

Универзитет у Београду
Математички факултет
Катедра за рачунарство и информатику



CMIS стандард и примена у електронској архиви истраживачког тима

МАСТЕР РАД

Студент: Никола Драгићевић, 1046/2011

Ментор: Проф. др Гордана Павловић-Лажетић

септембар 2014. год.

Математички факултет у Београду, Катедра за рачунарство и информатику

Захвалница:

Ову прилику користим да се захвалим свом ментору на помоћи, проф. др Гордани Павловић-Лажетић. Похађао сам код професора неколико курсева, између осталог и курс из Биоинформатике, што ми је дало идеју за примену проблематике којом се овај рад бави. Професоров приступ тематици и методологији, сугестије и савети, учинили су да овај рад буде још бољи. Посебно сам захвалан професору за изузетно стрпљење, разумевање и подршку коју ми је пружила у настајању овог мастер рада.

Математички факултет у Београду, Катедра за рачунарство и информатику

Такође бих се захвалио и асистенту Јовани Ковачевић на помоћи око дефинисања Алфресковог модела података и функционисања апликације која прати овај рад, као и господину Зорану Марјанову на идеји за саму тему.

Задовољство ми је и да се захвалим члановима комисије, др Саши Малков и др Јелени Граовац на посвећеном времену за преглед и оцену рада.

Апстракт

DMS (Document Management System) је рачунарски систем или скуп програма који се користе за складиштење, управљање и праћење електронских докумената, њихових верзија и корисника који производе те верзије. Алфреско је DMS отвореног кода, посебно погодан за имплементацију компликованих пословних процеса и сарадњу.

CMIS (Content Management Interoperability Services) је отворени стандард који дефинише извесни ниво апстракције за контролисање различитих DMS-ова, али и за размену електронских докумената, коришћењем веб протокола. Посебно је користан јер омогућава повезивање разнородних DMS-ова, затим, теоретски омогућава трансфер електронских докумената са њиховим метаподацима и самим тим представља будућност даљег развоја рачунарства у облаку.

Задатак овог рада је да кроз Алфреско прикаже: концепт DMS-а на моделу извесне биоинформатичке истраживачке групе и примену CMIS стандарда на архивирање докумената наведене групе, са фокусом на анализи конкретног недостатка CMIS-а, очувавању свих метаподатака приликом електронске размене докумената између два Алфреско DMS-а.

Садржај:

- 1) Увод..... 5
- 2) Технологије рада са електронским документима..... 6
 - 2.1. DMS..... 6
 - 2.2. Alfresco DMS..... 8
 - 2.3. CMIS стандард..... 12
 - 2.4. Проблем и значај очувања метаподатака..... 15
- 3) Очување метаподатака..... 18
 - 3.1. Могућа решења проблема..... 18
 - 3.2. Упоређивање могућих решења..... 19
 - 3.3. Образложење избора решења за имплементацију..... 25
- 4) Технички аспект решења..... 26
 - 4.1. Проблеми у вези са имплементацијом CMIS-а..... 26
 - 4.2. Детаљи и проблеми у вези са изабраном верзијом Алфреска за имплементацију..... 29
 - 4.3. Детаљи одабраног решења, предности и недостаци..... 38

5) <u>Имплементација.....</u>	
<u>..... 40</u>	
5.1. <u>Опис модела података и</u>	
<u>сценарија..... 40</u>	
5.2. <u>Операција</u>	
<u>архивирања..... 45</u>	
5.3. <u>Тестирање сценарија без примене</u>	
<u>имплементираних решења и</u>	
<u>резултати..... 49</u>	
5.4. <u>Тестирање сценарија са применом</u>	
<u>имплементираних решења и</u>	
<u>резултати..... 53</u>	
6) <u>Закључак.....</u>	
<u>..... 55</u>	
7) <u>Литература.....</u>	
<u>..... 57</u>	

1. УВОД

У савременом друштву, тимски рад је, практично, незамислив без употребе рачунара. Управо је групни рад иницирао настанак интернета седамдесетих година прошлог века, а потреба људи да брзо дођу до различитих информација је подстакла даљи развој, који и данас неометано траје. Тренутни развој веб технологија је у правцу и духу рачунарства у облаку.

Cloud computing, односно рачунарство у облаку, је мрежна технологија која обезбеђује флексибилан, од локације независан, приступ рачунарским ресурсима који се брзо и неприметно алоцирају и деалоцирају, према потражњи. Посебно је корисна у пракси, јер смањује трошкове одржавања по кориснику, убрзава корисничке процесе и лако је проширива. У прилог овој технологији иде и чињеница да је институцијама, организацијама и компанијама са великим бројем запослених, неопходан систем за управљање електронским документима који ће покривати све потребе пословних процеса наведених за имплементацију и сарадњу, а то је управо оно што је концепт *DMS-a (Document Management System)*. Алфреско је један такав управљачки систем који у потпуности подржава *CMIS (Content Management Interoperability Services)* протокол.

CMIS је отворени стандард који омогућава различитим *DMS*-овима међусобну интеракцију. Његова најчешће коришћена имплементација (*OpenCMIS*, *Apache Chemistry*) омогућава трансфер електронских докумената између два *DMS*-а, али тек са селективно сачуваним метаподацима, што је крупан недостатак.

Овај рад описује примену имплементације *CMIS* стандарда у електронској размени докумената, користећи два Алфреско *DMS*-а; први представља продукциони систем једне биоинформатичке, истраживачке групе, а други архиву њених електронских докумената. Кључна примена стандарда ће се огледати у анализи наведеног недостатка. За конкретни модел података истраживачке групе, биће понуђено неколико решења, а одабрано једно, које, иако је статичко, отклања проблем губитака метаподатака приликом трансфера докумената.

2. ТЕХНОЛОГИЈЕ РАДА СА ЕЛЕКТРОНСКИМ ДОКУМЕНТИМА

У данашњем друштву, када говоримо о папирном, правном, личном, документу у ширем смислу, значење је сасвим јасно. Међутим, није баш исти случај када је реч о електронском документу. Једноставно, није довољно рећи да је то било какав тип документа у електронском облику, као што саме речи сугеришу, јер би то била недовољна дефиниција. Неко прецизније објашњење би било да је електронски документ скуп података састављен од слова, бројева, симбола, графичких, звучних и(ли) видео записа, електронски израђен, дигитализован, послат, примљен, сачуван или архивиран на електронском, магнетном, оптичком или другом медију [1].

У почетку, подаци складиштени у сталној меморији рачунара су коришћени углавном за потребе рада самог рачунара, док су сами документи чувани на папиру. Временом, непрекидни развој интернета је све више утицао на потребу трајног складиштења разних типова података (докумената у електронском облику и слично) у меморију рачунара, како из личних разлога појединца, тако и због размене електронских докумената са другим људима. Једна од главних намена рачунара је управо чување и размена електронских докумената са другим појединцима и групама. Но, електронски

документи имају такође и широк спектар друштвене примене: у електронском учењу, управљању, здравству, пословању, банкарству, администарству, трговини итд. Како би се олакшало чување и унапредила размена електронских докумената (у аспекту њихове друштвене примене), осмишљен је и реализован концепт управљачког система за такав вид докумената. Један од новијих протокола за размену електронских докумената између тих система, односно *DMS*-ова, је већ споменути - *CMIS* протокол.

На следећим странама ће бити детаљно описан наведени концепт управљачког система, један одабрани систем (Алфреско *DMS*) за примену *CMIS* протокола, а неће бити изостављене ни појединости о самом протоколу.

2.1. DMS

Од периода настанка администрације, па све до данас, много тога се променило, но, суштина је остала непромењена - документовање људског рада. Као непосредну последицу повећања броја становника, убрзаног технолошког развоја и приступачне едукације у последњем веку, увиђамо чињеницу да данашње организације, предузећа, установе, компаније стварају никад већи број докумената, како папирних, тако и електронских. Разни формулари, понуде, потврде, уговори, рачуни, изводи, обавештења, извештаји, правни акти, писма, мејлови, факсови, скенирана документа и слично, само су део скупа докумената данашње просечне фирме. Са порастом броја запослених, расте и број докумената једне организације, па никаква новост није да радник или његов надређени добије главобољу тражећи неки затурени документ (који је врло вероватно и изгубљен), посебно ако је у питању папирни документ. По студији аналитичара компаније *Coopers & Lybrand* (данас - *PricewaterhouseCoopers*) [2] просечни папирни документ једног предузећа се копира 19 пута, приближно се сваки четрнаести документ изгуби, док 3% од преосталог броја чине „залутали“ документи.

Применом рачунара у административне сврхе, повећан је степен сигурности од губитка документа, документи су прегледнији, приступачнији, лако се могу изменити, претражити, допунити, направити, али ипак, све наведено, није довољно. Олакшана реорганизација, брзина проналаска документа и свих осталих побољшања која рачунар (са било којим оперативним системом) нуди, су довољно добра за свакодневне потребе појединцу, не и предузећу. Разлог? Рецимо само то да су на било ком оперативном систему управљачке могућности над документима скромне, метаподаци су најосновнији и немогуће је додати нове, сарадња више корисника је ограничена и верзионисање докумената не постоји. Непрекидна еволуција рачунара, апликација и система је обликовала и објединила решења за ове, али и многе друге проблеме у виду система за управљање документима (*Document Management System, DMS*). Но, пре него што концепт *DMS*-а буде објашњен, рећи

Ћемо о нешто уопштенијем појму, односно систему за управљање садржајем (*Content Management System, CMS*).

CMS представља систем за управљање, контролу и модификовање разноликог садржаја попут веб садржаја, докумената и садржаја рачунарства у облаку. Ипак овако представљена дефиниција обухвата доста тога, а нека од конкретизованих замисли је управо *DMS*. Укратко речено, *DMS* је управљачки систем (или скуп апликација) који служи за складиштење, контролу и праћење електронских докумената. До пре десетак година (када су овакви системи настајали), били су скупни за одржавање, а наведене три сврхе су им биле и једине. У међувремену, *DMS*-ови су пратили развој технологије, као и сами рачунари, постали су приступачни свакоме и добили су неке нове, ништа мање значајне намене [3]:

- **Олакшано процесирање скенираних докумената.**
- **Могућност закључавања докумената.**
- **Индексирање** значајно убрзава све критеријуме претраге и поједностављује праћење докумената.
- **Метаподаци.** Најједноставније речено, подаци о подацима. Неки примери метаподатака електронских докумената могу бити: опис документа, датум скенирања документа, последњи корисник који је мењао документ или име корисника који је документ креирао итд.
- **Интеграцију** са другим апликацијама подржава већина *DMS*-ова. Она подразумева приступ неке локалне или мрежне апликације електронским документима *DMS*-а, преко разних протокола попут *LDAP (Lightweight Directory Access Protocol)*, *WebDAV (Web Distributed Authoring and Versioning)*, већ споменутог *CMIS*-а...
- У *DMS*-овима, напредна **сарадња** корисника је подразумевана. Најновији управљачки системи пружају подршку за истовремену, али синхронизовану, промену садржаја неког документа.
- **Верзионисање.** Уопштена карактеристика *DMS*-а је да чува све претходне верзије неког документа, након мењања његовог садржаја. Ово омогућава корисницима преузимање неке претходне верзије документа и наставак рада на њој, али и увид у историју свих промена изабраног документа.
- **Валидација** било ког формата електронских докумената је чест атрибут новијих *DMS*-ова.
- **Безбедност** електронских докумената је свакако на вишем нивоу у сваком управљачком систему него ли на оперативном систему.
- **Радни токови.** Још једна од кључних предности савремених *DMS*-ова над оперативним системима, али такође и уопштенијим, наменски сличним *CMS*-овима. Наиме, реч је о имплементацији компликованог процеса међусобне пословне сарадње више корисника над документом или групом докумената или пак, пословног правила за напреднију

валидацију података или процеса. Размотримо следећи пример: Директор или шеф предузећа покрене радни ток у коме задуженим радницима прослеђује, у виду задатка, скуп електронских докумената који чине неколико пословних понуда. Конкретан задатак радника је да (након што прихвате задатак) независно једни од других, анализирају, простудирају те документе и надређеног обавесте о донесеном закључку, после чега директор или шеф може, (не)задовољан, да оконча консултацију, тј. затвори радни ток или пак, да проследи радницима нове документе, инструкције, захтеве итд.

Но, ово свакако није крај листе добрих својстава које управљачки системи поседују, а и нова се непрестано осмишљавају, зато што су управљачки системи важна компонента машинерије зване: „рачунарство у облаку“. Следеће странице су посвећене опису једног конкретног *DMS*-а који ће бити коришћен за потребе овог рада.

2.2. Alfresco DMS

Алфреско је бесплатан *ECM* (*Enterprise Content Management*) систем, отвореног кода. *ECM* је скраћеница за предузетничко управљање извесног типа садржаја, али могло би се рећи да је то скуп стратегија, метода и алата за чување, управљање, размену садржаја у некој организационој јединици [4]. Одавде се може закључити да је *ECM* примењени *CMS*, али у ствари, појам *ECM*-а је пре свега теоретски, а нека од његових конкретизованих примена се огледа управо у *DMS*-у. С обзиром да је софтвер Алфреско оријентисан ка *ECM*-у, он има и друге намене осим тога што је *DMS*, а то су управљање веб садржајем, сликама и записима. Алфреско има два радна модула (*Explorer* и *Share*), а може се преузети у три верзије:

- Верзија за заједнице (*community edition*).
- Верзија за предузећа (*enterprise edition*).
- Верзија за рачунарство у облаку (*cloud edition*).

С обзиром да је друга верзија комерцијална, а последња само за рачунарство у облаку, за потребе овог рада надаље ће бити речи искључиво о свеобухватној, првој верзији и то из перспективе најопштијег контекста управљања подацима, *DMS*-а.

Џон Њутн (*John Newton*) и Џон Пауел (*John Powell*) су основали фирму *Alfresco Software* 2005. године. Идеја им је била да тадашњим најновијим технологијама и аспектно оријентисаним принципима програмирања, створе флексибилан, скалабилни управљачки систем отвореног кода, који неће бити стриктно веб, већ пословно оријентисан, уз лаку примену у имплементацији пословних процеса [5]. Осам година касније, више од седам милиона људи у 75 земаља (укључујући и Србију) свакодневно користе *Alfresco ECM*. Са преко 150 стратешких партнера широм света, *Alfresco Software* је добро познато име у свету *ECM* система, а стратешки партнер код нас им је

Математички факултет у Београду, Катедра за рачунарску информатику

Београдски институт „Михајло Пупин“.

Пауел

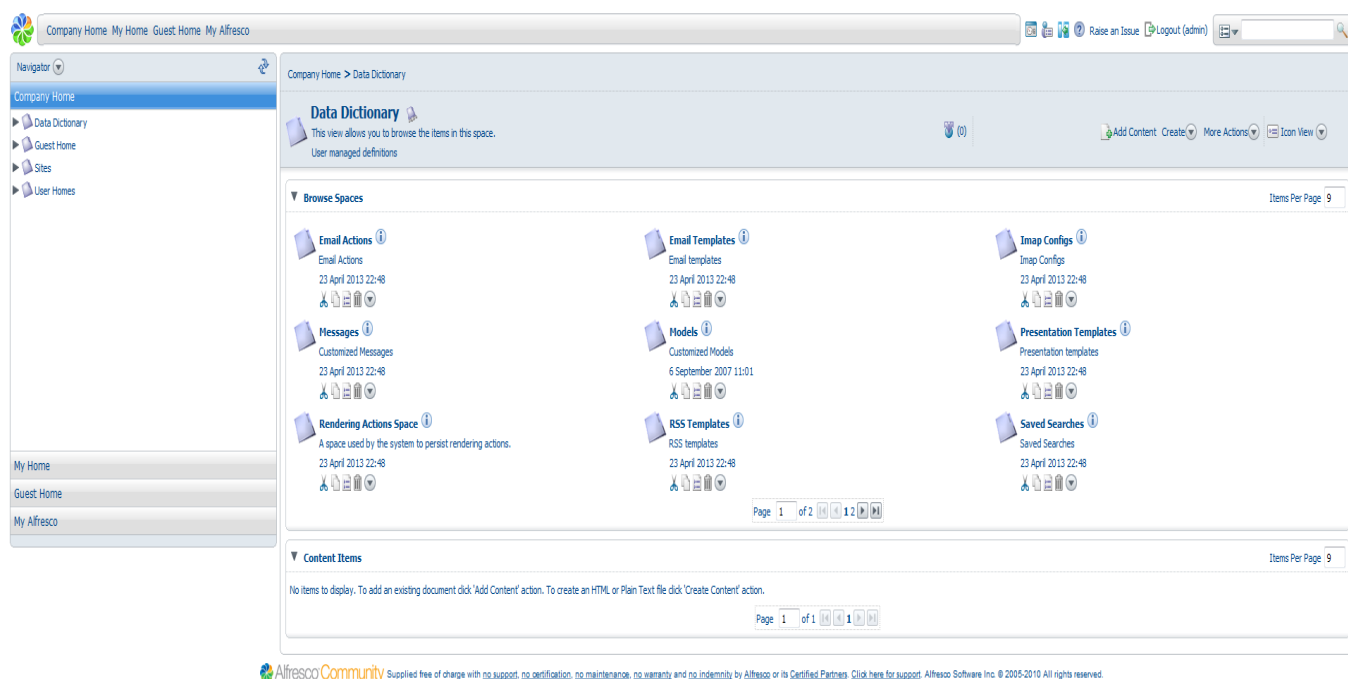
Опис особина Алфреско *DMS*-а свакако мора почети од складишта података, односно - **репозиторијума**, локације где корисници могу пронаћи све документе *DMS*-а. Репозиторијум није ништа друго до управљачко-навигациони алат који има сваки оперативни систем (попут *Windows Explorer*-а у *Windows*-у, *Konqueror*-а у *Linux*-у, *Finder*-а у *Mac OS*-у и слично), али је немерљиво обogaћен пропратним садржајем, могућностима, опцијама и веома је прилагодљив самој намени управљачког система, док се у исти мах приступање и манипулација документима унутар директоријума одвија на исти начин као у споменутом алатима. Садржај репозиторијума се смешта у „**просторе**“; они представљају напредне директоријуме Алфреско система. Репозиторијум и типови докумената и простора који могу постојати су дефинисани у Алфресковом **моделу података**, о чему ће касније бити више речи.

Осим споменутих метаподатака „напредност“ простора се огледа у могућности извршавања извесних активности, прецизније - **акција**, нпр. приликом додавања, ажурирања или брисања докумената (попут аутоматског попуњавања метаподатака, категоризације или, пак, разних обавештења корисницима). Још једна (на)предност је и могућност креирања шаблонских докумената и простора у репозиторијуму.

Претрага репозиторијума је базирана на претрази уобичајеног управљачко-навигационог алата, али је, такође, проширена трагањем по (шаблонском) типу документа и простора, а могуће ју је проширити и на претрагу по вредности појединачног метаподатка. Све наведене активности извршене од стране било ког корисника су забележене у тзв. **audit**-у Алфреска, додатном, хронолошком прилогу сваког простора и документа, чиме се осигурава одговорност корисника за своје поступке.

У Алфреско *DMS*-у постоје три типа корисника: администратор, обичан корисник и опционо, гост. Уобичајено, администратор је корисник који има видљивост над свим документима система и има дозволу да изврши било коју акцију над њима, док гост има јако сужену, само најуопштенију видљивост система и дозвољена му је једино могућност прегледа података. Корисници који не спадају у ове две групе, могу припадати било којим другим, унапред дефинисаним корисничким групама. Приступ документима и просторима, могућност извршавања акција и све остале привилегије се могу дефинисати на нивоу корисничке групе или појединачног корисника.





Слика 3: Поглед на репозиторијум Алфреско Експлорера и садржај једног простора.

Сарадња и комуникација између корисника се огледа у прилагодљивим форумима. Наиме, корисници могу креирати дискусију која се односи на конкретан докуменат или простор и она је смештена у самом простору, како не би гушила сам систем или глобални форум. Такође, могуће је креирати и тематске дискусије, али круну сарадње у Алфреску ипак чине **радни токови** - могућност доделе *ad hoc* задатка, као и преглед и одобравање садржаја (*review & approval*). У оба сценарија, корисник који започиње радни ток, може проследити документ (или неколико њих), као и извесне пропратне информације и очекује од корисника који прима(ју) задатак, некакве повратне информације.

