

Univerza
v Ljubljani
Fakulteta
za gradbeništvo
in geodezijo



Jamova cesta 2
1000 Ljubljana, Slovenija
<http://www3.fgg.uni-lj.si/>

DRUGG – Digitalni repozitorij UL FGG
<http://drugg.fgg.uni-lj.si/>

V zbirki je izvirna različica izdajatelja.

Prosimo, da se pri navajanju sklicujete na bibliografske podatke, kot je navedeno:

University
of Ljubljana
Faculty of
Civil and Geodetic
Engineering



Jamova cesta 2
SI – 1000 Ljubljana, Slovenia
<http://www3.fgg.uni-lj.si/en/>

DRUGG – The Digital Repository
<http://drugg.fgg.uni-lj.si/>

This is a publisher's version PDF file.

When citing, please refer to the publisher's bibliographic information as follows:

Šušteršič, K., Podobnikar, T. 2013. Prostorski ETL za boljšo medopravilnost v GIS = Spatial ETL tool for better interoperability in GIS. Geodetski vestnik 57, 4: 719-733.

DOI: <http://dx.doi.org/10.15292/geodetski-vestnik.2013.04.719-733>
<http://drugg.fgg.uni-lj.si/4593/>

Datum arhiviranja / Archiving Date: 14-09-2014

PROSTORSKI ETL ZA BOLJŠO MEDOPRAVILNOST V GIS

SPATIAL ETL TOOL FOR BETTER INTEROPERABILITY IN GIS

Katja Šušteršič, Tomaž Podobnikar

UDK: 528.88:711.4(497.4)

Klasifikacija prispevka po COBISS-u: 1.04

IZVLEČEK

V prispevku predstavimo prostorsko ETL-orodje, ki omogoča medopravilnost prostorskih in neprostorskih podatkov. Skovanka prostorsko ETL-orodje kaže, da orodje omogoča uporabnikom obdelavo in preoblikovanje prostorskih podatkov med podatkovnimi formati. Kratica ETL (angl. extract, transform, load) se nanaša na aktivnost izbora, preoblikovanja in nalaganja podatkov. Uporaba prostorskih podatkov je postala del vsakdanjika, vendar lahko le s pravilno uporabo dosežemo pravo vrednost, ki jo lahko ponudijo uporabnikom. Namen prispevka je predstavitev sposobnosti in uporabnosti prostorskega ETL-orodja, podrobnejša opredelitev ETL-postopka, seznanitev s prostorskim ETL-orodjem FME Desktop ter predstavitev uporabnosti na dveh primerih. V prvem primeru smo vzpostavili poenoteno podatkovno skladišče iz nehomogenih skladišč za ugotovitev vpliva geološke podlage na poškodovanost stavb pri potresu leta 2004. V drugem primeru smo opisali obveščanje lokalnih prebivalcev o predvidenih spremembah v prostoru. S tem smo potrdili fleksibilnost orodja pri urejanju in analizi prostorskih podatkov ter pokazali, da je prostorski ETL primeren za hitrejšo in učinkovitejšo rešitve pri delu s podatki.

KLJUČNE BESEDE

prostorski ETL, medopravilnost, ETL, FME Desktop, uporabnost prostorskega ETL-orodja

ABSTRACT

A spatial ETL tool that allows interoperability between spatial and non-spatial data is presented in this article. The primary goal of the tool is to provide spatial data processing and transformation among various data formats. This is made possible by the ETL process, which extracts, transforms and loads data. The use of spatial data has become significant in everyday life, because only correctly applying the data enables users to extract the true value spatial data offers. The main purpose of this article is to demonstrate the capability and usability of the spatial ETL tool, in order to introduce a more detailed definition of the ETL process to acquaint the reader with the FME Desktop tool, and to demonstrate the applicability of the tool in two case studies. In the first case, a unified spatial data warehouse is built from non-homogeneous data warehouses in order to assess the impacts and effects the geological basis had on the amount of damage to buildings in the 2004 earthquake. The second case demonstrates how the spatial ETL tool can be used to inform locals of predicted spatial changes in the area. The flexibility and the efficiency of the spatial ETL tool are successfully demonstrated in both cases; ETL turns out to be a robust tool for editing and analysing data.

KEY WORDS

spatial ETL, interoperability, ETL, FME Desktop, serviceability of spatial ETL tool

1 UVOD

Poznavanje in razumevanje uporabnosti prostorskih podatkov je v današnjem informacijskem svetu bistvenega pomena. S prostorskimi podatki ne usklajujemo le različnih interesov prostorskega razvoja, ampak so tudi pomemben člen pri varovanju okolja in naravnih vrednot, obvladovanju naravnih procesov ter ne nazadnje v pomoč pri varovanju človeka in njegovih

dejavnosti pred naravnimi nesrečami (Šušteršič, 2012). Preden uporabnik začne uporabljati prostorske podatke, se mora sreča s preprekami, kot so lastništvo podatkov, različni podatkovni formati, kakovost podatkov, namen, za katerega so bili podatki zajeti, programska oprema itd. (Safe Software, 2010). Dodatna težava je vedno večja količina podatkov, pri katerih pa je vprašljiva njihova dostopnost in uporabnost za posamezen namen. Postopek oziroma orodje ETL se uveljavlja kot podpora pri pripravi in urejanju (npr. medopravilnosti in harmonizaciji) različnih vrst podatkov. Angleška kratica ETL pomeni »extract, transform, load«, torej ponazarja proces, ki obsega aktivnosti izbora, preoblikovanja in dostave podatkov (Kimball in Caserta, 2004; Safe Software, 2012).

