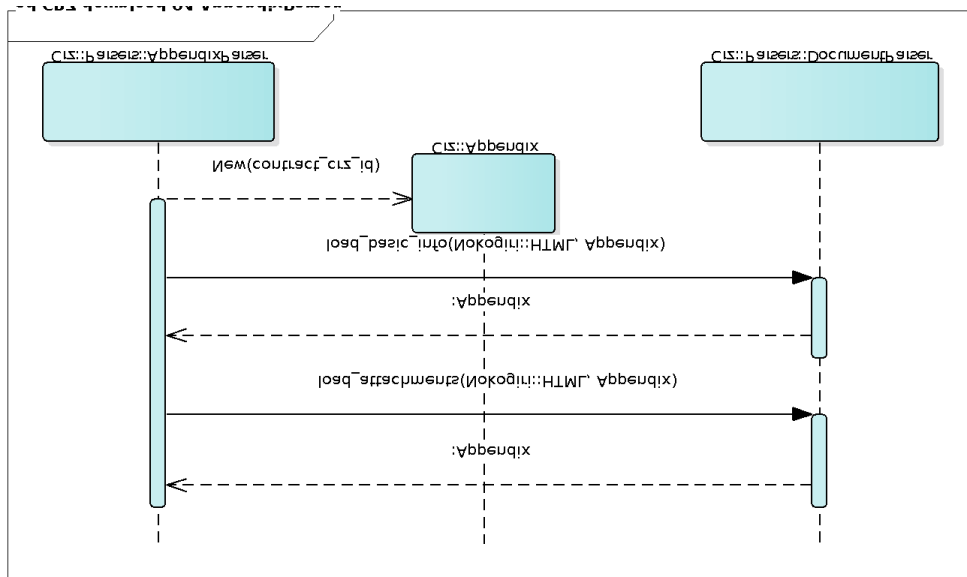
Popis k sd CRZ download 02

Pre každé id zmluvy sa následne spraví kontrola id, stiahne sa html stránka zmluvy a podľa obsahu sa použije buď AppendixParser alebo ContractParser, podľa toho, či sa jedná o novú zmluvu alebo nejaká zmena či dodatok zmluvy. Po tomto sa pre každú prílohu stiahne každá príloha, extrahuje sa z nej text a obrázky. Ak sa jednalo o Appendix a vytvorili sa pripojenia na ďalšie appendixy tak sa rekurzívne vykonáva proces tohto sekvenčného diagramu pre každé prepojenie. Nakoniec sa celý dokument (metadata) uložia do databázy.