Сваки корисник има јединствену, персонализовану **надзорну плочу** (*dashboard*), чији део може заузети жељени поглед над делом репозиторијума, пројеката који корисник има право да види, али део могу заузети и активности, преглед (активних) задатака корисника, линкова, календара итд. Другим речима, надзорна плоча представља конфигурабилан, прилагођен поглед на све оно што корисник сматра најважнијим (било то унутрашњи или спољашњи извор) у оквиру његових задужења, навика или само радних обавеза на систему.

У прошлом поглављу је било речи о томе које су намене *DMS*-а, а мора се додати да (у зависности од одабране верзије) Алфреско *DMS* пружа подршку и за [6]:

- Рад на више платформи.
- Једноставно преузимање и инсталацију.
- Никакве инсталације на клијентској страни нису неопходне.
- Управљање записима и сликама, у било ком формату.

- Динамичко верзионисање докумената, али и комплетног репозиторијума које може бити и аутоматско.
- Једноставно враћање репозиторијума на претходну верзију (у случају колапса система или било ког другог разлога).
- Виртуелне огранке репозиторијума за различите намене (рецимо, изолација докумената попут *Recycle Bin*-а у *Windows*-у).
- Управљачке системе за електронско учење.
- Вишејезичке садржаје.
- Интеграцију са *Microsoft Office*-ом, *OpenOffice*-ом и *LibreOffice*-ом.
- Широки спектар трансформација из једног формата документа у други.
- Више база података: *MySQL*, *PostgreSQL* (само у верзији за заједнице), *Oracle*, *IBM DB2*, *Microsoft SQL Server* (само у верзији за предузећа)...
- Додатно оверавање аутентичности: *LDAP (Lightweight Directory Access Protocol)*, *KERBEROS*, *CAS (Central Authentication Service)*, *NTLM (NT Lan Manager)*.
- Закључавање докумената.
- Трансакционалност за све врсте расположивих сервиса.
- Широки ниво конфигурабилности (као последицу модуларности и решења у духу аспектно оријентисаног програмирања).
- Приступ репозиторијуму преко неколико протокола: *FTP (File Transfer Protocol)*, *WebDAV*, *CMIS*, *CIFS/SMB (Common Internet File System/Server Message Block)*, *NFS (Network File System)*.
- Графички интерфејс интегрисан у веб претраживач.

Алфреско је реализован у програмском језику **Java**; већину докумената апликације чине Java зрна и конфигурациони XML-ови. Апликација је углавном базирана на најбољим алатима отвореног кода - окосницу чине *Spring Framework* и *Java Server Faces (MyFaces)*, веб сервери који представљају извршно окружење за Java зрна могу бити *Apache Tomcat* или *JBoss*, али (од до сад неспоменутих технологија) интегрисани су такође и *Hibernate* (користи се за објектно-оријентисано мапирање и рад са базом података), *Lucene* и *SOLR* (имплементирају индексирање и претрагу), *Freemarker* (генерише шаблоне разнолике намене, од најпростијих за обраду обичног текста, па све до *CSS* и *Java* извршног кода), *Activiti* и *JBPM* (имплементирају радне токове докумената) итд [7].

Све споменуте технологије обједињене у Алфреско *DMS*-у, чине овај систем изразито скалабилним и прилагодљивим еволуцији самог система са растом броја корисника и докумената које корисници обрађују.

Enterprise Content Management (ECM) систем се традиционално сматра за веома скуп, како за само лиценцирање, тако и за одржавање, а неретко захтева и скупе хардверске компоненте и пратеће софтвере. Верзија Алфреско *ECM*-а за предузећа је комерцијална - наплаћује се годишња претплата и захваљујући њој,

Alfresco Software се финансира и у могућности је да развија, обogaђује и дистрибуира остале, бесплатне верзије. Ипак, у односу на производе водећих компанија у свету *ECM*-а и *DMS*-а, попут *Documentum*-а, *OpenText*-а или *Microsoft Share Point*-а, цена Алфреско *DMS*-а (верзије за предузећа) је и до десетоструко мања [8]. За ту цену, односно, годишњу претплату, корисници имају следеће услуге:

- Техничка подршка развојног тима Алфреско *DMS*-а и одговор у највише два часа.
- Моментални приступ свим новим верзијама, надоградњама и поправкама.
- Лак прелазак са било које верзије апликације на жељену, новију.
- Подршка приликом развијања и модификације сопствене верзије *DMS*-а.
- Гаранција за производ и обештећење у случају било какве штете [9].

Овакав приступ омогућава Алфреску да понуди много јефтиније решење било којој организацији, компанији, предузећу које траже потпуно дефинисано, подржано и тестирано *ECM/DMS* решење.

Корисници *DMS*-а желе једноставно конфигурисање система, интегрисаних апликација и секторских решења, док корпорације желе сталну, складну контролу скалабилности и модуларности, што управо Алфреско *DMS* и нуди.

Захваљујући скромним хардверским захтевима (за 50 конкурентних корисника или 500 повремених препоручено је 1.5 GB RAM меморије за *Java* виртуелну машину и процесор са два језгра, док је за 200 конкурентних или 2000 повремених корисника препоручено 2.5 GB RAM меморије и 8 процесорских језгара [10]), високој поузданости и брзини рада, лакој руковању и одржавању, Алфреско је израстао у водећи *ECM*, односно *DMS*, отвореног кода.

2.3. CMIS стандард

Пре десетак година, ушли смо у нови миленијум, али и у еру (непрестане) експанзије система за управљање различитим садржајем. Упоредо са развојем *CMS*-ова текао је и развој других управљачких система, попут система за управљање документима, записима, сликама, веб садржајима итд. Примењене управљачке системе свакодневно виђамо у облику блогова, веб страна за разне рекламе, вести, спорт...Примену такође налазе и у школама, предузећима, истраживачким установама, интернационалним компанијама итд. Још 2010. су три најпопуларнија *CMS*-а (*WordPress*, *Joomla*, *Drupal*), од стотине постојећих, бележила преко 1.100.000 преузимања седмично [11] а данас, трећина свих веб страна користе неки од система за управљање садржајем [12].

Ипак, као и код других рачунарских система, од важности је и интеракција са свим осталим системима. Но, нажалост, овде се наилази на проблем - не постоји подршка за тако нешто. Управо ту на *CMS/DMS* сцену ступа *CMIS* протокол. **Content Management**

Interoperability Services, или скраћено - CMIS је отворени стандард који дефинише извесни ниво апстракције за контролисање различитих управљачких система, али и за размену електронских докумената, коришћењем веб протокола.

OASIS конзорцијум је јавности представио CMIS 1.0 спецификацију 1. маја 2010, а следећа, 1.1 (тренутно и најновија) верзија је одобрена 12. децембра 2012. године [13]. Идеја конзорцијума је била да се осмисли стандард за ECM системе који би био нешто што је SQL за релационе базе података, односно, такав стандард који би премостио разлике између система базираних на различитим платформама, програмским језицима и слично, са готово истом синтаксом на свакој од њих [14]. Од важности је споменути и имена чланова удружења - конзорцијума: *Adobe Systems Incorporated, Alfresco Software, EMC Corporation, eXo, FatWire Software, Hewlett-Packard, IBM, ISIS Papyrus, Liferay, Microsoft, Nuxeo, OpenText, Oracle, Newgen OmniDocs* и *SAP*.

Замишљен као интерфејс највишег нивоа неког CMS/DMS-а, CMIS има особину мапирања четири различита типа објеката у неком репозиторијуму: документа, директоријума, везе (између претходно наведених објеката) и политике. Последњи тип представља извесну административну политику (можда је и прецизније рећи дозволу или забрану за конкретну активност [15]) која се односи на неке од преостала три типа објеката.

Да би одређени систем интегрисао CMIS у своју архитектуру, неопходно је да претходно пружи подршку за његове градивне компоненте; динамичност у интеракцији са разноврсним системима је остварена спојем посебног модела домена, прилагођених веб сервиса и *Restful AtomPub* протокола (базираном на HTTP-у, а служи за креирање и ажурирање веб ресурса). Споменути модел разликује разне објекте, а тиме и типове докумената и директоријума, чиме не одступа од стандардне архитектуре управљачких система докумената, већ напротив, може се применити на било који модел података CMS-а и(ли) DMS-а. Придружени скуп веб сервиса омогућује читање, писање, прављење, ажурирање и брисање како самих докумената и директоријума, тако и њихових својстава и метаподатака, све уз помоћ *Restful AtomPub* протокола. Такође, скупу сервиса припадају и сервиси за навигацију и претрагу репозиторијума, верзионисање докумената итд.

За упите претраге над мапираним моделом, CMIS користи интерфејс базиран на језику који је сличан стандарду SQL-92. Интерфејс је заправо схема и погледи који јој припадају, генерисани у зависности од модела података система које је стандард мапирао. Језик интерфејса није проширив и има ограничену подршку за JOIN операције, док у потпуности не подржава GROUP BY и HAVING операције. Другим речима, SELECT и FROM су обавезне, а могу се користити и WHERE и(ли) ORDER BY кључне речи. У зависности од репозиторијума, језик може да изврши тзв. *full-text* претрагу и упите базирани на метаподацима (нпр. упит **SELECT F.cmis:name AS name, F.cmis:objectId AS id FROM cmis:folder as F** враћа све

називе и идентификаторе из погледа који садржи CMIS објекте типа директоријума).

Последња верзија CMIS стандарда доноси и значајни допринос динамичком концепту и корисне новине. Наиме, у изворној верзији CMIS-а, да би се неки тип документа ефикасно послао са једног DMS-а на други, неопходно је да модели података оба буду идентични. Неке од многих побољшања у верзији CMIS 1.1 су, рецимо, прилагодљивост која се огледа у томе да се модели података DMS-ова могу мењати у току самог трансфера, подршка за JSON стандард и могућност дописивања садржаја у документ приликом његовог креирања или ажурирања.

У овом тренутку, најшире коришћена имплементација CMIS-а (отвореног кода), *Apache Chemistry*, омогућава трансфер електронских докумената, али тек са селективно сачуваним метаподацима. Алфреско DMS користи библиотеке програмског језика *Java*, а *OpenCMIS* 0.12.0 је скуп најновијих алата и библиотека који имплементирају актуелни, CMIS 1.1 стандард на том језику [16]. Ипак, ово би требало прихватити са великом дозом резерве. Претходних једанаест верзија (које су сукцесивно објављиване у протекле четири године) су тек са верзијом *OpenCMIS* 0.8, може се рећи, у великој мери комплетирали, заокружили имплементацију CMIS 1.0 стандарда, па се са разлогом за приближно исто толико верзија (односно у сличном временском интервалу) очекује комплетно имплементирани CMIS 1.1 стандард. А до тада, програмери морају да импровизују решење за све недостатке имплементација CMIS протокола. Но, пре спомињања недостатака, морају се споменути и неке од предности стандарда:

- Омогућава трансфер електронских докумената са њиховим метаподацима.
- Као последица претходне врлине, представља садашњост и будућност програмирања у облаку.
- Остварује повезивање и интеракцију истородних, али и разнородних CMS-ова и DMS-ова.
- Као последица претходног својства, представља подршку за интеграцију различитих система за управљање подацима.
- Драстично упрошћен развој апликација за трансфер докумената.
- Као последица претходне врлине, доприноси расту корисничке, али и развојне заједнице како програмирања у облаку, тако и мрежног програмирања.
- Обухватио је већину важних предности ранијих протокола сличне намене, а поседује и спој технологија за сопствене будуће потребе и развој.
- Раздвојеност концепта садржаја и веб сервиса, чиме стандард добија на модуларности.
- Независан од платформе, програмског језика и модела података са којим је у контакту .

- Унапређивање постојеће пословне логике и решења за комерцијалну примену у локалном, регионалном и међународном економском сектору [17].

Ипак, како ниједан протокол није савршен, тако и *CMIS* стандард није савршен. Овај рад се неће фокусирати на све конкретне недостатке споменутог протокола, али ево највећих:

- Објекти трансакција су или документи или директоријуми, то јест, *CMIS* мапира свега два родитељска типа свих могућих изведених објеката.
- Уколико у интеракцији два репозиторијума постоје нека контрадикторна правила (рецимо, у једном репозиторијуму документи морају бити унутар неког, барем матичног (енгл. *root*) директоријума, а у другом не морају), *CMIS* ће прескочити тај случај.
- Без икакве подршке за рад са корисницима репозиторијума и њиховим (мета)подацима.
- Имплементација протокола не покрива комплетан протокол.

CMIS протокол је детаљно дефинисан само теоретски и тренутни, главни проблем даљег усавршавања стандарда се огледа у томе што имплементације исувише заостају за самом спецификацијом. Узимајући у обзир само друга два наведена недостатка, овај рад ће се у наставку усредсредити на анализу и превазилажење проблема који настају због споменутих тачака.

2.4. Проблем и значај очувања метаподатака

Подржан од стране свих водећих *ECM* корисника и осмишљен са намером да буде прихваћен као индустријски стандард, *CMIS* протокол је свакако значајан алат радионице зване - рачунарство у облаку. Ипак, са правом и није индустријски стандард (барем не још увек). Разлози који иду у корист последњој чињеници су наведени у претходном одељку, међу недостацима протокола (са нагласком на последња два).

Уопштено гледано, главни проблем (у даљем тексту, *Проблем*, који ће бити прецизније дефинисан у одељку 4.1) имплементације *CMIS* протокола је селективно сачувавање метаподатака, као последица непотпуног мапирања комплетног модела репозиторијума. Ако знамо да су метаподаци једна од кључних, употребних вредности неког *DMS*-а, јасно је колико је овај проблем велики недостатак самом *CMIS* стандарду и сваком *DMS*-у који зависи од трансфера података преко њега. Прошло је нешто више од четири године од објављивања стандарда, а имплементација и даље није комплетна, неизбежна су импровизирана решења за мање проблеме, тј. *Проблем*, што свакако није охрабрујуће за даљу промоцију, развој, усавршавање и коришћење самог протокола. Ипак с обзиром да су у целу ову причу укључени гиганти попут *Microsoft*-а, *Oracle*-а, *IBM*-а или *Adobe*-а и као једна од одлучујућих, динамичких технологија за даљи развој рачунарства у облаку (отелотворење даље визије

рачунарства и веб технологија), сигурно је да *CMIS* протокол има јаку потпору и светлу будућност иако његов потенцијал није искоришћен у потпуности. Стандард је релативно скоро дефинисан и извесно је да ће споменути проблеми пре или касније бити решени, али у овом тренутку, свакако нису за потцењивање.

А зашто и коме је важно да овај Проблем буде решен? Појединцу, свакако није. Но, будућност веб технологија лежи у домену развоја концепта рачунарства у облаку. Захтевнији управљачки системи, убрзан раст броја компанија, установа и фирми које их користе и њихов константно растући профит поспешују и осигуравају непрестан развој рачунарства у облаку.

Према званичним изворима Европске Уније, за период од 2010. до 2015. године, предвиђено је да ће рачунарство у облаку пет економски најјачих држава ЕУ (односно Немачке, Француске, Велике Британије, Италије и Шпаније) донети зараду од импресивних 763 милијарде евра и отвориће око два и по милиона радних места [18]. Узмимо у обзир да су наведене цифре можда само оптимистичка очекивања аналитичара, али чак и да је реч о омашци, и даље су у питању милиони људи и стотине милијарди евра. Сада се лако може доћи до закључка да даљи развој *CMIS* протокола, односно рачунарства у облаку, не само да иде у корист развоју веб технологија, већ и доноси профит, како чланицама Европске Уније, тако и осталим државама.



Слика 4: Дијаграм серверске апликације која као посредник користи CMIS протокол за размену докумената између различитих система.

Прецизније, донеће профит пре свега великим компанијама тих држава. Али каква је ту конкретна улога CMIS стандарда? Која је његова практична примена у којој би Проблем дошао у први план?

Слово *I* у називу CMIS протокола би осим „*Interoperability*“ могло да се тумачи и као „*интеракција*“; пример остваривања интеракције би била размена докумената између две или више фирми преко CMIS протокола који повезује њихове (разнородне) управљачке системе. Ипак, за пуни, тренутни ефекат лепезе услуга које CMIS пружа, неопходна је нека серверска апликација која ће, на изванредан начин (до комплетне имплементације верзије 1.1 протокола) обезбеђивати динамичност и независност очувања метаподатака од самог модела података разних DMS-ова укључених у комуникацији.

Други пример примене интеракције би била размена докумената између више рачунарских складишта у облаку. У већим фирмама, односно компанијама, се користи више DMS-ова, па би интеракција подразумевала и комуникацију свих управљачких система унутар саме те компаније. Ево и једног конкретног примера.

Претпоставимо да компанија „Компанија“ сав свој материјал складишти на три различита репозиторијума: Алфреско *DMS*-у, *Microsoft Share Point*-у и *IBM FileNet*-у. „Компанијин“ сектор продаје, ради рекламирања, тражи захтев за генерисање *PowerPoint* презентација, чији шаблон се налази у репозиторијуму *Share Point*-а, слике и остали ресурси су смештени у репозиторијуму *FileNet*-а, а коначно генерисана презентација би требало да се смести негде на Алфреско *DMS*-у. Пре појаве *CMIS* протокола, овакав проблем би се решавао уз минимум три различита АПИ-ја (апликациона програмска интерфејса), а сада је за то довољан свега један.

Зашто је последњи пример и реалан? По статистици *AIIM* (*Association for Information and Image Management*) старој три године, 72% великих организација и компанија имају три или више различита *ECM* система, док 25% има и пет или више [19]. Колико год то било нелогично, компаније и организације се могу наћи у оваквим ситуацијама из више разлога; понекад сваки сектор или огранак има свој *DMS*, који у зависности од извршних задатака сектора, има потребу за *DMS*-ом који се разликује од већине који се већ користе у компанији. Такође, како не постоји идеалан за све потребе, перформансе, прилагодљивост и ефикасност *DMS*-ова зависе од саме намене.

Годину дана након објављивања *CMIS* стандарда (2011), постојало је више од 50 имплементација [20], које су пре свега биле уско везане за неки од аспеката примене протокола, чиме се додатно успорило прихватање *CMIS*-а као диманичког решења. Ипак, данас постоји једна имплементација (*Apache Chemistry*) која се са оправданим разлогом највише користи, а број људи који свакодневно размењује документе преко *CMIS* протокола је у свакодневном расту, као и број компанија које су решиле да га користе, па је једини проблем који изискује моментално решење, Проблем очувавања метаподатака.

3. ОЧУВАВАЊЕ МЕТАПОДАТАКА

Једна од суштинских разлика између уобичајених оперативних система и *DMS*-ова је да сваки документ или директоријум једног система за управљање документима, може имати произвољан број жељених метаподатака. Уколико је неопходно да се одређени објекат, са свим својим метаподацима, премести на неки други *DMS* уз помоћ *CMIS* протокола (нпр. у новинарству, објављивање приватних чланака на јавно доступну веб апликацију или рецимо, архивирање докумената неке радне групе), ту настају извесни проблеми.

Као што је споменуто раније, *CMIS* (односно његова имплементација) не мапира, тј. не чита баш све метаподатке неког

документа које би требало, па их самим тим не може ни проследити *DMS*-у примаоцу. Величина и значај овог Проблема је описана у претходном поглављу, а чекање да се Проблем сам реши (другим речима, да у некој од идућих верзија имплементације, Проблем буде ревидиран и комплетно решен) се показало као неделотворно. Стога, развојни тимови корисника који су се одлучили за употребу *CMIS* стандарда, морају развити нека међурешења која би ублажила или неутралисала Проблем док не буде отклоњен у некој од будућих верзија имплементације.

На следећим страницама ће бити детаљно образложени проблеми и неколико могућих, теоретски изводљивих, привремених решења за очување метаподатака који настају приликом трансфера докумената преко *CMIS* протокола. Решења ће бити међусобно упоређивана и једно од њих ће бити изабрано за реализацију имплементације, са пропратним објашњењем самог избора.