Nadgradnja splošnega ETL je prostorski ETL kot podpora urejanju podatkov v geografskih informacijskih sistemih (GIS), oziroma še več, kot podpora vzpostavitvi prostorske podatkovne infrastrukture, kot je na primer INSPIRE. Uporabnikom prostorskih podatkov je torej v pomoč pri vgradnji podatkov različnih formatov v njihove lastne aplikacije. Poleg pretvarjanja med formati vsebuje tudi bogato zbirko orodij, ki omogočajo čiščenje, poenotenje in preoblikovanje podatkov, kar je privedlo do tega, da veliko uporabnikov izkorišča prostorsko ETL-orodje za obdelovanje podatkov v istih formatih (Murray, 2008). V prispevku obravnavamo prostorski ETL, ki se je inženirjem, prostorskim načrtovalcem, GIS-strokovnjakom ponudil kot orodje za enostavnejšo uporabo prostorskih podatkov za analizo, vizualizacijo, načrtovanje prostora ter se lahko uporablja kot posrednik za obveščanje javnosti pred spremembami v prostoru.

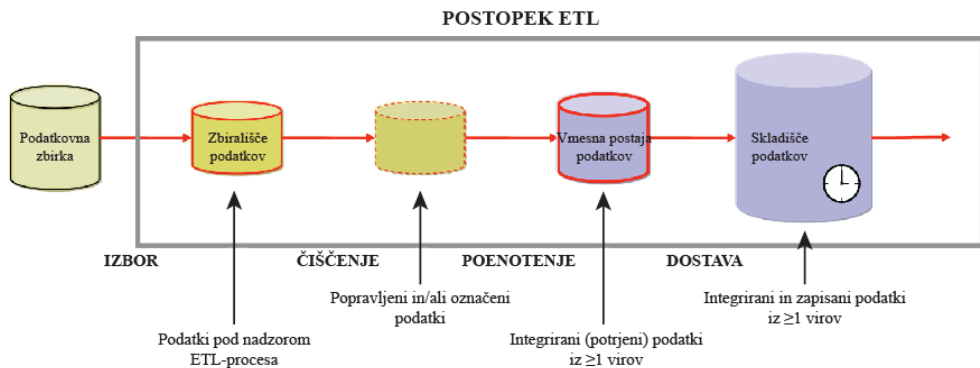
2 POSTOPEK ETL

Kot smo navedli, pomeni angleška kratica ETL aktivnost izbora, pretvorbe in dostave podatkov. Kimball in Caserta (2004) razčlenita postopek ETL na aktivnosti izbora, čiščenja, poenotenja in dostave podatkov (ECCD, angl. extract, clean, conform, deliver). Čeprav opisane stopnje postopka ne sovpadajo s kratico ETL, bomo ohranili to poimenovanje postopka namesto manj uveljavljene kratice ECCD in pri tem večinoma uporabljali razčlenjeni postopek. Na sliki 1 je prikazano polnjenje podatkovnega skladišča po postopku ETL.

Namen postopka ETL je realizacija zalednega sistema podatkovnega skladišča, ki zagotavlja vhodne podatke področnim podatkovnim skladiščem. V postopku transformacije podatkov, ki poteka od izvornih heterogenih podatkovnih virov do končnih podatkov, urejenih v večdimenzijskem podatkovnem modelu, je pomembno zagotoviti njihovo zaščito, vsebino transformacij pa dokumentirati. Kakovostno zasnovan ETL v aktivnostih čiščenja in poenotenja podatkom doda informacijsko vrednost ter zagotovi eno in edino različico (Kralj, 2010).

Sodobna ETL-orodja so tako v pomoč pri naslednjih fazah (Žefran, 2009):

- priključitev na poljuben vir ne glede na izvor podatkov;
- preverjanje kakovosti podatkov in združevanje podatkov iz večjih sistemov;
- hitrost prenosa podatkov (zmožnost paralelnega delovanja omogoča prenose ogromnih količin podatkov v kratkem obdobju);
- pregled izvora in celotnega tokokroga podatkov.



Slika 1: Postopek ETL (Marti, 2011, str. 4)

2.1 Opredelitev postopka ETL po korakih

Pravilno zasnovan ETL-postopek omogoča vnos podatkov iz različnih virov podatkovnih skladišč, s čiščenjem zagotavlja kakovost podatkov in skladnost standardov, omogoča poenotenje podatkov na način, da se lahko podatki iz različnih podatkovnih virov uporabljajo kot celota, ter nazadnje dostavi podatke v format po uporabnikovem izboru za nadaljnjo nadgradnjo in uporabo (Kimball in Caserta, 2004). V tem poglavju nekoliko podrobneje opisujemo štiri korake procesa ETL:

- izbor: zberemo podatke iz heterogenih podatkovnih virov;
- čiščenje: zagotovimo kakovost podatkov s čiščenjem;
- poenotenje: poenotimo mere in nazive atributov vseh uporabljenih podatkovnih virov;
- dostava: izvedemo dostavo podatkov v obliki večdimenzionalnega podatkovnega modela.

2.1.1 Aktivnost izbora podatkov

Aktivnost izbora podatkov izvedemo po nalogah, ki jih Kralj (2010) razdeli na:

- *Opredelitev virov podatkov:* odločimo se za uporabo tistih podatkov, ki bodo pripomogli k boljšemu končnemu odločanju uporabnikov.
- *Analiza kakovosti podatkov:* kakovost vsebine prostorskih podatkov oziroma podatkov na splošno lahko preverimo s pravilnostjo podatkov, ki se kažejo z njihovo skladnostjo z nominalo osnovo, z nedvoumnostjo (to razlagamo s preprečitvijo možnosti, da bi podatke razumeli napačno), s skladnostjo (ta določa, da ista zapisana informacija nima različnih vrednosti) ter s popolnostjo (tukaj preverjamo, da imajo podatki celotno vrednost posameznih atributov ali zapisov). Pri analizi kakovosti podatkov nam je v pomoč standard SIST EN ISO 19113, ki določa kakovostna načela za objektivno določanje kakovosti prostorskih podatkov. Postopke za ocenjevanje kakovosti podatkov pa najdemo v standardu SIST EN ISO 19114.
- *Pregled transformacij:* pomeni vpogled na nastavljene transformacije, pri katerih so že upoštevana prostorska pravila. Pozornosti ne posvečamo njihovem kodiranju.
- *Pregled dimenzijskega modela:* podatkovni model določimo že pred aktivnostmi ETL-postopka, v tej fazi preverimo, ali model izpolnjuje zahteve končnega podatkovnega skladišča.