Popis k sd CRZ download 03 ContractParser

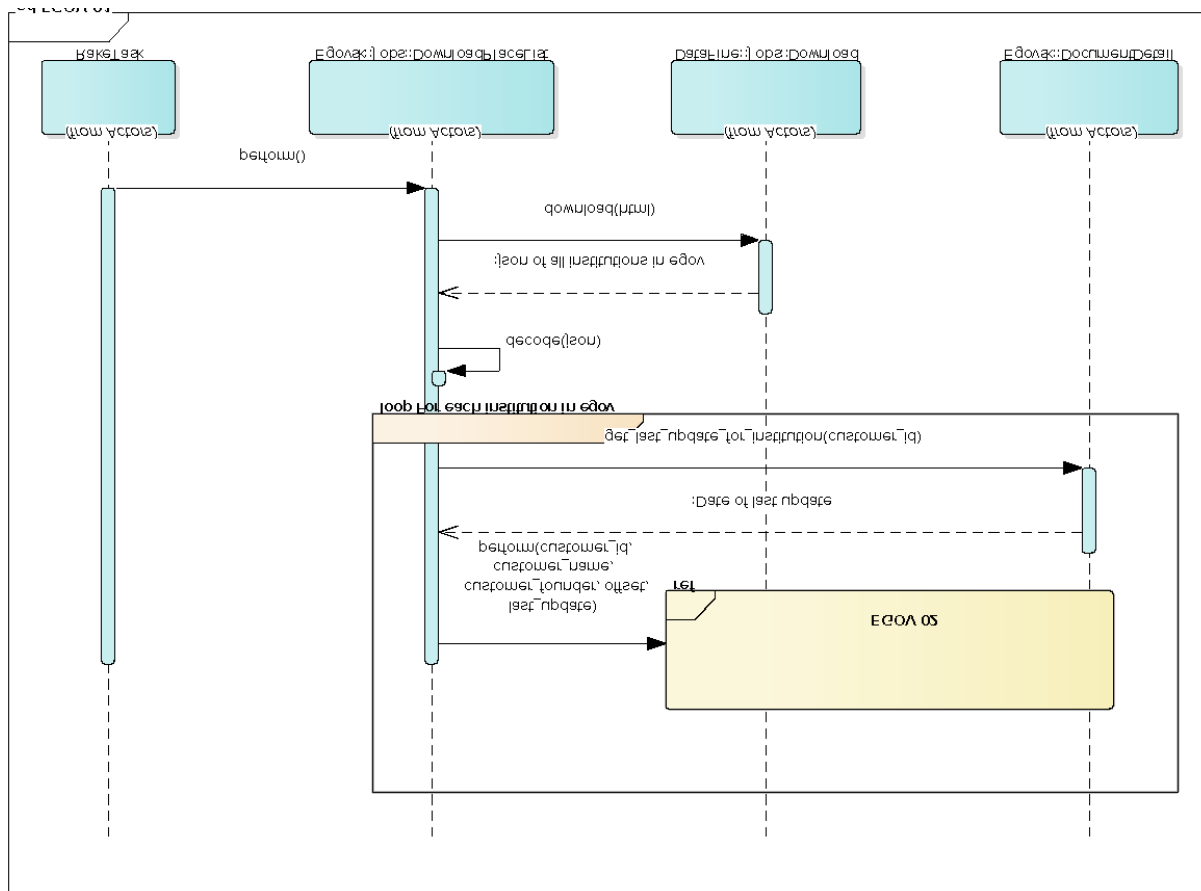
Ak sa jedná o novú zmluvu, tak ak sú k nej nejaké appendixy tak sa extrahujú url referencie, idečka a vytvorí sa prepojenia. Načítajú sa základné metadata z html stránky o zmluve a potom o prílohách k nej.