3.1. Могућа решења проблема

Као што је већ споменуто, свака установа, компанија или предузеће које користи неки од управљачких система, у зависности од пословне логике, има сопствени, специфични модел података. Такође, модели података сваког *DMS*-а су другачије дефинисани, типизирани и конфигурисани. Одавде се лако закључује да се може десити извесни конфликт приликом трансфера података преко, не баш перфектног, *CMIS* стандарда услед премошћавања свих тих разлика. Прецизније речено, у најблажем случају, конфликт се може испољити у изостанку очувавања свих метаподатака приликом трансфера, а у најгорем - у прекиду процеса трансфера. До конфликта одређеног обима ће сигурно доћи уколико су *DMS*-ови разноврсни или ако им се модел података, у било ком смислу - разликује, барем у оним типовима објеката који се шаљу/примају. Уколико, пак, не дође до конфликта, и даље преостаје проблем очувавања метаподатака, објашњен у одељку 2.4.

Разликоваћемо два општа случаја проблемске ситуације: први, када је модел података на страни пошиљаоца идентичан моделу података на страни примаоца и други, када се два наведена модела података репозиторијума *DMS*-а разликују. Први случај се у пракси виђа најчешће унутар једног малог или средњег предузећа, зато што је, иако можда на први поглед нелогично, тешко и неефикасно одржавати све моделе података идентичним у било ком пословном сценарију. Ипак, како фирма расте и прераста у компанију или установу, тако и њен софтвер (*DMS*) еволуира и модел података се мења, развија и грана, у зависности од потреба и пословних процеса и ту је реч о другом случају.

У првом општем случају, решења за све проблеме и конфликте могу бити имплементирана уз помоћ привремених (енгл. *temporary*) докумената. Ако су у трансфер укључени и ови документи, пракса је или да се сви неопходни подаци и команде сместе у тачно један или

да су генерисани (по потреби, од 0 до n) привремени документи у уској вези са документима који могу да изгубе метаподатке. Овај приступ игнорише (мањкаво) имплементирано мапирање *CMIS* протокола и сви метаподаци који би се иначе изгубили се уписују у привремене документе, који се, опет, шаљу заједно са свим преосталим објектима. Уколико је метода очувавања метаподатака уз помоћ привремених докумената (у првом општем случају) непожељна - информације, наредбе и сами изгубљени подаци се могу уписати у неки, за то предвиђени метаподатак или пак, сам садржај неког документа.

У другом општем случају, решења за проблеме и конфликте не могу бити једноставна, али могу бити имплементирана тако што ће се мапирати комплетан модел примаоца. Затим се слање података прилагођава мапираном моделу или се *DMS* прималац прилагођава моделу пошиљаоца. Такође, могуће решење је и серверска апликација која игра улогу посредника (слика број 4). Која год се стратегија изабрала, решење свакако не може бити тривијално; модел података није тек тако приступачан из сваког управљачког система и свакако ће бити неопходно инсталирање неког алата, плагина (*plug-in*), додатка (*add-on*) и слично.

У следећем одељку, предложени типови решења ће бити подробније објашњени и међусобно упоређени по одређеним критеријумима.

3.2. Упоредивање могућих решења

Свако од споменутих метода решења у претходном поглављу има своје предности и недостатке. Нека су динамичка и боља, али и скупља за реализацију са нимало занемарљивим временом израде. Друга су статичка, самим тим и лошија, али и она могу наћи своју примену, посебно зато што су јефтинија и имплементација се одвија брже. У следећим редовима ће бити представљено неколико примера теоретски могућих решења, са њиховим особинама.

Када је реч о идентичним моделима података *DMS*-а пошиљаоца и примаоца, једна од карактеристичних статичких метода подразумева слање привремених докумената који чувају (мета) податке, извршне наредбе, коментаре или било шта што је од суштинског значаја за сам трансфер и очување метаподатака објеката послатих преко *CMIS* протокола. Креирање и слање привремених докумената је јефтино решење, лако је за имплементацију и релативно брзо се извршава, уколико број објеката (директоријума и докумената) који се шаљу није огроман. Ипак, неке конвенције морају бити донете, попут генерисања имена привремених докумената и њиховог садржаја, но највећи посао представља имплементација механизма који чита ове документе, обрађује њихове садржаје, извршава пропратне акције и брише привремене документе. Неопходно је споменути да се све будуће промене модела података морају опрезно изводити, јер је метода

статичка и, да би исправно радила, мора бити синхронизирана са моделом. А ево и два конкретнија примера у вези са овом методом:

ПРИМЕР 1: Трансферу података се придружује тачно један привремени документ

Идеја је да се уобичајеном трансферу података преко *CMIS* протокола дода један привремени документ, који би садржао све информације (свих објеката) које ће се изгубити у самом трансферу. По завршетку размене података, привремени документ бива прочитан и сви метаподаци и остале информације се уписују тачно на оне објекте који су их изгубили, након чега се привремени документ брише. Овај пример функционише веома брзо и звучи привлачно, али у неким случајевима би се парсирало и по хиљаде линија привременог документа, а осим самог метаподатка и његове вредности, у њега се мора уписати и путања до објекта и можда још неки неопходни детаљи (тип метаподатка, кардиналност итд).

ПРИМЕР 2: Трансферу података се придружује онолико привремених докумената, колико је неопходно

Са сваким документом и директоријумом који је укључен у трансфер преко *CMIS* стандарда (наравно, уколико се неки метаподатак губи у процесу) се шаље и један привремени документ. Дакле, ово значи да уколико шаљемо са једног управљачког система на други, N ($N > 0$) објеката, у најгорем случају би било послато $2 * N$ објеката (укључујући привремене документе). Иако број привремених докумената теоретски може бити тако велики, у пракси је, свакако, приметно мањи, а и с обзиром да се у грубом просеку, сваки седми, осми, десети метаподатак неког објекта изгуби, ни њихова садржина неће достизати проблематичну величину. Са унапред договореним конвенцијама давања имена привременим документима, неће бити неопходно уписивање путање до објекта коме је неопходно успоставити изгубљени метаподатак, што чини парсирање и саму имплементацију лакшом него у претходном примеру, али ипак је неопходно послати изванредан број привремених докумената.

Када су два модела података синхронизирана, није неопходно уписивање информација које се губе у неки привремени документ, већ се могу проследити и на неки други начин. Следећа два примера су управо таква, теоретска решења:

ПРИМЕР 3: Метаподаци који ће бити изгубљени у трансферу неког објекта се уписују у неки предефинисани метаподатак самог објекта

Замисао је да се проблематични метаподаци упишу у неки предефинисани, динамички додат текстуални метаподатак на самом објекту. На тај начин, метаподатак ће сигурно бити мапиран од стране *CMIS*-а и сачуван у самом трансферу. Као и у претходним примерима, подаци из предефинисаног метаподатка ће бити парсирани и обрађени, након чега се динамички уклања метаподатак и успоставља стање какво је било у *DMS*-у пошиљаоцу. Ово је најелегантнији и веома брз, природан начин рада, али и даље се парсирање и конвенције не могу избећи, а и не пружају сви

Математички факултет у Београду, Катедра за рачунарство и информатику

управљачки системи подршку за споменуто ефикасну динамику преко *CMIS* протокола.

ПРИМЕР 4: Метаподаци који ће бити изгубљени у трансферу неког објекта се уписују у сам садржај

Једина разлика овог примера од претходног је што се за складиштење проблематичних метаподатака користи сам садржај неког документа у који се дописују подаци, одвојени јединственим знацима за парсирање. Иако је ово вероватно најбржа операција од досадашњих примера и процес не зависи од подршке за динамичку интеракцију *DMS*-ова и привремених докумената, парсирање је на још опаснијем нивоу него у пређашњим примерима, јер уколико дође до евентуалне грешке, сам садржај документа ће бити изгубљен, а проблематични метаподаци директоријума су у директној вези са неким документом, што није препоручљиво.

Ипак, у неупоредиво више случајева, модели података никако неће бити идентични. Компликованији случај подразумева и већи број конфликта и проблема који могу настати, па су самим тим неопходна и комплекснија решења од до сад споменутих. Следећа четири примера покривају управо тај сценарио.

ПРИМЕР 5: Модел података *DMS*-а примаоца се пре трансфера прво чита и обрађује, а након тога се испоручују прилагођени (мета)подаци

Читање и обрада самог модела података своди се на анализу, односно истраживање података (енгл. *data mining*). Наиме, када се преузме модел података *DMS*-а примаоца, извршава се истраживање података модела, након чега се објекти за слање шаљу у оне типове података који су највероватнији резултат предвиђања истраживања. Брзина рада овог решења долази до изражаја када се шаљу огромне количине података, јер се предвиђања врше само иницијално. Такође, никакво парсирање није укључено у трансфер и ова метода може дати бриљантне резултате. Но, с обзиром да је у игри одређена вероватноћа, може дати такође и катастрофалне, посебно ако се два модела података прилично разликују. Такође, примена овог решења мора заобићи *CMIS* објекте веза, јер је врло мало вероватно да ће ово предвиђање дати конзистентне резултате.

ПРИМЕР 6: Модел података *DMS*-а пошиљаоца се шаље и интегрише у модел података примаоца пре самог трансфера

Овај концепт је базиран на слању мини модела података пре самог трансфера података преко *CMIS* протокола. Мини модел је у ствари само неопходни, али довољни подскуп модела података управљачког система који игра улогу пошиљаоца. Такође, са њим се шаљу и конфигурациони подаци који чувају подешавања о томе како ће се пребачени објекти приказивати преко графичког интерфејса корисницима *DMS*-а. Трансфер података се одвија на следећи начин: одређивање минималног неопходног и довољног подскупа модела података *DMS*-а пошиљаоца, одређивање неопходних конфигурационих података за графички интерфејс, слање мини

Математички факултет у Београду, Катедра за рачунарство и информатику

модела и конфигурационих података (и њихова интеграција са системом примаоца) и последњи корак - уобичајен трансфер података преко *CMIS* стандарда.

Прва два корака представљају нетривијалан изазов за развојне тимове, а и слање мини модела није тривијално. Проблем који би настао током времена је и питање одлуке - шта радити са бројним мини моделима?

ПРИМЕР 7: Модел података DMS-а пошиљаоца се шаље и интегрише у модел података примаоца само приликом првог трансфера

Ова идеја је мање-више иста као претходна и разлика се огледа у томе да у овом случају нема мини модела, већ постоји само онолико додатних модела колико је различитих система у интеракцији. Модели овог случаја су приступачнији, могу се динамички ажурирати и, практично, њихов број је ограничен. Претходни пример је бољи уколико у реткој комуникацији учествује велики број различитих управљачких система, а текући је бољи уколико у фреквентној интеракцији учествује неколико различитих *DMS*-ова.

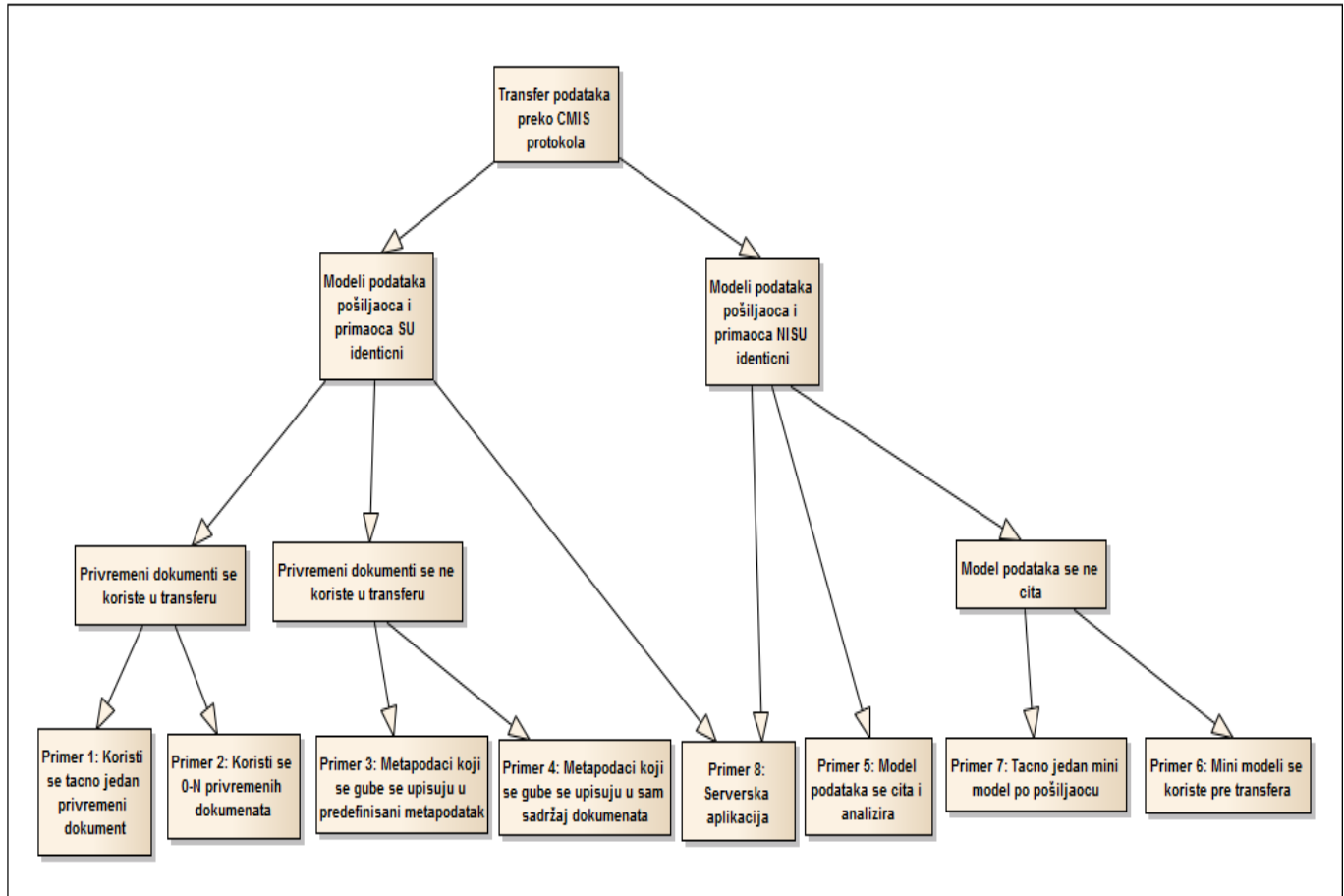
ПРИМЕР 8: Серверска апликација која се стара о решавању свих конфликта и проблема који настају приликом трансфера података

Ово решење је графички представљено на слици 4. Дакле, реч је о серверској апликацији која ће бити предефинисана да динамички комуницира са било којим управљачким системом и да исто тако буде посредник у комуникацији. Апликација би интегрисала све предности претходних решења уз минималне недостатке и очекивани профит на тржишту. Но, најкомплексније решење захтева да се највише времена посвети имплементацији, а главни недостатак је брзина трансфера. Јасно је да је трансфер успорен ако се одвија преко посредника, па је ово решење погодно уколико брзина преноса података није пресудна.

Оно што је заједничко за све сценарије је испољавање Проблема; проблема који преостаје услед недовољно доброг мапирања метаподатака пре самог процеса слања објеката преко *CMIS* протокола.

Не постоји листа управљачких система код којих се Проблем испољава, нити се овај рад бави таквом анализом. Ипак, оно што је дефинитивно је да је Проблем уочљив у Алфреско *DMS*-у, а готово је сигурно да постоји и у већини других система.

Дакле, да сумирамо, Проблем је само један у низу могућих практичних проблема, али се испољава у свим пословним сценаријима и случајевима и због тога је од велике важности да се у некој од следећих верзија протокола отклони, а до тад би се вероватно користила нека од привремених решења. Класификација примене наведених решења, у зависности од модела података, пословне логике и стратегије се може видети на слици 5.



Слика 5: Домен примена решења у зависности од модела података и стратегије, дрво приказа.

И на крају, следећа табела међусобно упоређује свих осам анализираних решења преко неколико критеријума. Параметри поређења су пажљиво одабрани; иако нису сви параметри истог приоритета ни значаја, унија потврдних одговора на те захтеве чини већину скупа жељених особина свеопштег решења. Овај преглед служи као категоризација и рекапитулација проблематике и сценарија, преглед предности и недостатака свих предложених решења, али и као увод за следећи одељак, где ће бити објашњени разлози за одабир одређеног решења за имплементацију.

<u>Пример/ предности</u>	<u>Пр. 1</u>	<u>Пр. 2</u>	<u>Пр. 3</u>	<u>Пр. 4</u>	<u>Пр. 5</u>	<u>Пр. 6</u>	<u>Пр. 7</u>	<u>Пр. 8</u>
<u>Балансирана брзина трансфера</u>	✗	✗	✓	✓	✓	✓	✓	✗
<u>Привремени документи нису укључени</u>	✗	✗	✓	✓	✓	✓	✓	✓
<u>Модел података се не мења/шаље</u>	✓	✓	✓	✓	✓	✗	✗	✗

<u>Операције парсирања су минималне</u>	✗	✓	✗	✗	✓	✓	✓	✓
<u>Пример функционише и за неидентичне моделе</u>	✗	✗	✗	✗	✓	✓	✓	✓
<u>Садржај докумената је непромењен</u>	✓	✓	✓	✗	✓	✓	✓	✓
<u>Додаци и плагинови нису неопходни</u>	✓	✓	✓	✓	✗	✗	✗	✗
<u>Непотребни помоћни документи се бришу</u>	✓	✓	✓	✓	✓	✗	✓	✓
<u>Додатни механизам код примаоца није неопходан</u>	✗	✗	✗	✗	✓	✓	✗	✓
<u>Пример је веома поуздан у сваком сценарију</u>	✓	✓	✓	✗	✗	✓	✓	✓
<u>Пример је флексибилан</u>	✗	✗	✗	✗	✗	✗	✗	✓
<u>Пример је лако проширив</u>	✗	✗	✗	✗	✗	✗	✗	✓

Табела 1: Преглед и поређење представљених примера кроз подршку за изабране критеријуме.

3.3. Образложење избора решења за имплементацију

CMIS протокол је од великог значаја за садашњост и будућност рачунарства у облаку. Ипак, у зависности од пословне логике и ситуације у којој се протокол користи, извесно је да ће се испољити неки проблеми или чак и конфликти, који ће онемогућити успешан трансфер (мета)података. У прошлим поглављима је било речи о конфликтима, проблемима, али и неким могућим решењима која су само (као и сам стандард) теоретска - дакле, неопходно их је имплементирати.

Нека наведена решења су боља и ефикаснија, нека лошија, спорија и скупља. Сценарији трансфера података између управљачких система који имају различите моделе података су најкомпликованији и самим тим их морају развијати тимови програмера, а не појединац. Последњи споменути пример односи се

на комплексну, серверску апликацију која би, за ефикасну глобалну употребу повезивала четрдесетак најпознатијих управљачких система, што представља огроман напор у сваком смислу. Због тога се за брзу, практичну имплементацију и демонстрацију (не)моћи *CMIS*-а, морају узети два *DMS*-а са истим моделом података. Пратећи слику 5 и приказано дрво, може се изабрати стратегија трансфера са или без привремених докумената који чувају изгубљене метаподатке. Слање додатних докумената не делује ефикасно и привлачно, али проблем са примером број 4 је што је неопходно мењати садржај докумената, што је веома ризично за резултате који се добијају том методом, док у примеру број 3 би морао да се мења сам модел података, парсирање није олакшано, а уз то је неопходно мењати саме метаподатке објекта који су у трансферу. Но, привремени документи су устаљена програмерска пракса за аналогну проблематику, па се некако логично намећу као избор. Једина преостала дилема је да ли решење Проблема очувавања метаподатака реализовати са једним или више привремених докумената.

Који год да је одабир начињен, садржај самог или свих привремених докумената (у случају да их је више) је сличан. Предност избора једног привременог документа је креирање и упис у само један додатни документ, али је истовремено неопходно уписивање више информација о метаподацима и процес претраге и парсирања дуже траје.