- *Kontrola merljivih dejstev in izračunov*: treba je preveriti izbor merljivih dejstev in izvesti kontrolo nastavljenih kalkulacij. V pomoč pri preverjanju je lahko končni uporabnik sistema, ki pomaga pri preverjanju, kajti pridobljeni podatki morajo biti dobra podlaga za nadaljnjo analizo.

2.1.2 Čiščenje podatkov

Čiščenja se lotimo s testiranjem, ki potekajo med samim postopkom ETL. Testiranja omogočijo odkrivanje neakovostnih ali napačnih podatkov, ki jih nato vsebinsko popravimo (Kimball in Caserta, 2004). Med napačne podatke spadajo ničelne vrednosti v atributih, kjer takšna vrednost tam ni dovoljena, numerične vrednosti izven dovoljenega intervala, izločitev vrednosti, ki smo jih predhodno označili kot nedovoljene, ter tekstovne napake, ki jih preverjamo s slovarjem. Ko preverimo posamezne vrednosti, je potreben pregled struktur, s katerim odkrivamo napačne ali manjkajoče povezave hierarhično urejenih podatkov in napačne vrednosti povezanih podatkov (Šušteršič, 2012).

2.1.3 Poenotenje podatkov

Poenotenje podatkov se od čiščenja najbolj razlikuje po tem, da gre tu zgolj za prenose podatkov in spreminjanje strukture, ki ne vplivajo neposredno na rezultate. Opravka imamo z dvema vrstama opravil. Prvo se nanaša na poenotenje dimenzijskih tabel, ki omogočajo povezovanje področnih podatkovnih skladišč. Dimenzijske tabele (angl. dimension table) so tabele v dimenzijskem podatkovnem modelu, ki opisujejo transakcije iz tabele dejstev. Tabela dejstev (angl. fact table) pa je središče dimenzijskega podatkovnega modela, njen osnovni element je zrno, ki določa velikost tabele (Kimball in Caserta, 2004). Kako zahtevno je lahko poenotenje, je odvisno od strukture in kakovosti podatkovnih virov. Izvaja se iz več tabel, ki so lahko v lasti različnih sistemov. Iz tabel nato izluščimo vse podatke, ki pripadajo posameznemu subjektu, in jih združimo v enoten zapis. Končni rezultat je poenotena dimenzijska tabela, ki zagotavlja skladno vsebino. Drugo opravilo se veže na poenotenje merljivih dejstev, ki ga lahko izvedemo v treh korakih. V prvem standardiziramo podatkovno skladišče, kar se izvede z opredelitvijo veljavnih vrednosti za vsakega izmed neskladnih atributov. Drugi korak je iskanje podvojenih zapisov, ki pa je v realnosti manj pogost pojav. V tretji korak spada združevanje zapisov, pri katerem se upoštevajo podvojeni zapisi. Tukaj določimo, katere attribute iz podvojenih vrednosti bomo obdržali in katere zavrgli (Šušteršič, 2012).

2.1.4 Dostava podatkov

Podatki so primerni za dostavo, ko so pretvorjeni v želeni format, želeni podatkovni model in koordinatni sistem. Podatki se po predhodnem definiranju v vnaprej določeni podatkovni format izpišejo v definirano podatkovno bazo ali v prikazno okno (angl. data view) (Safe Software, 2010).

Običajno večino delovnega napora vložimo že pri predhodnih aktivnostih. Ob upoštevanju pravil dimenzijskega podatkovnega modela vzame dostava podatkov manj časa (Šušteršič, 2012).