Popis k sd CRZ download 04 AppendixParser

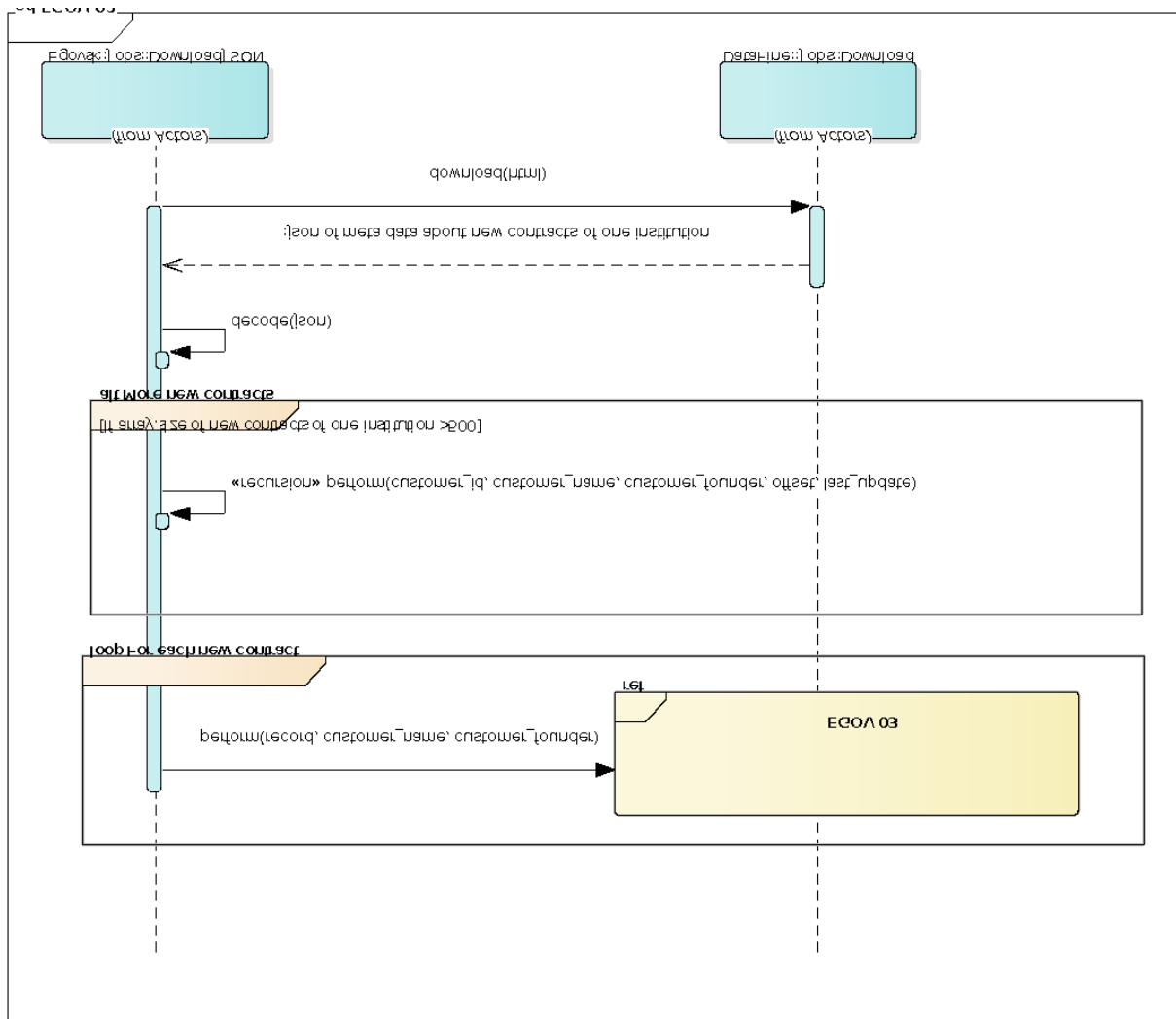
Vytvorí sa objekt Appendixu, načítajú sa o ňom základné metadata a taktiež aj k jeho prílohám.