Но, уколико из неког разлога ипак не буде могуће очување метаподатака, лепо би било када би ти изгубљени метаподаци могли да се виде као сачувани привремени документи на нивоу самог објекта, а не на једном, глобалном месту, јер то може бити веома непрегледно. Такође, када постоји више привремених докумената, механизам очувавања метаподатака код примаоца може у току самог трансфера да обрађује прослеђене информације, док у случају једног привременог документа, процес слања мора да се заврши, па се тек након тога успостављају изгубљени (мета)подаци.

Због свих претходних разлога, у следеће две главе ће имплементација благо модификоване верзије решења број 2 бити објашњена из аспекта одабраног управљачког система и протокола, уз осврт на сам модел података и сценарио примене.

4. ТЕХНИЧКИ АСПЕКТ РЕШЕЊА

У поглављу 2, „Технологија рада са електронским документима“, представљен је концепт управљачког система, *Document Management System*-а, затим и Алфреско *DMS* и *CMIS* протокол. То је било неопходно како би се објаснила суштина Проблема, представила нека од решења за њега и на крају - како би примена свих технологија могла адекватно да се представи. Како је одабрано конкретно решење за имплементацију, тако је сада неопходно изабрати и одређену верзију Алфреско *DMS*-а, као и

Математички факултет у Београду, Катедра за рачунарство и информатику

имплементацију (и њену верзију) *CMIS* стандарда за сценарио примене. Потребно је предложити и експлицитну спецификацију структуре технологија које ће се користити у имплементацији.

Следећи одељци ће бити посвећени управо аспектима структуре изабраних технологија. Изабрано решење ће бити подробније објашњено, и то у смислу односа према одабраним технологијама.

4.1. Проблеми у вези са имплементацијом *CMIS*-а

За потребе овог рада, неопходно је изабрати конкретну имплементацију стандарда. Као што је раније наведено, трансфер електронских докумената преко *CMIS* протокола се одвија између два Алфреско *DMS*-а. С обзиром на то да је Алфреско реализован у програмском језику *Java*, пожељна је имплементација *CMIS*-а у баш том програмском језику. Постоје десетине имплементација стандарда, али најприхваћенију и најбољу је развила фирма *The Apache Software Foundation*. Стога се за израду решења очувања метаподатака користи верзија имплементације *CMIS* протокола у програмском језику *Java*, фирме *Apache*, односно *Chemistry* пројекат - ***OpenCMIS***. Најновија тренутна верзија је 0.12.0, објављена 16.8.2014. године, а она подразумева скуп алата и библиотека чиме је, условно речено, имплементирана верзија 1.0 и започет рад на развоју реализације верзије 1.1 стандарда. Заправо, имплементација *CMIS* стандарда поседује неке од недостатака описаних у одељку 2.3, а *OpenCMIS* се не стара ни за Проблем очувавања метаподатака.

У чему се тачно огледа Проблем? Пре него што се одговори на ово питање, неопходно је анализирати детаљно структуру самог протокола.

Спецификација *CMIS* стандарда описује модел података, сервисе и начин повезивања који мора бити подржан од стране свих серверских апликација које пружају подршку за рад са стандардом. Ово је заједничка особина свих таквих апликација јер је то једини начин да се успостави динамичко решење које *CMIS* пружа, попут креирања докумената и директоријума, претраге и навигације кроз репозиторијуме или пак, верзионисања докумената и постављање дозвола над њима.

Изабрано динамичко решење је омогућено уз помоћ ***Restful AtomPub*** повезивања. Уз помоћ овог протокола је могуће извршити *HTTP* захтеве над репозиторијумом, и то: *GET*, *POST*, *PUT* и *DELETE*. Но, у било ком тренутку, уз помоћ сервиса *getRepositoryInfo* могуће је добити податке о конкретном репозиторијуму.

CMIS у себи пружа везивања и са АПИ-јем *SOAP* веб сервиса, али иако ти веб сервиси протокола пружају већи скуп функционалности, сам протокол је старији, а како су и сви напори развојних тимова наклоњени управо ширењу *REST* АПИ-ја, у наставку рада ће бити речи само о раду са *Restful AtomPub* протоколом.

Одговори на захтеве упућене преко *Restful AtomPub* протокола су, скоро без изузетка, увек у форми *Atom XML*-ова и подразумева се

да су неки од тагова специфични за *CMIS* протокол. За парсирање овако добијеног резултата, могу се користити *Atom* библиотеке, генерички *XML* парсери, али и својеручно написани програми.

Сервиси интегрисани у *CMIS* деле неке заједничке карактеристике, попут следећих:

- Скоро сваки сервис враћа као резултат неки тип листе.
- Сервиси који враћају *CMIS*-овски објекат, такође враћају и додатне информације о самом објекту, попут *CMIS* веза, листе овлашћења и дозвољених акција над објектом, као и начин презентације самог објекта.
- Сваки сервис може да врати грешку (енгл. *exception*).
- Уз објекат као резултат, сервис може да приложи и токен промене, како би се смањила вероватноћа да се ажурира неки објекат који је застарео. Наравно, ово се само подразумева у случају када се сервису претходно проследи токен промене.

Неки од сервиса које *CMIS* имплементација у себи мора да садржи су и:

- **Репозиторијумски сервиси.** Користе се за извлачење информација и могућности неког репозиторијума.
- **Навигациони сервиси.**
- **Сервиси за рад са објектима.** Служе за креирање, читање, ажурирање и брисање објеката.
- **Сервиси за вишеструко попуњавање.** Овај сервис је значајан уколико је неки објекат неопходно креирати истовремено у неколико директоријума.
- **Истраживачки сервиси.** Скоро сви упити претраге се реализују уз помоћ њих.
- **Сервиси за верзионисање докумената.**
- **Сервиси за везе.** Користе се када је неопходно извршити упите у потрази за објектима који су у *CMIS*-овској вези са неким другим објектима.
- **Сервиси за политике.** Служе за примену, отклањање и претрагу неке административне политике.
- **Сервиси за привилегије.** Овај скуп сервиса служи да управља привилегијама неког објекта, али такође могу да врате као резултат и привилегије неког објекта.

Математички факултет у Београду, Катедра за рачунарство и информатику

```
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:app="http://www.w3.org/2007/app" xmlns:cmisra="http://docs.oasis-open.org/ns/cmisa/restatom/200908/"
xmlns:cmis="http://docs.oasis-open.org/ns/cmisa/core/200908/" xmlns:alf="http://www.alfresco.org" xmlns:opensearch="http://a9.com/~spec/opensearch/1.1/">
<author> <name>System</name> </author>
<generator version="3.2.0 (r2 @build-number@)">Alfresco (Community)</generator>
<icon>http://localhost:8080/alfresco/images/logo/AlfrescoLogo16.ico</icon>
<id>urn:uuid:4eb6a431-3c56-4767-816a-4ceca2295ae2-children</id>
<link rel="service" href="http://localhost:8080/alfresco/s/cmisa" />
<title>Company Home Children</title>
<updated>2009-11-13T13:14:29.863-06:00</updated>
<opensearch:totalResults>11</opensearch:totalResults>
<opensearch:startIndex>0</opensearch:startIndex>
<opensearch:itemsPerPage>-1</opensearch:itemsPerPage>
<cmisra:numItems>11</cmisra:numItems>
<entry>
<author> <name>System</name> </author>
<content>8cd47010-0465-4e94-8a9d-e34417e63591</content>
<id>urn:uuid:8cd47010-0465-4e94-8a9d-e34417e63591</id>
<published>2009-11-13T13:14:40.843-06:00</published>
<summary>Site Collaboration Spaces</summary>
<title>Sites</title>
<updated>2009-11-13T13:14:40.869-06:00</updated>
<app:edited>2009-11-13T13:14:40.869-06:00</app:edited>
<alf:icon>http://localhost:8080/alfresco/images/icons/space-icondefault-16.gif</alf:icon>
<cmisra:object>
<cmis:properties>
<cmis:propertyId propertyDefinitionId="cmis:allowedChildObjectTypeIds" />
<cmis:propertyId propertyDefinitionId="cmis:objectId"> <cmis:value>F:st:sites</cmis:value> </cmis:propertyId>
<cmis:propertyString propertyDefinitionId="cmis:lastModifiedBy"> <cmis:value>System</cmis:value> </cmis:propertyString>
<cmis:propertyString propertyDefinitionId="cmis:path"> <cmis:value>/Sites</cmis:value> </cmis:propertyString>
<cmis:propertyString propertyDefinitionId="cmis:name"> <cmis:value>Sites</cmis:value> </cmis:propertyString>
<cmis:propertyString propertyDefinitionId="cmis:createdBy"> <cmis:value>System</cmis:value> </cmis:propertyString>
<cmis:propertyId propertyDefinitionId="cmis:objectId">
<cmis:value> workspace://SpacesStore/8cd47010-0465-4e94-8a9de34417e63591</cmis:value>
</cmis:propertyId>
<cmis:propertyDateTime propertyDefinitionId="cmis:creationDate"> <cmis:value>2009-11-13T13:14:40.843-06:00</cmis:value> </cmis:propertyDateTime>
<cmis:propertyString propertyDefinitionId="cmis:changeToken" />
<cmis:propertyId propertyDefinitionId="cmis:baseTypeId"> <cmis:value>cmis:folder</cmis:value> </cmis:propertyId>
<cmis:propertyDateTime propertyDefinitionId="cmis:lastModificationDate"> <cmis:value>2009-11-13T13:14:40.869-06:00</cmis:value> </cmis:propertyDateTime>
<cmis:propertyId propertyDefinitionId="cmis:parentId"> <cmis:value>workspace://SpacesStore/4eb6a431-3c56-4767-816a-4ceca2295ae2</cmis:value> </cmis:propertyId>
</cmis:properties>
</cmisra:object>
<cmisra:pathSegment>Sites</cmisra:pathSegment>
</entry>
```

Слика 6: Пример *Atom XML* одговора на захтев, упућени преко *Restful AtomPub* протокола.

На слици 6, може се видети скраћени пример *Atom XML*-а који је резултат извесног *CMIS* захтева (упита) преко *Restful AtomPub* протокола на Алфресковом репозиторијуму. Овде се јасно види да су тагови *XML*-а састављени од *Atom*-целина и дефиниција, уз додатак у виду *CMIS* и Алфреско надоградњи. Сви тагови који почињу са „**cmis:**“ су део *CMIS* надоградње. На пример, таг: **<cmis:propertyDateTime**

propertyDefinitionId="cmis:creationDate">, дефинише *CMIS*-овски метаподатак „датум креирања“, док његов унутрашњи таг **<cmis:value>**, чува саму, прочитану вредност тог метаподатка на Алфресковом репозиторијуму, за тражени захтев.

Тек сада се може и прецизно дефинисати шта је заправо проблем у Проблему очувавања метаподатака. Наиме, сви наведени недостаци *CMIS* протокола (мапирање свега два родитељска типа докумената и директоријума, контрадикторна правила различитих репозиторијума, одсуство подршке за рад са корисничким метаподацима, недовршена имплементација протокола), имају као последицу раније описан Проблем - немогућност мапирања. Другим речима, када се *Restful AtomPub* протоколу упути захтев за информацијама неког објекта, чији скуп информација садржи неке које директно рефлектују недостатке *CMIS*-а, тада резултат захтева у *Atom XML* облику неће уопште садржати тагове који описују проблематичне информације.

Овим се одржава стабилност и конзистентност (у случају конфликта), али се проблематичне информације/метаподаци - не прослеђују у резултату упита, односно, не мапирају се. Уколико се *CMIS* протокол користи за трансфер објеката између два *DMS* репозиторијума и ако се игнорише примена решења описаних у претходној глави, тада ће доћи до неповратног губитка информација на новонасталим објектима у *DMS*-у примаоцу. Конкретан пример за овај случај ће бити представљен у одељку 5.3, где ће се описати процес тестирања извесног сценарија слања електронских докумената преко *CMIS* протокола, без примене било ког решења за све недостатке протокола.

4.2. Детаљи и проблеми у вези са изабраном верзијом Алфреска за имплементацију

За ефективну примену *CMIS* стандарда, неопходно је одабрати платформу (или више њих) која пружа подршку за његов рад. Алфреско је управо једна таква платформа, тј. управљачки систем са придруженим репозиторијумом.

Ипак, за саму примену стандарда, мора да се изабере нека, конкретна верзија Алфреско *DMS*-а. Критеријум одабира је бесплатна, стабилна верзија, која не мора да пружа последње визуелне ефекте једне серверске апликације (јер то није приоритет овог рада), али мора да пружа подршку за рад са најновијом верзијом имплементације *CMIS* стандарда и да обухвата све системске исправке Алфреска у вези са контекстом *CMIS*-а који овај рад обухвата и, кроз пропратни сценарио, примењује.

С обзиром да је још у одељку 2.2 појашњено да ће у остатку рада бити једино речи о бесплатној верзији за заједнице (*community edition*), остаје само дилема да ли да се примена *CMIS*-а прикаже у радном модулу *Explorer* или *Share*.

Алфреско *Share* радни модул строго прати развој веб технологија, развојни тим Алфреска непрестано ради на његовим

новим верзијама, а и кориснички интерфејс је модернији. Насупрот томе, Алфреско *Explorer* је старији радни модул, развојни тим Алфреска га само одржава и оригинални кориснички интерфејс је лишен функционалности *jQuery* библиотека, што овакав интерфејс чини застарелим. Последња верзија за заједнице Алфреско *DMS*-а је 5.0.a, изашла 27.јуна 2014, али од верзије 4.0, тачније, од јануара 2012. године, Алфреско *Explorer* радни модул се више не развија, већ само одржава и временом ће вероватно бити одбачен, како од стране веб заједнице и корисника, тако и од стране самог Алфреска. Имплементација примене *CMIS* протокола ће се ипак одвијати у радном модулу Алфреско *Explorer*-а, и то у верзији 3.3.g, из два разлога. Прво, Алфреско *Share*, у најновијој верзији и верзијама уназад, не нуди стабилност у смислу свих суштинских, већ постојећих, функционалности у верзијама закључно са последњом пре верзије 4.0. И друго, фокус овог рада је пре свега на примени *CMIS* стандарда на стабилној верзији *DMS* система, а кориснички интерфејс није приоритет овог рада. У прилог овом избору иду и следеће чињенице:

- Функционалности се брже имплементирају у *Explorer* радном модулу.
- Изабрана верзија, иако стара скоро три године подржава рад са једном од најновијих верзија изабране имплементације *CMIS* протокола (*Apache Chemistry OpenCMIS*).
- Изабрана верзија је стабилна у смислу свих функционалности Алфреска неопходних за имплементацију примене протокола у овом раду.
- Ни у најновијој верзији Алфреско *Share* (5.0.a) нису отклоњени проблеми и конфликти који су од значаја за овај рад, било што се тиче самих Алфрескових функционалности или покривања недостатака *CMIS* протокола [21].

Дакле, у даљем наставку рада, Алфреско *DMS* подразумева Алфреско *ECMS*, верзију за заједнице, модул за рад *Explorer*, верзија 3.3.g.

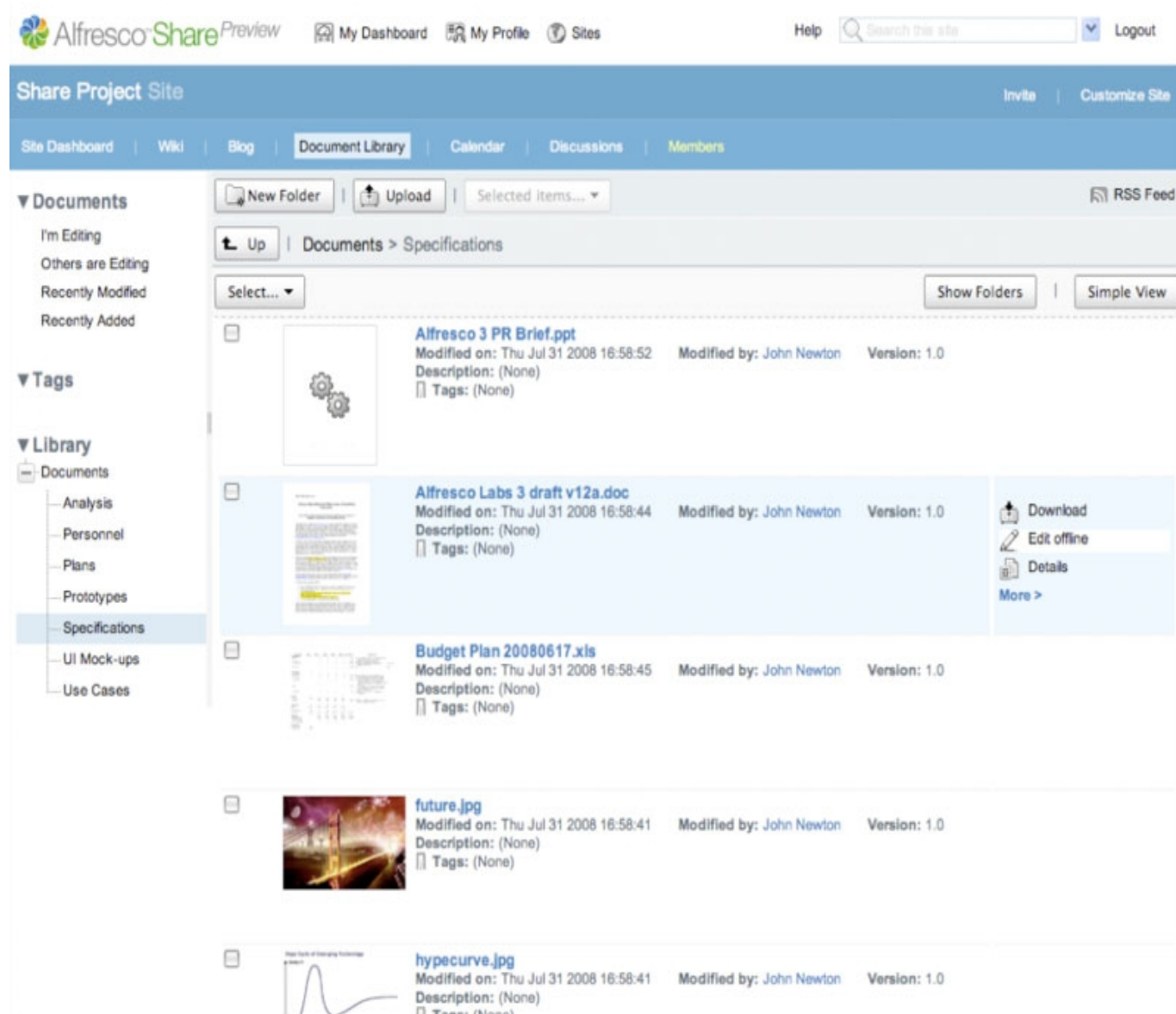
Пре него што се крене са описима проблема изабраног управљачког система, неопходно је детаљније објаснити саму његову структуру.

Multi-Tenancy карактеристике омогућавају Алфреско *DMS*-у да буде конфигуриран као јединствено окружење за вишеструке, независне корисничке апликације, на једној инстанци, која може бити инсталирана на једном серверу или на кластеру сервера. У исто време, та инстанца је логички партиционисана на такав начин да се појављује свакој корисничкој организацији или групи која јој приступа, као потпуно одвојена инстанца Алфреска [22]. Ова, као и многе друге позитивне особине Алфреско *DMS*-а, не би биле могуће без стабилне, али флексибилне структуре, која се лако може мењати, допуњавати и надограђивати.

Срж функционисања Алфреска чине *Java* зрна. За рад са Алфреском, неопходне су његове библиотеке: за *Java* (*Alfresco SDK*) и *Javascript* (*Alfresco Javascript API*). Природа програмског језика *Java*

Математички факултет у Београду, Катедра за рачунарство и информатику

пружа Алфреску подршку за рад на више платформи, а енкапсулирана зрна се могу мењати неким својеручно написаним зрнима. Исто важи и за ресурсне документе (серверских) веб апликација, попут мултијезичних ресурсних докумената, слика, извршних скрипти итд. Цео процес замене може бити аутоматизован уз помоћ неког алата, који је веома лако конфигурабилан, попут *Apache Ant*-а, било из командне линије или неког развојног окружења, попут *Eclipse*.