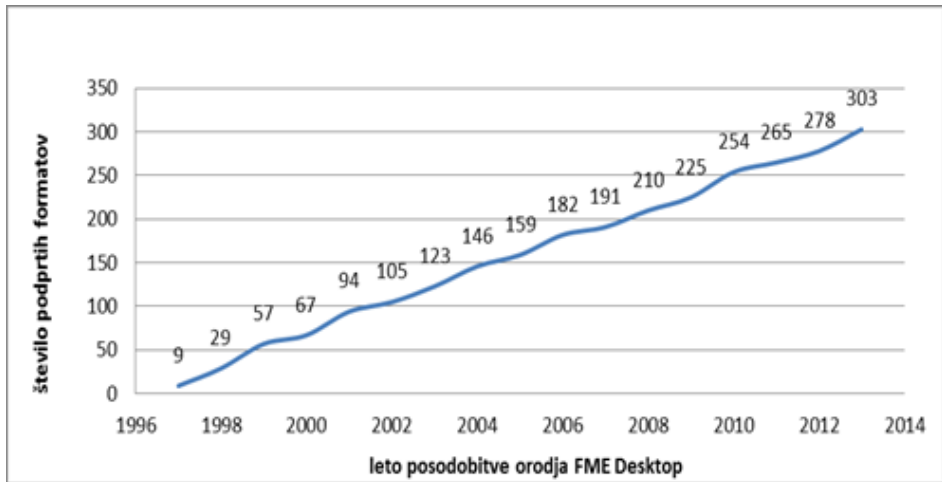
3 PROSTORSKO ETL-ORODJE

Prostorsko ETL-orodje ponuja prostorskim analitikom tisto, kar splošna ETL-orodja ponujajo poslovnemu svetu, tj. vključitev nepovezanih podatkovnih zbirk v enotno delovno okolje, kar okrepi poslovno znanje in posledično omogoča večjo konkurenčnost organizacije na trgu. Prostorsko ETL-orodje zagotavlja podobno funkcionalnost obdelave podatkov po načelu splošnega ETL, s primarnim poudarkom na upravljanju in preoblikovanju prostorskih podatkov. Prostorsko ETL-orodje je mehanizem, ki omogoča prostorskim načrtovalcem vpogled v ozadje različnih podatkovnih baz, kar omogoča uporabnikom pogled v lastnosti podatkov. Ključno uporabnost prostorskega ETL-orodja najdemo v njegovi medopravilnosti. Njeno bistvo je v izmenjavi podatkov, ki jih delimo ali sprejemamo od različnih uporabnikov, ter sposobnosti, da lahko dane podatke pregledno uporabljamo. Veliko prostorskih podatkov, ki jih imamo na voljo, namreč ni v pomoč, če jih ne znamo pravilno izkoristiti (Cragg, 2009a). Medopravilnost opredelimo kot tehnično sposobnost različnih sistemov in organizacij, da med seboj sodelujejo in si izmenjujejo podatke (Wikipedia, 2013). V standardu SIST EN ISO 19119 je geoinformacijska medopravilnost opredeljena kot tehnična sposobnost informacijskega sistema, da lahko:

- uspešno komunicira na različnih ravneh povezovanja med uporabniki in drugimi sistemi;
- prosto izmenjuje razne vrste in oblike prostorskih (geografskih) podatkov o objektih oziroma pojavih na zemeljskem površju, nad ali pod njim;
- povezano prek omrežij uporablja in poganja programsko opremo, ki je sposobna upravljati porazdeljene prostorske podatke.

GIS-tehnologija potrebuje ETL-orodje tudi za omogočanje medopravilnosti ali izmenjave podatkov med različnimi strokami, kot so kartografija, daljinsko zaznavanje, geodezija in prostorsko načrtovanje. Tako si lahko različne organizacije prek prostorskega ETL-orodja nemoteno izmenjujejo podatke z različnimi formati (ESRI, 2012). V praksi se lahko prostorsko ETL-orodje uporablja za preverjanje obstoječih prostorskih okolij državnih okvirov, ki morajo ustrezati evropskim ali svetovnim normativom. Primer takšnega projekta je GeoTest, ki so ga izvedli na Švedskem (Östman et al, 2009). Kot bistvena za uspeh prostorskega ETL-orodja se je pokazala njegova nevsiljiva tehnologija, ki omogoča podatkom, da ostanejo, kjer so, kajti orodje ne zahteva njihovega fizičnega premika za namen pretvorbe (Murray, 2008).

Sprva je bil namen prostorskega ETL-orodja preprosto nalaganje, pretvarjanje in dostava podatkov med različnimi formati. Z razvojem in večanjem potreb uporabnikov po bolj zmogljivih programskih orodjih pa so postala današnja ETL-orodja veliko več. Zanimivo je tudi, da vedno več uporabnikov uporablja prostorsko ETL-orodje za preoblikovanje podatkov znotraj istega formata (Murray, 2008). Kljub močnim orodjem, ki omogočajo uporabnikom preoblikovanje podatkov, je glavna funkcija ETL-orodja še vedno število raznovrstnih formatov, ki jih posamezno ETL-orodje vključuje. Trend naraščanja vgrajevanja novih formatov je razviden na sliki 2.



Slika 2: Naraščanje vgradnje števila formatov po letih v prostorsko ETL-orodje FME Desktop

3.1 Izbor programskega orodja

Na spletnem strežniku smo poiskali najbolj razširjena prostorska ETL-orodja. Za primerjavo smo vzeli naslednje programske pakete: FME Desktop, Intergraph Geospatial, GeoServer, GeoKettle, Spatial Data Integrator.

Za lažji pregled nad ponujenimi prostorskimi ETL-orodji smo upoštevali splošne lastnosti obravnavanih programov (leto izdaje, zadnjo različico programa, ki je na voljo na trgu, navedbo avtorja orodja), opis značilnosti programskega paketa (splošen opis programa, odprtost programskega orodja), navedbo spektra formatov, ki jih program podpira, vrsto uporabnika glede na raven zahtevnosti uporabe programa¹ ter vrsto licence za odprtokodna programska orodja.

3.1.1 FME Desktop

Leto izdaje: 1996.

Zadnja različica izdaje: FME Desktop 2013.

Avtor orodja: Safe Software.

Značilnosti: Prostorsko fleksibilno močno programsko orodje, narejeno za hitro pretvarjanje med formati, transformacijo različnih formatov in integracijo podatkov v različne aplikacije. Program je vgrajen v Autodesk AutoCAD, Bentley, ESRI ArcGis, Google Earth/Maps, MapInfo Professional itd. (Safe Software, 2013).

Dokumentacija: brezplačno spletno učenje, <http://fmepedia.safe.com/knowledgeDocumentation>.

Podpora formatov: omogoča branje in pisanje podatkov v več kot 300 prostorskih in neprostorskih formatih².

¹ Uporabniška raven: začetnik (branje), izkušen uporabnik (urejanje, preproste analize), strokovnjak (vse vrste analiz), raziskovalec (programiranje).

² Celoten seznam formatov FME Desktop: <http://www.safe.com/fme/format-search/>.

Uporabniška raven: od začetnika do strokovnjaka.

Licenca za programsko opremo: lastniški paket.

3.1.2 Intergraph Geospatial

Leto izdaje: 1997.

Zadnja različica izdaje: Intergraph Geospatial 2013.

Avtor orodja: Intergraph.