2. Sťahovanie zmlúv z portálu <http://zmluvy.egov.sk/>



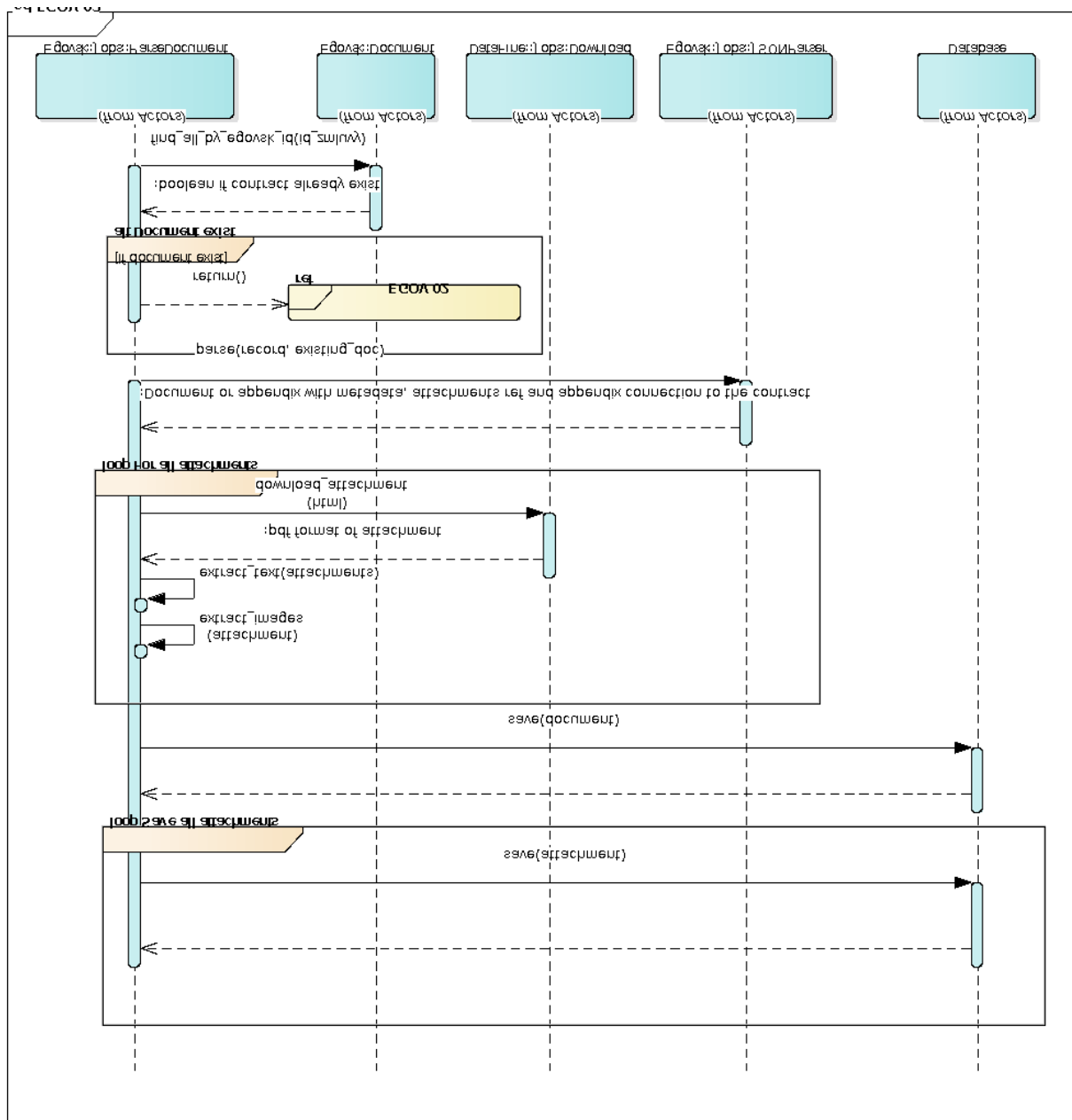
Popis k sd EGOV 01

V procese sťahovania nových zmlúv z portálu egov sa najprv stiahne json zoznám všetkých inštitúcií, ktoré uverejňujú zmluvy na tomto portáli. Z databázy sa načíta dátum posledného updatu pre každú inštitúciu.



Popis k sd EGOV 02

Následne sa vyskladá url adresa pre danú inštitúciu napríklad: <http://zmluvy.egov.sk/egov/jsonExport/place:105/offset:0/since:2017-01-01>. Stiahnu sa metadáta o zmluvách. Ak sa v danom json súbore nachádza viac ako 500 záznamov posunie sa offset o 500 a zavolá sa rekurzívne vykonanie tohto sekvenčného diagramu.



Popis k sd EGOV 03

Pre každú novú zmluvu sa uložia do objektu Document metadata a buď sa jedná o objekt Appendix alebo Contract podľa toho či sa jedná o novú zmluvu alebo dodatok a pod. V prípade Appendixu sa navyše vykoná len prepojenie na originál zmluvy. Načítajú sa url referencie na attachments z ktorých sa extrahuje text a obrázky. Tie sa uložia na filesystem a následne sa všetky metadata o zmluve a prílohách uložia do databázy.