Слика 7: Поглед на репозиторијум Алфреско Share и садржај једног простора.

Скуп података у Алфресковом репозиторијуму чине **чворови** (енгл. *nodes*). Модел и подаци су тако дефинисани да сваки објекат, било да је реч о кориснику, директоријуму, документу, административном правилу (па и сама веза између чворова) и слично, чини неки чвор. Овај апстрактни концепт је уведен због лакшег наслеђивања; сви објекти имају неку заједничку особину,

попут имена, адресе у репозиторијуму или описа, па је концепт интерфејса и матичног чвора, у том смислу, задовољен, а принцип наслеђивања је исти као код објектно-оријентисаног програмирања.

Ипак, да би све наведено било упаковано у компактну целину, неопходни су документи који дефинишу и чувају структуру апликације, односно у овом случају, конфигурабилни XML документи. И они су такође замењиви и могуће их је проширити замислима сопствене пословне логике, јер свака апликација има своје захтеве у погледу структуре садржаја.

Осим дефиниција радних токова или уобичајених шаблона за различиту употребу, Алфрескови XML документи служе и за подешавања: аутентикације, *audit*-а, (аутоматизованих) акција над репозиторијумом, обавештења, сервиса, итд.

```
<type name="cm:folder">
  <title>Folder</title>
  <parent>cm:cobject</parent>
  <archive>true</archive>
  <associations>
    <child-association name="cm:contains">
      <source>
        <mandatory>false</mandatory>
        <many>true</many>
      </source>
      <target>
        <class>sys:base</class>
        <mandatory>false</mandatory>
        <many>true</many>
      </target>
      <duplicate>false</duplicate>
      <propagateTimestamps>true</propagateTimestamps>
    </child-association>
  </associations>
</type>

<type name="cm:content">
  <title>Content</title>
  <parent>cm:cobject</parent>
  <archive>true</archive>
  <properties>
    <property name="cm:content">
      <type>d:content</type>
      <mandatory>false</mandatory>
      <index enabled="true">
        <atomic>true</atomic>
        <stored>false</stored>
        <tokenised>true</tokenised>
      </index>
    </property>
  </properties>
</type>
```

Слика 8: Дефиниције чворова који представљају директоријуме и документе у *contentModel.xml*-у.

Један од свакако најзначајнијих је и **faces-config-custom.xml**. Суфикс „*custom*“ подразумева да је ово надограђена верзија оригиналног конфигурационог XML-а који се добија уз инсталацију

Математички факултет у Београду, Катедра за рачунарство и информатику

Алфреска. Овај документ је одговоран за дефиницију свих својеручно израђених *Java* зрна.

```
<config evaluator="string-compare" condition="Content Wizards">
  <content-types>
    <type name="my:sop" />
  </content-types>
</config>

<config evaluator="node-type" condition="my:sop">
  <property-sheet>
    <show-property name="mimetype" display-label-id="content_type"
      component-generator="MimeTypeSelectorGenerator" />
    <show-property name="size" display-label-id="size"
      converter="org.alfresco.faces.ByteSizeConverter"
      show-in-edit-mode="false" />
    <show-property name="my:publishedDate" />
    <show-association name="my:signOff" />
    <show-property name="my:authorisedBy" />
    <show-child-association name="my:processSteps" />
  </property-sheet>
</config>

<config evaluator="aspect-name" condition="my:imageClassification">
  <property-sheet>
    <show-property name="my:width"/>
    <show-property name="my:height"/>
    <show-property name="my:resolution"/>
  </property-sheet>
</config>

<config evaluator="string-compare" condition="Action Wizards">
  <aspects>
    <aspect name="my:imageClassification"/>
  </aspects>
</config>

<config evaluator="string-compare" condition="Advanced Search">
  <advanced-search>
    <content-types>
      <type name="my:sop" />
    </content-types>
    <custom-properties>
      <meta-data type="my:sop" property="my:authorisedBy" />
      <meta-data aspect="my:imageClassification" property="my:resolution" />
    </custom-properties>
  </advanced-search>
</config>
```

Слика 9: Пример конфигурирања приказа оригиналних Алфрескових чворова у `web-client-config.xml`-у.

Ипак, кључни конфигурациони документ је, свакако, **contentModel.xml**. Овај документ и сви остали чији се називи завршавају суфиксом „Модел“ (*forumModel*, *systemModel*, *tojNoviModel.xml*) формирају скуп који је одговоран за Алфресков **модел података**, односно, речник података (енгл. *Data Dictionary*). На слици 8 се могу видети дефиниције директоријума (*cm:folder*) и документа (*cm:content*) у оригиналном *contentModel*-у. Он и служи за дефинисање основних типова чворова, док надограђени модели

(*forumModel*, *systemModel*, *mojNoviModel.xml*) служе за дефинисање чворова изведених из основних типова.

Документ **web-client-config**, односно **web-client-config-custom.xml** пак, чувају дефиниције самог приказа. На слици 9 се могу видети неке од дефиниција (чаробњак за садржаје, класификација слика, чаробњак за акције, напредна претрага) из првог документа, који заправо служи за дефиниције изгледа графичког, корисничког интерфејса саме Алфреско апликације, док његова надограђена верзија (*web-client-config-custom*) дефинише само приказ садржаја надограђеног модела (*forumModel*, *systemModel*, *mojNoviModel.xml*).

Од велике је важности да се оригинални конфигурациони документи не мењају, већ ако је и нека промена у њима неопходна, она се мора унети у надограђеним моделима. Без оригиналних конфигурационих докумената, Алфреско би био мање-више само још једно складиште података и ни у ком смислу не би могао да разликује, рецимо тип „број“ од типа „датум“ [23].

Када је реч о моделу података, оно што је за овај рад (а и иначе) најбитније су типови података, у ширем смислу, аналогни са директоријумима и документима - оним што *CMIS* протокол мапира. На слици 8 су дате и њихове дефиниције у моделу, а типови изведени из њих се дефинишу на сличан начин. Оно што је заједничко за чворове Алфреска изведене из чворова „*folder*“ и „*content*“ (који представљају уобичајене директоријуме и документе на какве смо навикли) је скуп метаподатака који им се може придружити и он се састоји из различитих:

- **Особина** (енгл. *properties*).
- **Асоцијација** (енгл. *associations*).
- **Аспеката** (енгл. *aspects*).

Особина је једноставан метаподатак о чвору и најчешће је реч о неком текстуалном, бројчаном податку (опис, назив, идентификациони број) или пак, датуму, односно о некој листи таквих података.

Асоцијација представља показивач, референцу на неки други чвор у репозиторијуму са којом је чвор који је садржи у некаквој вези. У асоцијацији се разликују *извор* (енгл. „*source*“) и *мета* (енгл. „*target*“). Асоцијација је *једносмерна веза*, па (функционални и видљиви) подаци о асоцијацији се искључиво налазе на чвору извора. Другим речима, асоцијацију представљају три чвора - чвор извора, чвор мете и чвор саме везе, референце. Асоцијације над директоријумом или документом не морају бити искључиво са истим таквим типом, већ са било којим типом чвора, попут, рецимо корисника и деле се на обичне и асоцијације потомака. Ова подела је извршена због аутоматизма одређених функционалности које су у аналогји са каскадним операцијама у базама података. На пример, када се у Алфресковом репозиторијуму креира директоријум, сав његов будући садржај (у виду нових директоријума и докумената) ће бити у асоцијативној, потомачкој вези са њим. Ово је обавезно, јер омогућује аутоматизовано, каскадно брисање целокупног садржаја

директоријума када се **он** обрише. На слици 8 се могу видети примери дефиниција особина и асоцијације потомака за конкретан тип податка.

Аспекти чине скуп сличних особина и асоцијација удружених у једну целину, како из функционалних и организационих, тако и због визуелних, естетских разлога. Све горе наведено за особине и асоцијације важи и за особине и асоцијације унутар неког аспекта. Особине и асоцијације неког типа чвора (директоријума или документа) се морају навести у самој дефиницији и то је заједничко за све чворове тог типа, док се аспект у моделу података дефинише независно од било ког типа, чиме се постиже флексибилност у смислу да не морају сви чворови који припадају истом типу да поседују и исте аспекте, а како се исти аспекти са својим метаподацима могу придружити и различитим типовима чворова, ово је огромна предност Алфреска над обичним складиштем података. Такође, аспект може садржати у себи и неке друге (не)обавезне аспекте. Слика 10 приказује системски аспект који садржи једну особину и једну асоцијацију потомака.

```
<aspect name="cm:preferences">
  <title>Preferences</title>
  <properties>
    <property name="cm:preferenceValues">
      <type>d:content</type>
    </property>
  </properties>
  <associations>
    <child-association name="cm:preferenceImage">
      <source>
        <mandatory>>false</mandatory>
        <many>>false</many>
      </source>
      <target>
        <class>cm:content</class>
        <mandatory>>false</mandatory>
        <many>>false</many>
      </target>
    </child-association>
  </associations>
</aspect>
```

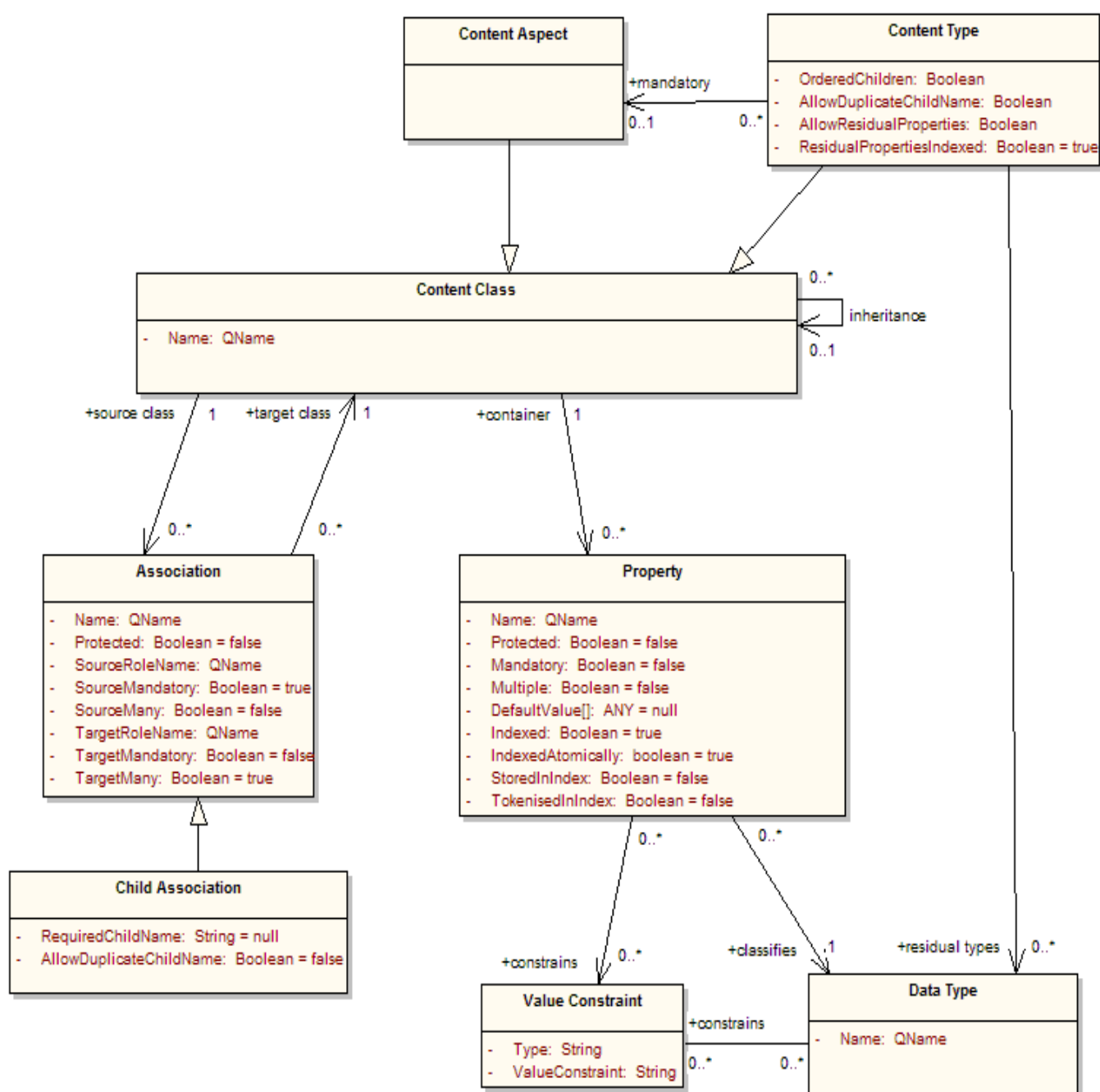
Слика 10: Пример дефиниције оригиналног Алфресковог аспекта „preferences“ у contentModel.xml-у

Како ово функционише на конкретном примеру? Дефиниција модела садржаја у Алфреску је слична *DDL (Data Definition Language)* дефиницији објеката у базама података и аналогно са тим, као што је очекивано да свака апликација има различит модел базе података, као и да је тај модел подложен променама, тако је и при дефиницији модела садржаја неопходно водити рачуна о могућности за флексибилно проширење и надоградњу [24]. На следећој слици је приказан дијаграм зависности, али и параметри подешавања типа садржаја, аспекта, асоцијације, објеката итд. Колико год се хијерархија и структура Алфреска надограђивала и мењала у

зависности од потреба апликације, ове зависности увек остају исте и на снази.

За потребе „интелигентног“ репозиторијума, Алфреско је додатно обогаћен и са још два веома битна својства: **понашањем** и **правилном**.

Понашање је *Java* зрно које обухвата цео репозиторијум и пасивно прати дешавања у репозиторијуму. Састоји се од услова и акције; када се услов испуни, акција се извршава. Услов може бити нпр. креирање неког типа документа, додељивање аспекта документу, а акција је, рецимо, подешавање наслова новокреираног документа или брисање садржаја директоријума.



Слика 11: Мета модел Алфресковог речника података у UML нотацији. Слика модела је преузета из рада наведеног у литератури [24].

Правило се подешава на нивоу директоријума и може рекурзивно утицати на комплетно стабло садржаја или само на директне потомке. Као и понашање, састоји се од услова и акције, с тим да је за разлику од понашања, акција овде готово увек дефинисана у *Javascript*-у. Одавде се може извући закључак да се за решавање неког проблема, у зависности од ситуације и пословне логике некад бира понашање, а некад правило. Сами програмски језици имплементације акције су различити, али и њихове Алфреско библиотеке нуде неидентичне скупове функција, па је јасно да се не може избећи један од два алата.

Напокон, сада, када је укратко објашњена структура одабране верзије Алфресковог репозиторијума, може се приступити проблемима који постоје у овој (и каснијим) верзији, а тичу се подршке у раду са *CMIS* протоколом.

Прва од ствари које се лако приметите је да Алфреско DMS игнорише недостатке CMIS протокола описане у претходном одељку и да пружа подршку за рад са њим, таквим какав је. Другим речима, чак и у трансферу директоријума и докумената између две инстанце Алфреско *DMS*-а се дозвољава губљење (мета)података о корисницима, у трансферу могу учествовати само типови чворова изведени из чвора директоријума или документа, а и након скоро пуне три године од изласка верзије 3.3.g, развојни тим Алфреска ни у једној верзији није покрио недовршену имплементацију *CMIS* стандарда! Списак недостатака које Алфреско игнорише укључује и поновно успостављање верзионисања, мултијезичности, кодних шема и *MIME* типова за документе.

Ипак, највећи проблем за трансфер података преко *CMIS* стандарда, о ком се (ни сам *CMIS*, а ни) Алфреско *DMS* не стара, јесу - **асоцијације**. У току самог трансфера, уколико извор или мета Алфреско асоцијације нису укључени у трансфер, асоцијација се неповратно губи. *CMIS* протокол мапира тип објеката звани - веза, што је аналогно са Алфресковом асоцијацијом, али није у стању да сачува прави показивач, везу или асоцијацију коју је мапирао, на неки ресурс који се не налази у самом *DMS*-у примаоцу. Такође, верзија 1.1 *CMIS* протокола подразумева постојање аспеката и њихово мапирање, а најновија верзија имплементације *OpenCMIS* 0.12.0, увелико то и омогућава, али...Уколико је асоцијација придружена аспекту, она уопште неће бити мапирана, њени показивачи ће бити неповратно изгубљени и ова чињеница свакако задаје главобољу развојним тимовима који раде на Алфреско *DMS*-у, а користе *CMIS* протокол. А ево и практичног примера.

Аспект „Аутори“ је дефинисан са једном асоцијацијом: „АуториДокумента“. Како неки документ може имати неодређен број аутора, асоцијација је прави избор и то асоцијација где је тај документ извор, а мета је неки корисник (односно, више њих) Алфреско *DMS*-а који ради на садржају тог документа. Уколико се

овакав документ проследи преко *CMIS* протокола, без решења за Проблем очувавања метаподатака, новонастали документ ће имати само аспект „Аутори“, без асоцијације и мете, јер *CMIS* и Алфрескова подршка за рад са протоколом не мапирају корисничке чворове, а такође ни асоцијације унутар аспекта. Дошло је до двоструког проблема за реалан, чест пример у пракси.

Мноштво споменутих проблема су минорни и ниског приоритета за Алфресков развојни тим, али као целина, никако нису занемариви, тим пре јер су ту присутни већ дужи временски период. Ипак, неке је неопходно решити на нивоу имплементације протокола, друге је боље решити на самом управљачком систему, али док се то не деси, неопходно је импровизовати брзо, функционално и елегантно решење (колико је то могуће) за Проблем очувавања метаподатака.

4.3. Детаљи одабраног решења, предности и недостаци

Као и за већину отворених проблема, ни за представљени проблем не постоји идеално решење. Чак и оно решење које је, гледано из функционалног аспекта, идеално, скупо је и за његову израду је неопходно највише времена. *CMIS* протокол је осмишљен тако да његове предности могу да искоришћавају како велике (програмерске) компаније и фирме, тако и мале фирме и појединци. Но, док се велике корпорације „не договоре“, у ком ће се тачно правцу одвијати развој и сврха стандарда и појединци се морају изборити са главним тренутним проблемима.

Решење за очување метаподатака приликом трансфера електронских докумената између два управљачка система уз помоћ привремених докумената је, највероватније, најлакше за разумевање. Много пута је опробано у програмерској пракси у разним ситуацијама, и данас се користи, па се и логично намеће за главну опцију. У овој конкретној ситуацији, као предности решења могу се навести следеће:

- **Не зависи од модела података.** Уистину, мора постојати механизам читања метаподатака у првом репозиторијуму управљачког система и механизам уписивања метаподатака у другом репозиторијуму, али привремени документи ће лако прихватити било какву конвенцију за било какав модел података из простог разлога јер садржај тих докумената може бити било шта, обичан текст, *CSV* документ, *XML* конфигурациони документ и слично.
- **Сам садржај докумената који се шаљу се не мења.** Интегритет докумената који се шаљу је ненарушен.
- **Нису потребни плагинови и други софтверски додаци.** Помоћни, привремени документи морају бити прочитани, обрађени и обрисани, али за имплементацију тих функционалности, довољан је и сам Алфреско АПИ.

- **Није потребно заузимање додатне меморије.** Привремени документи се након обраде бришу, па се додатна меморија на репозиторијуму не заузима. Ипак, ово се мора узети са резервом, јер због недостатака самог протокола, уколико је жеља да се сачувају изгубљени подаци, они се морају на неки начин сачувати.
- **Решење је поуздано.** Иако је неопходно водити доста рачуна о свим детаљима имплементације, поузданост решења у пракси је загарантована.
- **Решење је јефтино, брзо и лако за имплементацију.** За комплетан развој решења довољан свега један програмер.
- **Без последица по систем уколико дође до било какве грешке приликом трансфера.** Чак и ако се трансфер не имплементира као трансакција, довољно је само обрисати све креиране документе и покренути трансфер поново.