Značilnosti: Intergraph Geospatial 2013 je celovita rešitev za uspešno integracijo prostorskih podatkov iz različnih prostorskih strok, kot so: fotogrametrija, daljinsko zaznavanje in GIS. Uporabniku je na tako na voljo ogromna količina podatkov iz vsebin različnih podatkovnih virov, ki jih nato lahko hitro in varno deli ter jih dostavi končnim uporabnikom kot zanesljive in točne podatke (Intergraph, 2013a; Intergraph, 2013b).

Dokumentacija: plačljivi spletni tečaji.

Podpora formatov: široka podpora vektorskih in rastrskih formatov³.

Uporabniška raven: od začetnika do strokovnjaka.

Licenca za programsko opremo: lastniški paket.

3.1.3 GeoServer

Leto izdaje: 2001.

Zadnja različica izdaje: 2.3.3.

Avtor orodja: OpenGeo, GeoSolutions, Refrations Research.

Značilnosti: Orodje GeoServer je odprtokodni program, zapisan v Javi, ki omogoča ogled in urejanje prostorskih podatkov. Prednost orodja je prav v odprtokodnosti, saj lahko uporabnik razpolaga s kartami iz vseh aktualnih kartografskih aplikacij, kot so Google Maps, Google Earth, Yahoo Maps in Microsoft Virtual Earth. Prav tako se lahko poveže z licenčnimi programi, kot je ESRI ArcGIS. Uporablja integrirano knjižnico kart Openlayers in Geotools kot GIS orodje (GeoServer, 2013).

Dokumentacija: bogata dokumentacija priročnikov; <http://docs.geoserver.org/>.

Podpora formatov: knjižnica GDAL.

Uporabniška raven: od začetnika do raziskovalca.

Licenca za programsko opremo: GPL.

³ Seznam formatov, ki jih podpira orodje Intergraph Geospatial: <http://oasis.massey.ac.nz/erdas-apollo/welcome/en/acknowledgements.html>.

3.1.4 Spatial Data Integrator

Leto izdaje: 2007.

Zadnja različica izdaje: 5.3.0.

Avtor orodja: Talend.

Značilnosti: ETL-orodje, opredeljeno kot prostorski povezovalnik podatkov. Programska orodja temeljijo na Talend Open Studio (TOS) za branje in pisanje GIS-formatov, manipulacijo funkcij, metapodatkovno ustvarjanje in objavlanje. Program nima tako bogate podpore zbirk GIS-formatov (Spatial Data Integrator, 2013; Talend, 2013).

Dokumentacija: spletni priročniki s primeri; <http://www.talendforge.org/wiki/doku.php?id=sdi:mainpage#tutorials>.

Podpora formatov: seznam formatov, ki jih podpira orodje⁴.

Uporabniška raven: od začetnika do raziskovalca.

Licenca za programsko opremo: GPL.

3.1.5 GeoKettle

Leto izdaje: 2009.

Zadnja različica izdaje: 2.5.

Avtor orodja: Pentaho Data Integration.

Značilnosti: Močno, metapodatkovno usmerjeno prostorsko ETL-orodje, namenjeno vključevanju različnih prostorskih podatkovnih virov za izgradnjo in posodobitev prostorskih podatkov za skladiščenje. GeoKettle omogoča pridobivanje podatkov iz podatkovnih virov, preoblikovanje podatkov za namen odpravljanja napak, čiščenje podatkov, spremembo strukture podatkov. Uporablja odprtokodne knjižnice, kot so: JTS, GeoTools, OGR in Sextante (GeoKettle, 2013).

Dokumentacija: bogata spletna dokumentacija s priročniki; <http://wiki.pentaho.com/display/EAI/Latest+Pentaho+Data+Integration+%28aka+Kettle%29+Documentation>.

Podpora formatov: SHP, GML, KML, OGR-knjižnica.

Uporabniška raven: od začetnika do raziskovalca.

Licenca za programsko opremo: LGPL.

3.2 Predstavitev orodja FME Desktop

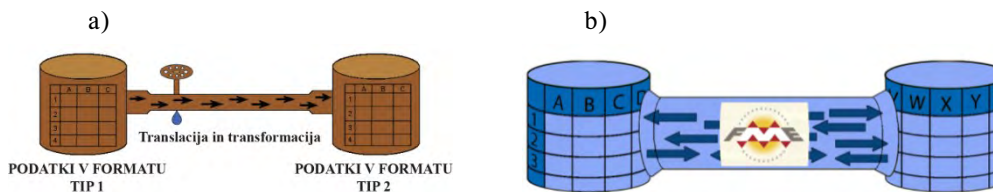
Pred drugimi prostorskimi orodji prevladuje Feature Manipulation Engine (v nadaljevanju: FME), ki je delo Safe Softwara. Posebna obravnava orodja FME Desktop temelji na dejstvu, da je vgrajeno v ArcGIS, ki je najbolj razširjen komercialni program za delo z GIS-i. FME Desktop je produkt podjetja Safe Software, ki je skovalo poimenovanje prostorskega ETL-

⁴ Celoten seznam formatov SDI: <http://www.talendforge.org/wiki/doku.php?id=sdi:geocomponentlist>.

orodja, saj so pionirji v prostorski ETL-tehnologiji. Orodje ponuja pretvorbo med več kot 300 prostorskimi in neprostorskimi formati, neomejeno prilagodljivost pri prestrukturiranju, predelavi formatov in integracijo lastnih prostorskih podatkov. Za povrh omogoča prostorsko ETL-orodje medopravilnost vseh teh različnih podatkov v enem samem delovnem okolju (Safe Software, 2012), v našem primeru v FME Desktop. Prav različni podatkovni formati so velika tehnična težava za prostorske načrtovalce in analitike, zato je takšen program odgovor na težave, ki jih povzročajo različni podatkovni formati. FME Desktop je vključen v številna programska orodja, kot so: Autodesk AutoCAD, Bentley, ESRI ArcGis, Google Earth/Maps, Intergraph GeoMedia Professional, MapInfo Professional, OpenStreetMap, Smallworld, Trimble itd. Orodje vsebuje veliko podprtih prostorskih in neprostorskih formatov ter knjižnico z več kot 400 pretvorniki (angl. transformers), ki dajejo uporabnikom veliko možnosti za preoblikovanje njihovih podatkov. FME je tudi zelo prilagodljiva in celovita tehnologija, zato je smiselna njena vključitev pri dodajanju novih funkcij v druga programska orodja (Safe Software, 2012).