Ипак, понуђено решење (као и свако) има извесне недостатке. Они се огледају у:

- **Слању сувишних докумената.** Теоретски, у најгорем случају, број докумената који се шаљу може бити и удвостручен.
- **Брзина трансфера је успорена.** Ово је директна последица претходног недостатка. Привремени документи немају велику садржину, али ипак је неопходно извесно време да се они креирају и да се њихов садржај проследи и на крају - обришу, а како је сасвим могуће да их има велики број, трансфер у неким случајевима може трајати и до упола дуже.
- **Неизбежне операције парсирања.** У коме год да су формату привремени документи креирани, они морају бити прочитани и обрађени.
- **Неопходности постојања механизма за обраду.** Као последица претходног недостатка, иако није потребан никакав плагин или додатак, ипак је неопходан некакав механизам на репозиторијуму који прима податке, како би поново успоставио изгубљене метаподатке и релације.
- **Решење не ради за различите моделе података.** Решење је примењиво само када се два модела података поклапају на подскупу (мета)података који се шаљу. Генерално, ово јесте крупан недостатак, али у неким случајевима и није (нпр. сарадња између више одељења исте фирме, архива фирме и сл).
- **Решење није флексибилно и лако прошириво.** Уколико се пословна логика прошири или једноставно промени, мора се променити и логика решења. Ово мора бити јако деликатно и прецизно урађено.

Најпроблематичнији метаподаци за очување приликом трансфера у Алфреско *DMS*-у су **асоцијације**. Замисао овог решења је да се у привремене документе (који су најобичнији текстуални документи) упишу сви подаци о асоцијацијама, са јединственим идентификаторима. На овај начин, механизам у *DMS*-у примаоцу, у

зависности од идентификатора, зна о ком виду асоцијације је реч и самим тим зна како да поново успостави асоцијацију. Уколико је из неког разлога, немогуће успоставити асоцијацију, привремени документ који чува тај податак се не брише, како би корисници могли да виде која је асоцијација у репозиторијуму пошилаоца постојала. У супротном, након читања и обраде, привремени документ се брише.

Механизам за решавање *проблема* у Алфреско *DMS*-у примаоцу би укључио комбиновано Алфресково понашање и правило. Само понашање је довољно за успостављање асоцијација, али за постављање пар метаподатака који се неповратно губе - неопходан је АПИ *Javascript*.

Процес слања и примања докумената садржи још доста детаља, али они ће подробније бити објашњени у следећем поглављу.

5. ИМПЛЕМЕНТАЦИЈА

У претходним поглављима, било је речи о протоколу, окружењу, *проблеми* очувавања метаподатака, различитим решењима, као и њиховим дефиницијама, описима, упоређивањима са другим теоретским могућностима и слично. Наредна поглавља ће представити један практичан пример решења. Иако далеко од идеалног или најбољег, изабрани пример јасно осликава комплетну проблематику у пракси.

Овај рад прати имплементација, односно две инстанце Алфреско управљачког система које размењују садржај преко *CMIS* протокола. Пример представља део радног окружења једне биоинформатичке, истраживачке групе (прва инстанца, *DMS* пошилалац), као и њену електронску архиву (друга инстанца, *DMS* прималац). Комплетан сценарио и алгоритам архивирања, као и поступак архивирања са и без примене имплементационог решења биће представљени на следећим страницама.

5.1. Опис модела података и сценарија

У одељку 4.2 је представљен Алфресков модел података и начин како се креира нови, за потребе сопствене пословне логике. Имплементација која прати овај рад садржи две инстанце Алфреско *DMS*-а и обе имају идентичан модел података. Прва инстанца представља свакодневно, радно окружење - полигон за израду научних радова, пројеката једне биоинформатичке групе. Друга представља електронску архиву ове групе, складиште у које би се смештали радови и пројекти са прве инстанце који су застарели или више нису од великог значаја. Групу чине две биоинформатичке

Математички факултет у Београду, Катедра за рачунарство и информатику

подгрупе: подгрупа која се бави геномиком и подгрупа која се бави протеомиком.

Замисао пословне логике у овом раду подразумева чињеницу да биоинформатичка група има доста истраживача и сарадника, да је веома продуктивна и да, временом, репозиторијум продукционе инстанце постаје загушен количином докумената. Логична последица је измештање старијих, завршених или објављених пројеката и научних радова који се ређе користе на репозиторијум архивске инстанце. И ту на сцену ступа (не)савршени протокол *CMIS*. Управо његова имплементација се користи за ефикасан и безбедан трансфер докумената и (мета)података на архивску инстанцу Алфреско *DMS*-а.

Примена *CMIS* протокола у овом раду је осмишљена као архивирање електронских докумената споменуте научне групе. Прави начин да се представи *CMIS* протокол јесте његова примена у корелацији са Алфреско *DMS*-ом, док је за конкретнију примену у биоинформатичком окружењу, неопходно обогатити и само окружење управљачког система, чији опис следи.

Што се тиче корисника архивске инстанце, с обзиром на то да је и у пракси, уопштено речено, архива једно складиште не баш актуелних докумената, идеја је да постоји само један корисник, односно администратор (**admin**). С друге стране, продукциона инстанца Алфреско *DMS*-а је свакако богатија како садржајем тако и пословним процесима. Овде је идеја да постоје корисници који би припадали само једној од две биоинформатичке групе и корисници који би радили у обе. Исто тако, корисници подгрупа би били подељени у обичне и напредне кориснике. Представници обичних корисника (**clanGrupeGenomika**), би били студенти мастер студија чији рад покрива неку биоинформатичку проблематику, затим сарадници са других факултета и научних установа, истраживачи и слично, док би представници напредних корисника били асистенти, докторанти, професори, али и други аутори научних радова. И наравно, слично као и у архивској инстанци, и овде би постојао администратор (**admin**), али и руководилац целе биоинформатичке групе (**sefGrupe**) који би могао да буде супервизор свих кључних акција и дешавања на продукционој инстанци.

Имплементирана је комплетна вишејезичност обе инстанце у смислу промене активног језика саме Алфреско апликације. Ово подразумева да се активни језик апликације може променити иницијално или током рада, али не подразумева и промену језика метаподатака и садржаја. Последње споменуто је такође могуће, али је у Алфреско *DMS*-у то независна целина од активног језика апликације и да би то уопште било могуће променити, мора се у моделу података назначити да је неки метаподатак „вишејезичан“ и приликом уноса метаподатака или креирања докумената у овом управљачком систему, мора се унети више вредности за различите језике које систем подржава. У овом раду, омогућен је избор језика апликације и то су: енглески, српска латиница и српска ћирилица.

Сав садржај је на енглеском језику из разлога што се научни радови углавном пишу на овом језику.

Вратимо се сада на заједнички модел (мета)података обе инстанце. Сваки (управљачки) систем података има документе и директоријуме и они имају неке своје метаподатке, на пример: назив документа, датум креирања, креатор документа итд. У одељку 4.2 је детаљно објашњено шта заправо представља и шта чини један модел података Алфреско система. Имајући то у виду, имплементације које прате овај рад у моделу података садрже следеће својствене типове директоријума и докумената:

- Тип директоријума „**Група**“, који заправо представља матични, „*root*“ директоријум једне од две биоинформатичке подгрупе. Дефиниција метаподатака садржи број актуелних пројеката те групе.
- Тип директоријума „**Пројекат**“, који, како само име каже, представља пројекат, али онај тип пројекта или научног рада на коме ради нека (или обе) биоинформатичке подгрупе. Представља матични, „*root*“ директоријум за све ресурсе и документе који су (били) неопходни за израду или су у вези са самим пројектом, док се дефиниција метаподатака своди на два Алфреско аспекта: „*Аспект пројекта*“ и „*Аспект уопштеног пројекта*“.
- Тип документа „**Спољни Пројекат**“, који представља документ пројекта или научног рада на коме НИЈЕ радила биоинформатичка група, већ нека друга научна установа или група научника, истраживача и сл. Дефиницију метаподатака заокружује аспект: „*Аспект уопштеног пројекта*“ (детаљније на слици 12).
- Тип документа „**Долазни документ**“, који представља документ који је неким видом слања (поштом, електронски и сл) биоинформатичка група примила од неког појединца или научне групе. Ово могу бити различите преписке у виду мејлова, порука (са својим додацима) или уобичајени документи. Дефиницију метаподатака чини скуп информација попут: кључних речи, датума и начина пријема долазног документа, прилога итд.

Лако је уочљиво да су метаподаци код кључних типова директоријума обједињени у једну целину - Алфреско аспект. Ово је урађено како би ниво искористивости био што је могуће виши и како би се смањила величина самих дефиниција. Модел података Алфреско инстанци садржи још и три аспекта, од којих су два већ споменута:

- „**Аспект пројекта**“ представља скуп метаподатака који је уско везан за рад биоинформатичке групе на самом пројекту. Садржи информације попут: датума почетка и објаве пројекта/научног рада, контакт особу и преостале ауторе, нека евентуална запажања, као и подгрупу (или подгрупе) које су радиле на пројекту.

- **„Аспект уопштеног пројекта“** чине метаподаци о неком пројекту, где при томе пројекат не мора нужно бити пројекат на коме ради биоинформатичка група. Дакле, овај аспект носи са собом информације попут: метода, резултата, закључка научног рада, датум објављивања и научни часопис у коме је рад објављен, ауторе, референце и додатне документе.
- Аспект **„Архивирано“** су обједињени метаподаци који држе вредности о самом процесу архивирања. Овај аспект се додељује пројекту након архивирања, као и свим документима и директоријумима у електронској архиви. Реч је о: иницијатору архивирања, особи која је одобрила архивирање, датуму архивирања, линку до архивираног пројекта (на продукционој Алфреско инстанци) или линку до оригиналног пројекта (на архивској Алфреско инстанци), времену трајања самог архивирања (продукциони Алфреско) или датуме креирања и модификације или креатора и особе која је последња ажурирала пројекат (архивски Алфреско).

Овим је дефинисан модел података биоинформатичке групе. Могуће је уконфигурисати и приказ модела кориснику, али како кориснички интерфејс није приоритет овог рада, модел је довољан за рад биоинформатичке групе, тј. могу се креирати сви описани типови докумената и директоријума, чиме је први део проблематике решен. Други део представља сам процес слања.

Имплементација *CMIS* протокола је у претходним поглављима била детаљно описивана, као и њени недостаци, па ће се у имплементацији овог рада користити таква каква јесте, односно већ споменуте библиотеке **OpenCMIS**, али не најновије верзије (0.12.0), већ нека од старијих верзија која је компатибилна са изабраном, такође старијом верзијом управљачког система, односно верзија 0.2.0. Ово није ограничење, јер за потребе овог рада, наведена верзија садржи све предности и недостатке које има и најновија.

Другим речима за трансфер електронских докумената и директоријума, неопходно је још омогућити слање из продукционог Алфреско система, као и примање у архивском *DMS*-у (само због раније наведених недостатака протокола).

У првом случају, слања из продукционог окружења у електронску архиву, за демонстрацију Алфрескових могућности имплементирана су два начина слања: слање у електронску архиву преко Алфрескове акције и трансфер уз помоћ Алфреско радног тока. Алфреско акција се извршава из контекста жељеног пројекта, може је извршити било ко и архивирање се извршава само користећи имплементацију протокола, без старања о недостатку самог протокола. Јасно је да овај сценарио слања није пожељан у стварном окружењу, али у овом случају ће сликовито приказати све раније исказане тврдње. Алфреско радни ток за слање у архиву (видети слику 13) започиње на иницијативу било ког корисника који има и најмање право (читања) над самим пројектом и извршава се над самим пројектом. Замисао је да такав корисник може да предложи архивирање, а један од аутора пројекта би архивирање морао да

одобри, као и супервизор, односно, руководилац биоинформатичке групе. Одлука руководиоца је финална и након његовог/њеног одобравања - иницијализује се акција архивирања, овог пута акција која се стара да отклони недостатке протокола који се користи.

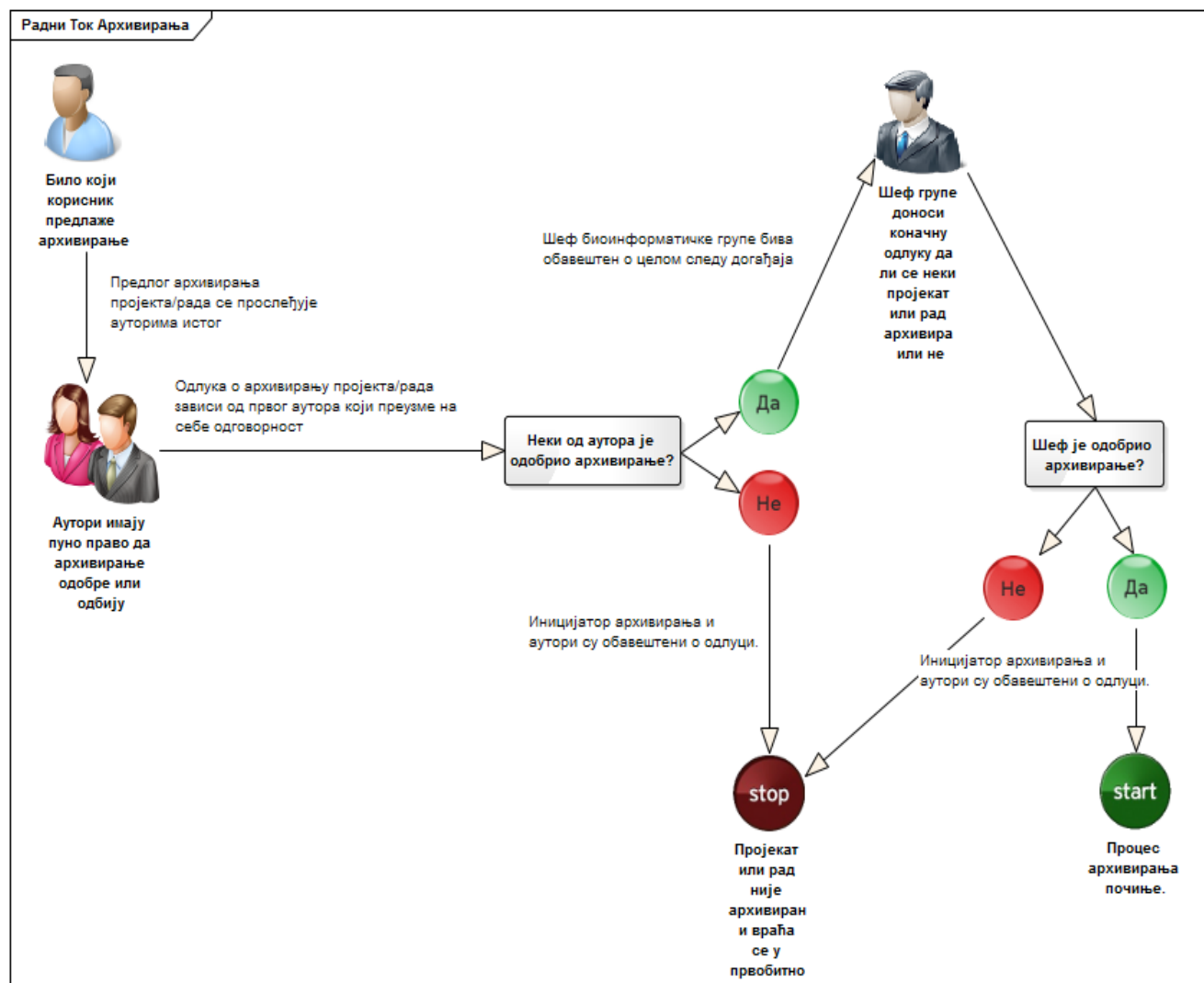
Док траје слање у електронску архиву, док траје радни ток одобравања или када је пројекат већ архивиран, немогуће је поново иницирати слање у архиву на било који начин. Уколико аутор или руководилац групе не одобре архивирање, пројекат се враћа у пређашње стање, тј. могуће га је поново архивирати. Чин архивирања блокира даље активности корисника у систему. У правом радном окружењу, архивирање би морало бити извршено у позадинском процесу, али имплементирано је на овај начин само због презентације рада и визуелног индикатора (учитавања) у прегледачу, да архивирање још увек траје.

Други случај, односно пријем пројекта у архивском репозиторијуму, пак, демонстрира Алфрескове могућности у виду понашања и правила (описани у одељку 4.2). У случају архивирања без очувања метаподатака (архивирање преко акције) ниједан од ова два механизма не функционише, јер за то и нема потребе. Механизми ослушкују цео архивски репозиторијум и када се деси извесни догађај, окине се акција у вези са њима. Све претходно, уз одабрано решење за очување метаподатака (креирање и слање привремених докумената) је неопходно како би се направила идентична копија пројекта у електронској архиви. У следећем одељку ће детаљније бити објашњен сам процес и алгоритам слања пројекта у електронску архиву, рад акције за слање на продукционом репозиторијуму, као и рад правила и понашања на архивском.

```
<type name="bio:ExternalProject">
  <title>Тудји научни пројекти</title>
  <parent>cm:content</parent>
  <mandatory-aspects>
    <aspect>bio:GeneralProjectAspect</aspect>
  </mandatory-aspects>
</type>
</types>

<aspects>
  <aspect name="bio:GeneralProjectAspect">
    <properties>
      <property name="bio:GeneralProjectResults">
        <title>Rezultati projekta</title>
        <type>d:text</type>
      </property>
      <property name="bio:GeneralProjectMethods">
        <title>Metode Projekta</title>
        <type>d:text</type>
        <multiple>true</multiple>
      </property>
      <property name="bio:GeneralProjectConclusion">
        <title>Zakljucak projekta</title>
        <type>d:text</type>
      </property>
      <property name="bio:GeneralProjectPublicationDate">
        <title>Datum objavljivanja projekta</title>
        <type>d:date</type>
      </property>
      <property name="bio:GeneralProjectSCIPaper">
        <title>Naucni casopis u kom je objavljen projekat</title>
        <type>d:text</type>
        <constraints>
          <constraint ref="bio:SCIList" />
        </constraints>
      </property>
      <property name="bio:GeneralProjectAuthors">
        <title>Autori Projekta</title>
        <type>d:text</type>
        <multiple>true</multiple>
      </property>
    </properties>
    <associations>
      <association name="bio:GeneralProjectReferences">
        <title>Reference projekta</title>
        <source>
          <many>true</many>
        </source>
        <target>
          <class>cm:content</class>
          <many>true</many>
        </target>
      </association>
      <association name="bio:GeneralProjectAdditionalDocuments">
        <title>Reference projekta</title>
        <source>
          <many>true</many>
        </source>
        <target>
          <class>cm:content</class>
          <many>true</many>
        </target>
      </association>
    </associations>
  </aspect>
```

Слика 12: Део модела (мета)података Алфреско DMS-а који дефинише тип документа „Спољни Пројекат“, као и (припадајући му) аспект „Аспект уопштеног пројекта“.



Слика 13: Алгоритам Алфреско радног тока за слање пројекта у електронску архиву.

5.2. Операција архивирања

Архива, сводила се она на електронску или папирну, свакако представља важну ставку (пре свега у организационом и смислу прегледности) било које институције, предузећа, канцеларије и сл. Док у неким мањим предузећима одлагање докумената у архиву и може да се сведе на одлуку једног човека, најчешћи случај, када је реч о тимском раду, је управо консултација сарадника да ли би неки документ или група докумената требало да се архивира или не.