FME vsebuje tri ključne značilnosti, ki pomagajo obvladovati težave, običajne pri podatkovnih pretvorbah (Cragg, 2009b):

- Omogoča enostavno dodajanje novih formatov na seznam obstoječih formatov v programskem orodju.
- Ima določene lastnosti semantičnosti, kar pomeni, da lahko razlikuje med različnimi podatkovnimi tipi in formati. Ta lastnost mu omogoča izvedbo postopkov, pri katerem se med izvajanjem pretvorbe ne izgubljajo podatki.
- FME opisujemo kot »debelocevno prevajanje« (angl. thick pipe translations) podatkov, kar se nanaša na dejstvo, da se lahko podatki preoblikujejo in prestrukturirajo že med samim procesom prevajanja. Na sliki 3 sta predstavljena dva načina prevajanja podatkov.



Slika 3: Prikaz prevajanja in preoblikovanja podatkov iz formata 1 v format 2 v tanki (a) in debeli (b) cevi (Cragg, 2009b)

3.3 Primera uporabe FME Desktop

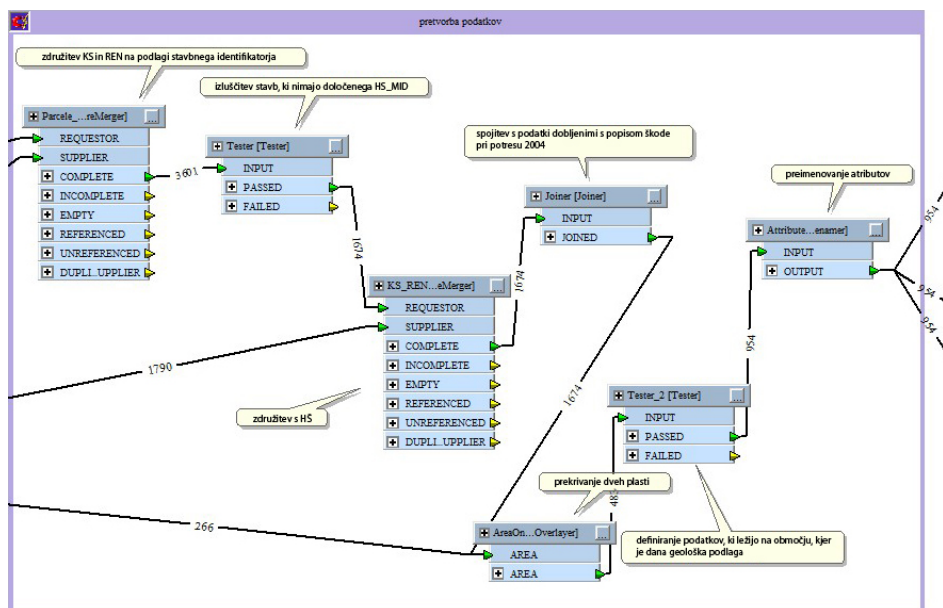
V prispevku smo obravnavali dva primera uporabe orodja FME Desktop. V prvem obravnavamo vpliv geološke podlage na poškodovanost stavb v občini Bovec za potres iz leta 2004. Ugotavljali smo ga na podlagi primerjave in poenotenja različnih podatkovnih skladišč, ki smo jih pridobili prek različnih organizacij. Drugi primer je uporaba orodja FME Desktop kot posrednega sredstva za namen obveščanja o predvidenih prostorskih spremembah na območju občine Kilkenny na Irskem.

3.3.1 Preverjanje vpliva geološke podlage na poškodovanost stavb v občini Bovec za potres iz leta 2004

Preverbo smo izvedli z izgradnjo poenotnega podatkovnega skladišča, ki smo ga zgradili z orodjem FME Desktop. Orodje omogoča vnos podatkov različnih formatov v eno samo delovno okolje, kar uporabnikom omogoča enostavno uporabo široke palete prostorskih in neprostorskih podatkov. V preglednici 1 so povzeti podatki, ki smo jih pridobili prek različnih ustanov.

Upravljevac vira	Podatkovno skladišče	Format
Uprava RS za zaščito in reševanje	podatki o hišnem naslovu s podatki o številu oseb s stalnim prebivališčem v posamezni stavbi	XLS
	vrednost ocenjene škode posameznega objekta za potres iz leta 2004	
Geološki zavod	geološka karta 1 : 10.000 za Bovško kotlino	DWG
GURS	hišne številke	SHP
	kataster stavb	SHP
	register nepremičnin	SHP

Preglednica 1: Prostorski podatki in njihovi viri (Šušteršič, 2012)