Оба ова сценарија слања су имплементирани (Алфреско акција и радни ток) у продукционом Алфреско управљачком систему који прати овај рад. Разликује их одвијање два корака (четвртог и шестог, у наставку). Када се архивирање извршава преко радног тока, приликом копирања извесног објекта (типа директоријума или документа), креирају се помоћни документи који служе упису неповратно изгубљених информација које имплементација протокола не мапира, док се на архивској страни, ти документи

читају, парсирају и обрађују. У случају архивирања преко Алфреско акције, не генеришу се привремени документи, самим тим ни у архиви нема њихове обраде. Другим речима, за оба начина архивирања сви кораци су идентични осим наведених. Архивирање је успешно ако и само ако су сви кораци завршени успешно. Ево и детаља корака слања која се одвијају унутар *Java* зрна, на продукционој Алфреско инстанци:

- **Први корак:** креирање *CMIS* сесија (енгл. *Session*) за рад са овом Алфреско инстанцом, али и за рад са физички удаљеном, архивском инстанцом. Сесија, уколико се успостави са корисником који има администраторска права, омогућује пуну контролу над креирањем, ажурирањем и брисањем садржаја било ког репозиторијума, што је веома моћна, али и опасна чињеница, те се зато све активности морају одвијати са великом дозом опрезности.
- **Други корак:** валидација свих неопходних предуслова за слање. На пример: да ли је пројекат којим случајем обрисан у међувремену, док се чека на потврду аутора или руководиоца групе, да ли постоје документи у пројекту који су системски закључани (самим тим се не могу копирати ни брисати), да ли корисници система укључени у архивирање још увек постоје итд.
- **Трећи корак:** рекурзивни пролазак кроз пројекат и прикупљање неопходних информација (за сам трансфер) о свим поддиректоријумима и документима. Реч је о попуњавању *Java* објеката у виду мапа и листа које се тичу списка свих објеката које би требало послати, њихових адреса на локалном систему и адреса где би их требало прекопирати, корисничких имена креатора и особе која је последња ажурирала објекат, датума креирања и последњег ажурирања итд.
- **Четврти корак:** креирање неопходне хијерархије директоријума у архиви у коју се пројекат смешта (уколико таква не постоји) и копирање „љуштуре“ пројекта, тј. самог директоријума пројекта користећи *CMIS* библиотеке и, евентуално, креирање привременог документа (уколико је то неопходно) у истом директоријуму у који се смешта копија пројекта, а који служи механизму за очување метаподатака. Овај привремени документ се креира само у случају архивирања преко Алфреско радног тока.
- **Пети корак:** рекурзивно копирање комплетног садржаја пројекта, користећи *CMIS* библиотеке. Аналогно као у претходном кораку, генерише се и привремени документ у архиви (у случају слања пројекта у архиву преко Алфреско радног тока и уколико је то неопходно. Наиме, могуће је да документ или директоријум који се шаље има комплетан скуп метаподатака који имплементација протокола уме да мапира, а самим тим и да их поново придружи новокреираном објекту у архиви. Овим се смисао привремених докумената губи, па се у

том случају они и не креирају и не шаљу, што не мора бити редак случај) у вези са ископираним објектом.

- **Шести корак:** додељивање аспекта „Архивирано“ директоријуму пројекта у продукционој инстанци Алфреска и рекурзивно додељивање истог Алфреско аспекта (уз помоћ *CMIS* библиотека) новокреираном пројекту и објектима које он садржи у електронској архиви. Вредности метаподатака овог аспекта су у уској релацији са самим пројектом, што је већ описано у претходном поглављу. *Овај корак је окидач Алфреско понашања архивског управљачког система*, али о томе ће бити више речи на странама које следе.
- **Седми корак:** брисање садржаја оригиналног пројекта. Уколико се дошло до овог корака, то значи да је садржај у целости безбедно и успешно пребачен у архиву и простор на продукцији може бити ослобођен.
- **Осми корак:** ажурирање метаподатака аспекта „Архивирано“ оригиналног пројекта који се тичу линка до архивираног пројекта и времена архивирања. Ово је уједно и последњи корак, након чега се ослобађају ресурси и систем је спреман за следећу операцију архивирања.

Уколико се неки корак не изврши успешно, процес архивирања се прекида, корисник бива обавештен о грешци, а пројекат се враћа у првобитно стање. Исто тако, пројекат који се креирао у архиви, о ком год неуспешном кораку да се ради, се брише. Другим речима, може се рећи да се архивирање обавља у трансакцији. Овај процес је такође неповратан, тј. није замишљено, ни имплементирано враћање пројекта из архиве у продукцију.

Важно је напоменути да би привремени документи требало да се зову слично као и документ или директоријум са којим су у вези. Но, исто тако приликом креирања привремених докумената теоретски се може десити да у неком директоријуму већ постоји документ таквог имена. Иако (поново) у теорији, због овога никада не можемо потпуно бити безбедни, неопходно је смислити конвенцију давања имена привременим документима која ће у скоро сто процената случајева функционисати без грешке. Најчешћа програмерска пракса је додавање „**тајмстемп**“ (енгл. *timestamp*) суфикса који представља тренутни датум и време форматиран на жељени начин. Ипак, с обзиром да ни у том случају, теоретски, нисмо увек безбедни, а свакако је лакше да тај додатак буде нека константа, у овој имплементацији се за привремене документе користи префикс и то „**^^^BioTempAssoc^^^**“.

Опишимо сада процес пријема пројекта у архиви, који је неопходан због несавршености протокола. Ти механизми раде само у случају ако постоје привремени документи.

Алфреско инстанца која представља електронску архиву у себи има имплементирано Алфреско правило и понашање који чине споменути механизам. Правило подразумева извршавање неке системски дефинисане акције након неког догађаја у репозиторијуму. У овом случају, наведена акција је извршавање

скрипте (писане у програмском језику *Javascript*), а догађај који се ослушкује (над директоријумом где се пројекти смештају) је креирање документа чије им почиње са „**^^^BioTemp**“, тј. ослушкује се креирање *привремених* докумената. Скрипта чита привремени документ, из њега извлачи метаподатке (ако постоје) о кодној шеми (енгл. *encoding*) и *MIME* типу документа и поставља их на документ у истом директоријуму и истог имена, само без наведеног префикса.

Алфреско понашање, након неког догађаја у репозиторијуму, позива извршавање изабраног *Java* зрна. У случају пријема архивираног пројекта, споменути догађај је шести корак слања, односно додавање аспекта „*Архивирано*“ неком документу или директоријуму, а акција ради следеће:

- Проналази се, уколико постоји, привремени документ који је у релацији са објектом коме је аспект додељен (у истом је директоријуму, истог имена са константим префиксом). Уколико такав привремени документ не постоји, то значи да је протокол мапирао и ископирао све метаподатке и стога се овде рад механизма завршава.
- Привремени документ се чита, складиштени подаци се парсирају и информације и асоцијације се поново успостављају. Овде је битно напоменути да у уопштеним случајевима асоцијацију није могуће поново успоставити и то нпр. ако „мета“ (*target*) асоцијације није учествовала у трансферу. У овом конкретном случају, то не може да се деси, јер сви корисници који постоје у продукционом управљачком систему, постоје и у архиви, само ради демонстрације, па асоцијације у којима су они „мета“ (рецимо у аспекту „**Аспект пројекта**“, асоцијације контакт особе и преосталих аутора пројекта) се успешно рекреирају, а имплементирано је и архивирање, односно копирање осталих Алфреско чворова (докумената) који су у асоцијацији са неким објектом у пројекту или самим пројектом, а не припадају директоријуму пројекта. Јединственост имена новокреираних докумената се овог пута постиже уобичајеном праксом, односно, тајмстемпом.
- Уколико је операција читања и очувавања свих метаподатака завршена успешно, привремени документ се брише. Уколико, ипак из неког разлога, није могуће поново успоставити вредности метаподатака изгубљених у трансферу или ако је дошло до неке грешке у раду механизма, привремени документ се не брише. Он преостаје да служи као носач изгубљених информација које администратор може да покуша ручно да унесе.

Када механизам обради и последњи објекат унутар архивираног пројекта којем је аспект „*Архивирано*“ додељен (чиме се механизам и окида), процес пријема пројекта у архиви је завршен.

Лако је уочити да је архивирање, када се посматра из контекста продукционог окружења, активан процес, који у потпуности и једини имплементира протокол и држи пуну контролу и валидацију трансфера. С друге стране, процес пријема пројекта у

електронску архиву из контекста друге Алфреско инстанце је прилично статичан и пасиван процес. Нема никаквих акција и контроле, нема коришћења протокола, једино постоји ослушкивање уобичајених догађаја у репозиторијуму, који се (у другим пословним логикама) могу десити и без архивирања и било каквог трансфера података и протокола.

Преостало је још споменути почетна стања оба Алфреско репозиторијума пре првог архивирања. Електронска архива је иницијално празан репозиторијум. *CSS (Cascade Style Sheets)* позадине је промењен у бресква-наранџасту нијансу, како би се архива и визуелно разликовала од светлих, бело-сивих нијанси оригиналног Алфреско окружења представљеног у другој инстанци.

Продукционо окружење садржи два директоријума која представљају матичне директоријуме биоинформатичких група. Свака од њих садржи по десет пројеката, али само два пројекта у подгрупи „*Геномика*“ (са редним бројем 1 и 2) представљају симулације пројеката које ће бити коришћене у тестирању сценарија архивирања са и без очувања метаподатака. Оба пројекта су заправо део симулације израде једног правог пројекта, научног рада: „*Bioinformatics analysis of SARS coronavirus genome polymorphism*“, аутора Гордане Павловић-Лажетић, Ненада Митића и Милоша Бељанског, објављеног у часопису „*BMC Bioinformatics*“, 25. маја 2004. године [25]. Метаподаци пројеката су одабране целине рада, додате су асоцијације на документе у прилогу и ауторе, а додато је и још неколико имагинарних детаља.

Оба Алфреско репозиторијума су подигнута на истом серверу, али на различитим портловима и протокол приступа архивском из локала, чиме је брзина трансфера, самим тим и архивирања, максимална. Овим је омогућена неометана проба описаног сценарија. У следећа два поглавља биће представљена и посматрана тестирања имплементације протокола у виду оба начина архивирања са својим предностима и недостацима.

5.3. Тестирање сценарија без примене имплементације решења и резултати

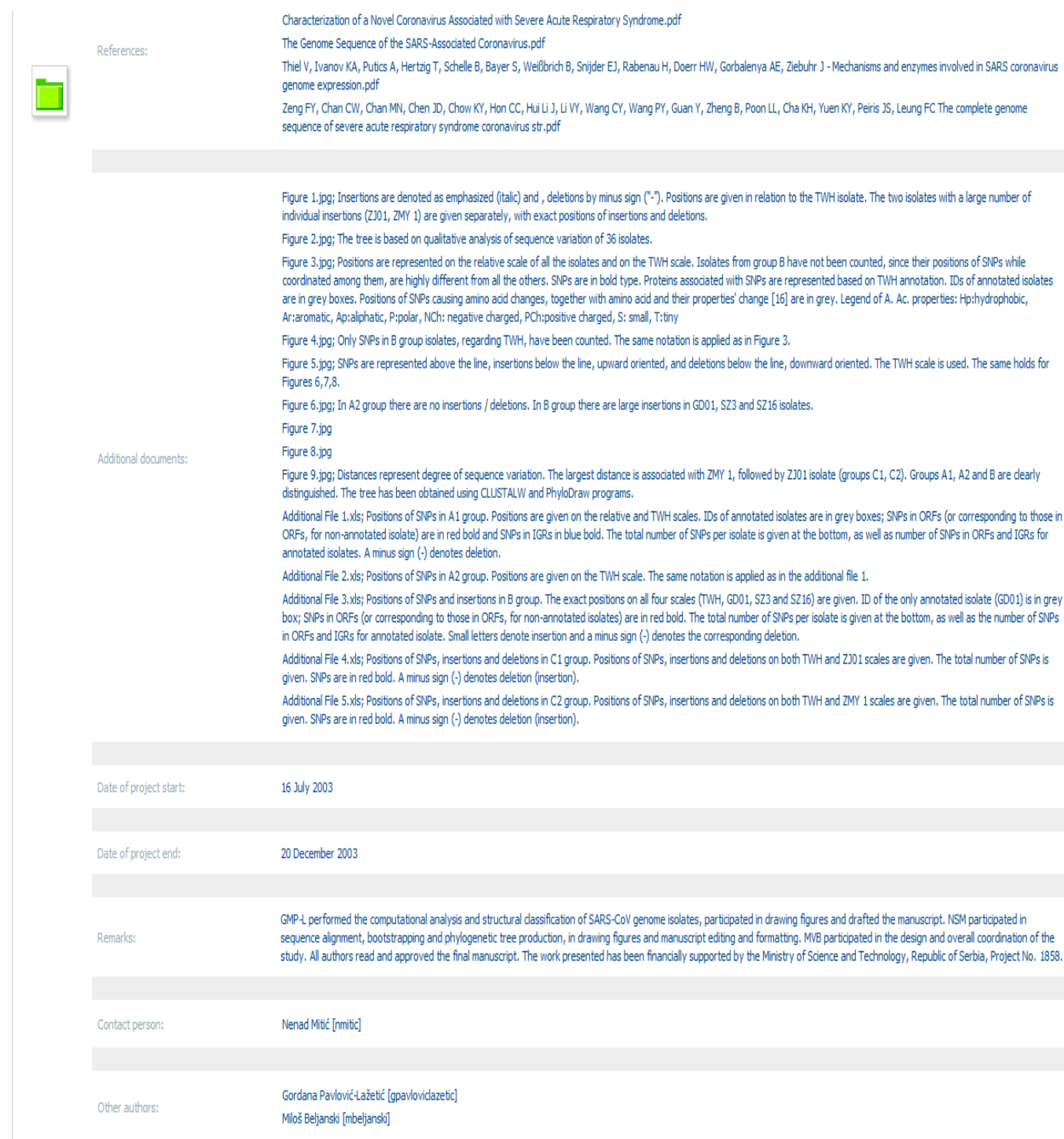
Последња фаза анализе *CMIS* протокола представља демонстрацију и пун приказ моћи и немоћи у пракси, ове будућности развоја програмирања у облаку. Сценарио који следи представља архивирање пројекта биоинформатичке групе користећи изворну имплементацију протокола. Овим ће постати уочљиви недостаци имплементације у виду Алфреско асоцијација које се нису поново успоставиле у архиви, иако оба Алфреско чвора (документ или корисник) постоје у електронској архиви.

Пре изношења резултата вишеструких архивирања, прво статистички подаци: у сценарију се архивирају *22 документа* и *3 директоријума* (укључујући и директоријум самог пројекта) укупне величине нешто мање од **4 мегабајта** (око 3906 килобајта). У

стварности, пројекат би заузимао барем три-четири пута више, али за потребе свих аспекта овог рада, ово је оптимална величина.

На слици 14 се види део метаподатака оригиналног пројекта пре архивирања. Видљиве су релације са бројним документима, као и са корисницима. Из контекста Алфреска, то су Алфреско асоцијације, а из контекста *CMIS* стандарда, то су *CMIS* везе. Након архивирања без примене решења за очување метаподатака, ови метаподаци се неповратно губе, што се лако може уочити упоређујући ову слику са следећом сликом (15). Разлог зашто се ови подаци губе лежи у одсуству мапирања Алфреско асоцијација које су унутар неког Алфреско аспекта. Да би ова тврдња била доказана, коришћен је програм **Cygwin** са инсталираним алатом **cURL**.

Cygwin [26] је конзолна апликација за *Windows* платформе која омогућава интеграцију апликација те платформе са *Unix* окружењем. У овом случају, желимо да добијемо *Atom XML* захтева са управљачког система Алфреско, подигнутог на *Windows* платформи, уз помоћ **cURL** алата. **cURL** [27] је бесплатан алат отвореног кода за трансфер података уз помоћ *URL* синтаксе. Другим речима, датим линком добија се жељени документ. *Atom XML* било ког захтева није доступан директно преко



Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome.pdf
The Genome Sequence of the SARS-Associated Coronavirus.pdf
Thiel V, Ivanov KA, Putics A, Hertzog T, Schelle B, Bayer S, Weißbrich B, Snijder EJ, Rabenau H, Doerr HW, Gorbalenya AE, Ziebuhr J - Mechanisms and enzymes involved in SARS coronavirus genome expression.pdf
Zeng FY, Chan CW, Chan MN, Chen JD, Chow KY, Hon CC, Hui Li J, Li VY, Wang CY, Wang PY, Guan Y, Zheng B, Poon LL, Cha KH, Yuen KY, Peiris JS, Leung FC. The complete genome sequence of severe acute respiratory syndrome coronavirus str.pdf

References:

Figure 1.jpg; Insertions are denoted as emphasized (italic) and deletions by minus sign ("−"). Positions are given in relation to the TWH isolate. The two isolates with a large number of individual insertions (ZJ01, ZMY 1) are given separately, with exact positions of insertions and deletions.
Figure 2.jpg; The tree is based on qualitative analysis of sequence variation of 36 isolates.
Figure 3.jpg; Positions are represented on the relative scale of all the isolates and on the TWH scale. Isolates from group B have not been counted, since their positions of SNPs while coordinated among them, are highly different from all the others. SNPs are in bold type. Proteins associated with SNPs are represented based on TWH annotation. IDs of annotated isolates are in grey boxes. Positions of SNPs causing amino acid changes, together with amino acid and their properties' change [16] are in grey. Legend of A. Ac. properties: Hp:hydrophobic, Ar:aromatic, Ap:aliphatic, P:polar, NCh: negative charged, PCh:positive charged, S: small, T:tiny
Figure 4.jpg; Only SNPs in B group isolates, regarding TWH, have been counted. The same notation is applied as in Figure 3.
Figure 5.jpg; SNPs are represented above the line, insertions below the line, upward oriented, and deletions below the line, downward oriented. The TWH scale is used. The same holds for Figures 6,7,8.
Figure 6.jpg; In A2 group there are no insertions / deletions. In B group there are large insertions in GD01, SZ3 and SZ16 isolates.
Figure 7.jpg
Figure 8.jpg
Additional documents:
Figure 9.jpg; Distances represent degree of sequence variation. The largest distance is associated with ZMY 1, followed by ZJ01 isolate (groups C1, C2). Groups A1, A2 and B are clearly distinguished. The tree has been obtained using CLUSTALW and PhyloDraw programs.
Additional File 1.xls; Positions of SNPs in A1 group. Positions are given on the relative and TWH scales. IDs of annotated isolates are in grey boxes; SNPs in ORFs (or corresponding to those in ORFs, for non-annotated isolate) are in red bold and SNPs in IGRs in blue bold. The total number of SNPs per isolate is given at the bottom, as well as number of SNPs in ORFs and IGRs for annotated isolates. A minus sign (−) denotes deletion.
Additional File 2.xls; Positions of SNPs in A2 group. Positions are given on the TWH scale. The same notation is applied as in the additional file 1.
Additional File 3.xls; Positions of SNPs and insertions in B group. The exact positions on all four scales (TWH, GD01, SZ3 and SZ16) are given. ID of the only annotated isolate (GD01) is in grey box; SNPs in ORFs (or corresponding to those in ORFs, for non-annotated isolates) are in red bold. The total number of SNPs per isolate is given at the bottom, as well as the number of SNPs in ORFs and IGRs for annotated isolate. Small letters denote insertion and a minus sign (−) denotes the corresponding deletion.
Additional File 4.xls; Positions of SNPs, insertions and deletions in C1 group. Positions of SNPs, insertions and deletions on both TWH and ZJ01 scales are given. The total number of SNPs is given. SNPs are in red bold. A minus sign (−) denotes deletion (insertion).
Additional File 5.xls; Positions of SNPs, insertions and deletions in C2 group. Positions of SNPs, insertions and deletions on both TWH and ZMY 1 scales are given. The total number of SNPs is given. SNPs are in red bold. A minus sign (−) denotes deletion (insertion).

Date of project start: 16 July 2003

Date of project end: 20 December 2003

Remarks: GMP-L performed the computational analysis and structural classification of SARS-CoV genome isolates, participated in drawing figures and drafted the manuscript. NSM participated in sequence alignment, bootstrapping and phylogenetic tree production, in drawing figures and manuscript editing and formatting. MVB participated in the design and overall coordination of the study. All authors read and approved the final manuscript. The work presented has been financially supported by the Ministry of Science and Technology, Republic of Serbia, Project No. 1858.

Contact person: Nenad Mitić [nmitic]

Other authors: Gordana Pavlović-Lažetić [gpavlovidlazetic]
Miloš Beljanski [mbeljanski]

Слика 14: Поглед на део метаподатака оригиналног пројекта пре архивирања.