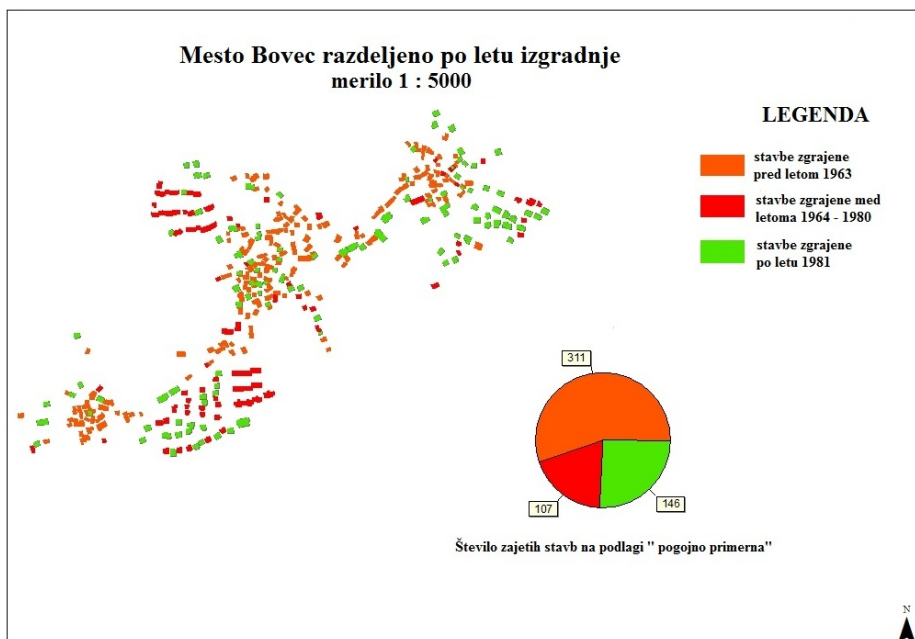


Slika 4: Prikaz pretvornikov, potrebnih za poenotenje v prostorskem orodju FME Desktop (Šušteršič, 2012).

Poenotenje nehomogenih podatkov iz različnih podatkovnih skladišč v orodju FME Desktop je prikazano na sliki 4. Z združitvijo skladišč smo dosegli, da ima vsak poligon stavbe v občini Bovec dodeljene attribute, ki so v pomoč pri nadaljnji analizi. Nadaljnje poizvedovanje po vzpostavljeni bazi podatkov smo izvedli v orodju ArcMap in analizirali dane podatke za dokazovanje vpliva

geološke podlage na poškodovanost stavb pri potresu. Predhodno smo obravnavali študijo, ki je bila predstavljena v članku Seizmološki in geotehnični pogoji gradnje v zgornjem Posočju (Ribičič et al., 2000) in na podlagi katere smo določili primernost posamezne podlage za gradnjo objektov na območju občine Bovec. Za primerjalno vrednost nam je služila ocenjena vrednost škode na posameznem objektu, ki so jo določili cenilci pri popisu škode leta 2004. Za analizo smo uporabili prostorska in statistična orodja, ki so na voljo v ArcGISu, ter pri rezultatih upoštevali tudi dejavnika *leto izgradnje* in *material gradnje*.

Za potrditev teze o vplivu geološke podlage na poškodovanost stavb smo poizvedovali po različnih tipih podlage, primernih za gradnjo. Na slikah 5 in 6 sta prikazani analizi za *pogojno primerno* podlago. V obravnavo smo vzeli 564 stavb ter preverili leto izgradnje stavbe in material gradnje.



Slika 5: Mesto Bovec, razdeljeno po letu izgradnje (Šušteršič, 2012).

Rezultati analize so presenetljivo pokazali, da je največ škode nastalo na območju podlage, ki je na inženirsko-geološki karti označena kot zelo primerna in uporabna. Nadaljnje poizvedovanje razkrije, da več kot polovica teh stavb ni protipotresno zgrajena. Na pogojno primerni podlagi so rezultati analize pokazali, da je bila ocenjena vrednost poškodb na objektih, ki so bili protipotresno grajeni (po letu 1981), približno enaka vrednosti poškodb objektov, pri katerih protipotresna gradnja ni bila zakonsko predpisana (pred letom 1963). Rezultati so sicer zanimivi, vendar je pri njihovem interpretiranju treba upoštevati, da so bili v analizo vključeni podatki le iz potresa iz leta 2004. Za večjo zanesljivost analize bi potrebovali podatke iz več predhodnih potresov (Šušteršič, 2012). Tako smo na preprostem primeru dokazali uporabnost ETL-orodja ter prikazali medopravnost podatkov in njihovo uporabnost v prostorskih analizah. Dolgoročni učinek takšnega vzpostavljenega podatkovnega skladišča je preglednejša in enostavnejša

uporaba prostorskih podatkov pri upravljanju podatkov, občini je lahko vzpostavljena baza v pomoč pri načrtovanju dopolnitev občinskih prostorskih načrtov in pri načrtovanju komunalne infrastrukture.



Slika 6: Graf povprečne škode glede na leto izgradnje stavb, zgrajene na pogojno primernih tleh (Šušteršič, 2012).

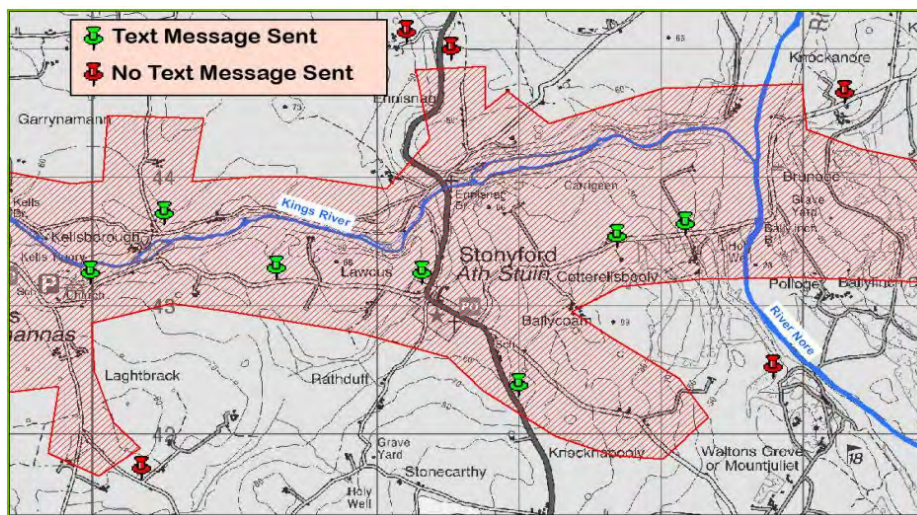
3.3.2 Primer obveščevalnega sistema za predvidene spremembe v prostoru