линка из Алфреска, стога се тај XML документ мора затражити од Алфреска исто онако како га CMIS протокол добија. Тиме ћемо уједно добити и начин на који имплементација стандарда „види“ садржај у Алфреско репозиторијуму.

Дакле, неопходно је аутентификовати се уз помоћ **Cygwin**-а на Алфреско управљачки систем и задати прецизну **cURL** команду како би се добио жељени одговор у XML формату. Након успешног приступа Алфреску, извршавамо команду:

curl USERNAME:PASSWORD

"http://HOST:PORT/alfresco/service/api/node/workspace/spacesStore/NODEREF/rels?includeSubRelationshipTypes=true"

где су **USERNAME** и **PASSWORD** - корисничко име и лозинка корисника којима смо остварили приступ Алфреско *DMS*-у, **HOST** и **PORT** - сервер и порт на коме је Алфреско подигнут и **NODEREF** представља јединствени идентификатор Алфреско чвора чије метаподатке желимо да испитамо. Овом командом се аутентификујемо на Алфреско и тражимо све могуће *CMIS* везе датог Алфреско чвора.

Слика 16 приказује одговор добијен после извршења ове команде за Алфреско чвор, тј. пројекат са слике 14. У тагу `<opensearch:totalResults>0</opensearch:totalResults>` се јасно види број *CMIS* веза, које имплементација протокола мапира. На слици 14 видимо двадесетак релација пројекта са његовим документима и корисницима система. Од двадесетак релација, не мапира се ниједна!

Другим речима, од приказаних седам, свега три метаподатка су сачувана и то она, која нису Алфреско асоцијација. На глобалном нивоу пак, ситуација је доста боља; шест докумената (од 22) и преостала два директоријума нису изгубила ниједан метаподатак, док су преостали објекти, сви скупа изгубили око 10% првобитних метаподатака.

Након двадесет архивирања истог пројекта без очувања метаподатака, просечно трајање архивирања је око два минута и четрнаест секунди. Неки други алгоритам архивирања могао би да буде примеренији и додатно оптимизован, али свакако не би био драматично бољи и бржи од имплементације. Ипак, са свим додатним операцијама, трансфер од 4 МВ на локалном рачунару траје преко два минута. Узмимо у обзир да су обе инстанце Алфреско подигнуте на истом рачунару и да протокол приступа локално архивском репозиторијуму, а не са неке удаљене локације, што би додатно успорило трансфер.

Уочљиво је да због својих предности да паметно мапира, сачува, обави трансфер и поврати оригиналне податке, скуп информација које *CMIS* протокол шаље је велики, а самим тим је и процес слања - спор. Може се претпоставити да ће даљим развојем протокола и његових имплементација, трансфер бити још оптимизованији и ефикаснији.

Резултати овог тестирања не иду у прилог овом протоколу. У овом сценарију тестирања примећене су следеће предности:

- Иако су неки објекти изгубили метаподатке, сви директоријуми и документи се у целости архивирају коришћењем чисте имплементације протокола. Садржај докумената се не губи.
- Коришћење оригиналних библиотека имплементације *CMIS*-а за трансфер је релативно једноставно.
- Нема осмишљавања нових алгоритама или корака за сам трансфер.

- Алфреско инстанца електронске архиве је растерећена, јер се механизми пријема никада не активирају у овом случају.

References:

Additional documents:

Date of project start: 16 July 2003

Date of project end: 20 December 2003

Remarks:

GMP-L performed the computational analysis and structural classification of SARS-CoV genome isolates, participated in drawing figures and drafted the manuscript. NSM participated in sequence alignment, bootstrapping and phylogenetic tree production, in drawing figures and manuscript editing and formatting. MVB participated in the design and overall coordination of the study. All authors read and approved the final manuscript. The work presented has been financially supported by the Ministry of Science and Technology, Republic of Serbia, Project No. 1858.

Contact person:

Other authors:

Слика 15: Поглед на део метаподатака пројекта архивираног без решења за очување садржаја.

```
<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:app="http://www.w3.org/2007/app" xmlns:cmisra="http://docs.oasis-open.org/ns/cmris/restatom/200908/"
xmlns:cmis="http://docs.oasis-open.org/ns/cmris/core/200908/" xmlns:alf="http://www.alfresco.org" xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/">
<author><name>genomikaVodja</name></author>
<generator version="3.3.0 (g 2860)">Alfresco (Community)</generator>
<icon>http://localhost:8080/alfresco/images/logo/AlfrescoLogo16.ico</icon>
<id>urn:uuid:480b7725-1ac8-4569-829a-33fec3f195e0-relationships</id>
<link rel="service" href="http://localhost:8080/alfresco/service/cmris"/>
<link rel="self" href="http://localhost:8080/alfresco/service/api/node/workspace/spacesStore/480b7725-1ac8-4569-829a-33fec3f195e0/rels?includeSubRelationshipTypes=true"/>
<link rel="via" href="http://localhost:8080/alfresco/service/cmris/s/workspace:spacesStore/i/480b7725-1ac8-4569-829a-33fec3f195e0"/>
<title>Projekat GE-000001 Relationships</title>
<updated>2014-09-06T20:44:33.039+02:00</updated>
<opensearch:totalResults>0</opensearch:totalResults>
<opensearch:startIndex>0</opensearch:startIndex>
<opensearch:itemsPerPage>-1</opensearch:itemsPerPage>
<cmisra:numItems>0</cmisra:numItems>
</feed>
```

Слика 16: AtomPub XML одговор на CMIS захтев за бројем CMIS веза једног Алфреско чвора (пројекат са слике 14).

Ипак, недостаци су крупнији, а огледају се у:

- Непотпуној имплементацији протокола. То се лако уочава при слању у архиву само једног документа, а огледа се у немогућности мапирања Алфреско асоцијација у аспектима, самим тим и њиховом прослеђивању у податке за трансфер.

- Последици претходног недостатка - неповратном губитку око 10% оригиналних метаподатака. Подаци који се губе у сценарију су Алфреско асоцијације, тип кодирања и тип документа. Тип документа и понеке Алфреско асоцијације би можда и могле да се у неким случајевима у пракси интуитивно, ручно повежу, али то није ни смисао, а ни сврха протокола.
- Спором трансферу података, као последици слања комплексних објеката.
- Након трансфера, операција рекурзивног додавања аспекта садржају архивираног пројекта уз помоћ имплементације протокола траје око двадесет секунди, тј. скоро објекат по секунди, што је веома споро.

У следећем поглављу видећемо други сценарио тестирања који (иако није отклонио све недостатке протокола) је донео знатна побољшања у раду, уз додатну цену за то.

5.4. Тестирање сценарија са применом имплементираних решења и резултати

Сценарио слања пројекта биоинформатичке групе у електронску архиву са решењем за очување метаподатака само се незнатно разликује од претходног сценарија. Једини кораци разлике су креирање привремених докумената који носе неповратно изгубљене метаподатке и њихово копирање које окида механизам, део корака за успостављање метаподатака у архивском репозиторијуму.

Као и у претходном тестирању и овде се архивира идентичан број директоријума (3) и докумената (22), али мора се додати то да се број помоћних докумената који се креира мери у десетинама, односно, једнак је броју докумената који су изгубили неки метаподатак у трансферу. Иако *привремени* документи заузимају свега неколико килобајта меморије, ипак је потребно кратко, али извесно време за њихово креирање у архиви уз помоћ *CMIS*-а, па је већ у овом тренутку јасно да ће спорост трансфера у овом примеру бити још израженија.

У прилог овоме иде и статистика; двадесет архивирања овим сценаријем је резултовало просечним временом архивирања од око два минута и четрдесет секунди. Ово је нешто мање од пола минута дуже од претходног сценарија, а то је заправо нешто мање од петине укупног времена. С обзиром да време креирања докумената, постављање њихових садржаја, постављање аспеката преко *CMIS*-а и друге акције протокола функционишу истом брзином и када су продукција и архива на једном рачунару или на другим крајевима света (једино што зависи од брзине протока на мрежи је слање самог *Atom XML*-а), сценарио са имплементираним решењем за очување метаподатака је *за око 20% спорији од сценарија без тог решења*.

Насупрот прошлом сценарију, овај сценарио имплементира решење за очување метаподатака, чиме се сав садржај пројекта и сви метаподаци успешно архивирају и нема никаквог губитка

информација, што је од суштинског значаја за ову биоинформатичку или било коју другу истраживачку, научну или радну групу која користи неки управљачки систем.

Summa summarum, предности овог сценарија су:

- Сви директоријуми и документи су у целости премештени, без икаквих ограничења или оштећења.
- Коришћење оригиналних библиотека имплементације *CMIS*-а за трансфер је релативно једноставно.
- Сви метаподаци свих Алфреско, односно *CMIS* објеката су успешно сачувани на архивском репозиторијуму.
- Представља пуноправан, сликовит и реалан приступ проблематици у пракси.

С друге стране, недостаци се огледају у:

- Додатном алгоритму за генерисање привремених докумената који се шаљу у архивски репозиторијум.
- Спором трансферу података, као последици слања комплексних објеката и привремених докумената.
- Након трансфера, операција рекурзивног додавања аспекта садржају архивираног пројекта уз помоћ имплементације протокола траје око двадесет секунди, тј. скоро објекат по секунди, што је веома споро.
- Неопходности постојања механизма у архивском репозиторијуму за потребе процесуирања привремених докумената и поновно успостављање метаподатака које имплементација *CMIS* стандарда није успела да одради.

Упоредјујући два сценарија тестирања, може се закључити да уколико фактор брзине операције архивирања није пресудан, сценарио архивирања са решењем за очување метаподатака је значајно бољи, ближи пословној логици једне праве научне групе, па је због тога и логичан избор. У пракси, брзина процеса је један од најнижих приоритета самог концепта смештања у архиву неких докумената. Архивирање може бити позадински процес који не омета рад ниједног корисника управљачког система. Такође, да би документи уопште били кандидати за архивирање, мора проћи извесно време у току ког они нису били од велике користи или пак ни од каквог значаја, па се претпоставља да се може потрошити и много више времена на архивирање него што је овим сценаријима неопходно, а посебно је то битно овом последњем, временски захтевнијем сценарију. Другим речима, цена која се плаћа (20% спорије архивирање) је свакако, у већини практичних случајева, вредна коначног резултата - постојању идентичности између првобитног и архивираног материјала.

6. ЗАКЉУЧАК

У двадесет и првом веку је настављен, обogaћен и убрзан непрестани научно-технолошки развој. Упоредо с тим, рачунари су све бољи, бржи, мањи, приступачнији. Све је мањи број папирних докумената, а насупрот томе, све је више електронских. Као последица свега наведеног, електронске обраде захтева су сваког дана све учесталија пракса. На пример, некада је за подношење разних пријава било неопходно приложити неколико докумената и чекати у реду сатима, а данас је то могуће обавити за неколико минута, не излазећи из своје куће. Такви документи се чувају у облаку, на неком се управљачком систему или веб порталу обрађују, док се размена врши уз помоћ неког протокола за трансфер података.

У сваком предузећу, фирми, компанији, установи, институту, факултету или писарници, рад са документима, папирним или електронским, је неминовност. Временом, број ових докумената се не смањује, напротив, непрестано расте, чиме се ствара потреба да се сувишни, застарели или документи који су ретко у употреби, негде преместе ради бољег квалитета рада. Ни у истраживачкој групи ово правило није изузетак.

Управљачки систем Алфреско, осим што је једно од најбољих бесплатних окружења отвореног кода за тимски рад, игра одличну улогу и као електронско складиште докумената, односно, у овом конкретном случају, електронску архиву. Оба контекста - инстанце апликације су обухваћена у овом раду и представљене су неке од бројних могућности наведеног *DMS*-а.

Ипак, архивирање не би било могуће без самог трансфера докумената. *CMIS* протокол (односно, прецизније речено, његова најбоља имплементација), једна од најмлађих технологија у успону, искоришћен је за овај трансфер. Иако се најбољи пример искористивости овог протокола огледа у примени програмирања у облаку, он ефикасно одговара и на друге захтеве. Но, како постоји приметни раскорак између дефиниције и имплементације протокола, проћи ће извесно време док *CMIS* не оствари свој зенит.

Највећи недостаци имплементације стандарда су уочени већ при првим корелацијама са две Алфреско инстанце. У изложеном сценарију примене, *CMIS* служи као једносмерна спрега између једног Алфреско система, који представља радно окружење извесне биоинформатичке групе која се строго бави израдом научних радова и пројеката, и другог Алфреско система који представља архиву радова ове групе. Иако је ефикасно одговорио на већину изазова, ипак је показао да га у тој ефикасности не краси и брзина, а у неким случајевима је и потпуно подбацио, као рецимо у мапирању релација директоријума и докумената, а самим тим и у очувању свих оригиналних метаподатака. Но, уз малу дораду, решења за очување метаподатака постоје, а са њима се драстично поправља пољуљани утисак о примени овог протокола.

Једно од решења је и примењено у овом раду. Реч је о коришћењу привремених докумената који чувају информације неповратно изгубљене у трансферу. Ови документи бивају

процесирани од стране механизма архивирања у електронској архиви, након чега се успоставља првобитно стање (у овом сценарију, у потпуности, а иначе, у мери колико је то могуће).

Након обимнијег тестирања показало се да су *CMIS* споре, да примењено решење за очување метаподатака заиста функционише, али додатно успорава архивирање за око 20%. Утврђено је и да имплементација протокола може да се носи са највећим изазовима у интеракцији између управљачких система, али да је далеко од идеалне и да ју је неопходно даље развијати ради бољих резултата.

Замисао рада биоинформатичке групе (која се огледа у додатим опцијама активности и садржаја Алфреско *DMS*-а) је можда преуско сагледана, тј. просечна група се бави и многим другим активностима осим израде научних радова и пројеката, али довољна је да се уочи да је Алфреско и те како примењив у научно-истраживачке сврхе, да је осмишљен да олакша тимски рад и да га је могуће мењати и прилагођавати потребама сопствене пословне логике.

Чињеница је да све више компанија, фирми и предузећа, али и научно-истраживачких установа, института и факултета прелази са папирног на претходно описани, електронски приступ. До пре, рецимо, десет година, ово можда не би толико важило за државу Србију. Међутим, примећује се да и ми у све већој мери прихватамо најновије технологије. Иако мало и у сталној кризи, наше тржиште је препознало ову потребу; технологије су приступачне и свакако доносе велику уштеду у смислу времена и новца.

У овом раду смо видели примену *CMIS* стандарда у некомерцијалне, научне сврхе једне биоинформатичке групе. Док управљачки систем Алфреско (који групи пружа систем за тимски рад богат бројним функционалностима) служи за складиште електронских докумената, стандард пружа ефикасан трансфер истих било где - у овом случају у електронску архиву групе. Користећи бесплатни управљачки систем и *CMIS* протокол, биоинформатичка (или било која друга научна) група је ефикаснија, сам рад је богатији детаљима, лакши, бржи и прегледнији, а потреба за силним ормарима и простором за документе, били они нови, стари или архивски, је сведена на минимум. Пре или касније, свака продукциона, радна заједница ће имати исти или сличан систем, што овим новим технологијама пружа већу примену, а самим тим и светлу будућност.

7. ЛИТЕРАТУРА:

- [1] „Закон о електронском документу“, Службени гласник Републике Србије, бр. 51/2009
- [2] Званична страница фирме „Questy Solutions“, http://www.questyssolutions.com/pdfs/docs/Document_Imaging_Whitepaper.pdf, 5.6.2013.
- [3] Дефиниција Управљачког система докумената на енглеској википедији, http://en.wikipedia.org/wiki/Document_management_system, 6.6.2013.
- [4] Дефиниција Предузетничког управљачког система докумената на званичном сајту асоцијације AIIM (*Association for Information and Image Management*), <http://www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx>, 9.6.2013.
- [5] Tony Byrne, „*Alfresco tries to repeat history*“, <http://www.kmworld.com/Articles/News/News-Analysis/Alfresco-tries-to-repeat-history-14544.aspx>, 9.6.2013.
- [6] Дефиниција Алфреско софтвера на енглеској википедији, [http://en.wikipedia.org/wiki/Alfresco_\(software\)](http://en.wikipedia.org/wiki/Alfresco_(software)), 10.6.2013.
- [7] Munwar Shariff, Amita Bhandari, Vinita Choudhary, Pallika Majumdar, „*Alfresco Enterprise Content Management Implementation*“, Packt publishing, Бирмингем, Енглеска, 2008.
- [8] Alfresco White Papers: „*Total Cost of Ownership for Enterprise Content Management*“, Alfresco Software, Беркшир, Енглеска, 2008.
- [9] Sebastian Selhorst, „*The Total Economic Impact Of Alfresco Enterprise Content Management Solution*“, Forrester Research, Кембриџ (Масачусетс), Сједињене Америчке Државе, 2011.
- [10] Препоручени хардверски ресурси за Алфресков сервер на енглеској Алфреско википедији, http://wiki.alfresco.com/wiki/JVM_Tuning, 23.6.2013.
- [11] Matt Krautstrunk. „*Comparing the Numbers: WordPress, Joomla, Drupal*“, <http://www.resourcenation.com/blog/comparing-the-numbers-wordpress-joomla-drupal/31988/>, 26.7.2013.
- [12] Званична страница резултата веб истраживања фирме Q-Success, http://w3techs.com/technologies/overview/content_management/all, 26.7.2013.

- [13] Званична страница OASIS конзорцијума, <https://www.oasis-open.org/news/announcements/content-management-interopability-services-cmis-version-1-1-approved-and-publis>, 8.8.2013.
- [14] Florian Müller, Jay Brown, Jeff Potts, „*CMIS and Apache Chemistry in Action*“, Manning publications Co, Њујорк (Њујорк), Сједињене Америчке Државе, 2013.
- [15] Jeff Potts, „*Getting started with CMIS*“, Optaros, Бостон (Масачусетс), Сједињене Америчке Државе, 2009.
- [16] Страница огранка фирме „*The Apache Software Foundation*“, „Apache Chemistry“, <http://chemistry.apache.org>, 8.8.2013.
- [17] Oasis Cover Pages: „*Vendors Publish Content Management Interoperability Services (CMIS) Standard.*“, <http://xml.coverpages.org/ni2008-09-10-a.html>, 10.8.2013.
- [18] Tim Weber, „*Cloud computing could give EU 763bn-euro boost*“, <http://www.bbc.co.uk/news/business-11931841>, 14.8.2013.
- [19] Званична страница резултата истраживања фирме AIIM „*State of the ECM Industry 2011*“, http://www.aiim.org/pdfdocuments/IW_ECM_State-of-Industry_2011.pdf, 18.8.2013.
- [20] Stefan Waldhauser, „*5 myths about the CMIS standard*“, <http://www.digitallandfill.org/2011/08/5-myths-about-the-cmis-standard.html>, 18.8.2013.
- [21] Званични преглед Алфреско багова и недостатка на JIRA-и, профилиран за потребе овог рада, <http://goo.gl/U0yIls>, 23.9.2013.
- [22] Зоран Марјанов, Марија Минић, Небојша Тешић, Јелена Јовановић-Бабић, „Алфреско 3.3, преглед“, Институт „Михајло Пупин“, Београд, Србија, 2009.
- [23] Jeff Potts, „*Alfresco Developer Guide*“, Packt Publishing, Бирмингем, Енглеска, 2008.
- [24] Јелена Јовановић-Бабић, Марија Минић, Зоран Марјанов, Небојша Тешић, Јасмина Стевановић, „Пример модела складишта података у Алфреско CMS“, Infotech 2009, Београд, Србија, 2009.
- [25] Gordana M Pavlović-Lažetić, Nenad S Mitić and Miloš V Beljanski, „*Bioinformatics analysis of SARS coronavirus genome polymorphism*“, BMC Bioinformatics, 2004, 5:65.
- [26] Дефиниција *Cygwin* софтвера на енглеској википедији, <http://en.wikipedia.org/wiki/Cygwin>, 6.9.2014.

Математички факултет у Београду, Катедра за рачунарство и информатику

[27] Званична страница *cURL* софтвера, <http://curl.haxx.se/>, 6.9.2014.