Drugi primer uporabe orodja FME Desktop je rešitev, ki jo je leta 2009 Brendan Cunningham predstavil prebivalcem v občini Kilkenny na Irskem. V orodju FME Desktop je našel rešitev, ki omogoča obveščanje o spremembah v prostoru v občini. Prebivalci imajo možnost brezplačnega vpisa v občinski obveščevalni sistem prek spletne strani. Prek aplikacije Google Maps določijo koordinate svojega položaja, nato se registrirajo v bazo. Glede na podane koordinate FME Desktop primerja dane poligone glede na uporabnikovo lokacijo. Uporabijo se slednji pretvorniki:

- izvedba prekritja točkovnega sloja z območjem (angl. point-on-area-overlay);
- prostorski filter in prostorski povezovalnik (angl. spatial filter and spatial relator);
- orodje najbližji sosed (angl. nearest neighbour);
- pretvornik vmesno območje (bafer).

V primeru preteče nevarnosti ali spremembe v prostoru se uporabnik sistema obvesti o spremembi, ki se bo zgodila v prostoru. Obvestijo se le uporabniki, ki so neposredno ogroženi (uporaba funkcije bafer). Za obveščanje uporabnikov se uporabi SQL Server Trigger, sistem za samodejno pošiljanje tekstovnih sporočil na mobilni telefon (slika 7). Takšen način obveščanja

prihrani marsikatero nevšečnost, saj velikokrat spregledamo obvestila o zaprtju vodovoda, zaprtju ali obvozu posamezne ceste itd. (The FME Evangelist, 2013). Obveščanje se lahko nadgrajuje globalno, saj bi lahko prebivalce na širšem območju opozorili na nevarnosti cunamija, požarov in drugih nesreč. Do podobne ugotovitve so prišli tudi razvijalci aplikacije MapAlerter, ki temelji na FME Desktopu. Aplikacijo uporablja kar pet občin na Irskem (MapAlerter, 2013).



Slika 7: Območje nevarnosti s prikazom uporabnikov, ki so SMS obvestila dobili ali ne (Cunningham, 2009).

4 SKLEP

V prispevku je predstavljeno prostorsko ETL-orodje, njegove značilnosti in uporabnosti. Pri tem smo izpostavili razširjen postopek ETL, ki je sestavljen iz štirih korakov – aktivnosti izbora, čiščenja, poenotenja in dostave podatkov. Glavna sposobnost prostorskih ETL-orodij je v njihovi medopravilnosti, saj tako ETL-orodje omogoča uporabnikom uporabo širokega spektra prostorskih in neprostorskih podatkov. Prikazana je tudi primerjava med različnimi prostorskimi orodji, prek katerih smo se odločili za predstavitev orodja FME Desktop. Orodje omogoča pretvarjanje kar med 300 različnimi formati. Njegova moč se kaže tudi v vgradnji mnogih formatov v eno programsko orodje ter v številnih pretvornikih, ki omogočajo preoblikovanje podatkov ter ne nazadnje izvoz oziroma pretvorbo podatkov v format, ki ga izbere uporabnik.

Na preprostem primeru smo predstavili uporabnost orodja FME Desktop za primer poenotenja različnih podatkovnih skladišč v povezavi z vplivom geološke podlage na poškodovanost stavb v občini Bovec za potres iz leta 2004. Potrdili smo enostavnost poenotenja podatkov različnih formatov, kot so XLS, SHP in DWG, v enotno podatkovno skladišče. Dodatno uporabnost orodja prikazuje primer iz občine Kilkenny na Irskem, kjer je prostorsko ETL-orodje FME posrednik, ki omogoča prebivalcem občine obveščanje o spremembah v prostoru. Tukaj se izkaže dodatna vrednost in uporabnost orodja v uporabi z družabnimi omrežji, ki jih je mogoče izkoriščati kot obliko obveščanja.

Prihodnost za prostorski ETL je spodbudna, saj zasnova omogoča vgradnjo najrazličnejših funkcionalnosti medopravilnosti v programska orodja. Krog uporabnikov prostorskega ETL-orodja je vedno širši, saj se uporabniki zavedajo moči, ki jo premorejo podatki, če so pravilno uporabljeni.

Literatura in viri:

- Cragg, R. (2009a). *ArcGIS Data Interoperability Basics*. Safe Software. Vancouver. Kanada.
- Cragg, R. (2009b). *Data Transformation with ArcGIS Data Interoperability Spatial ETL Tools*. Safe Software. Vancouver. Kanada.
- Cunningham, B. (2009). *Using FME to Drive a Location-Based Email / Cell Phone Notification Service*. FME International user conference. Whistler. Kanada.
- ESRI (2012). *ArcGIS Data Interoperability Extensions*. <http://www.esri.com/software/arcgis/extensions/datainteroperability/key-features/spatial-etl> (16. 8. 2012).
- FMEinsider (2013). *FME 2013 Has Landed*. <http://www.safe.com/about/newsletter/2013/Winter/founders-perspective-missions-mars-and-fme-2013/> (10. 2. 2013).
- FMEpedia (2012). *FME Desktop Documentation*. <http://fmeopedia.safe.com/knowledgeDocumentation> (12. 6. 2012).
- GeoKettle (2013). *A powerful open sourcespatial ETL tool*. <http://www.spatialytics.org/projects/geokettle/> (12. 8. 2013).
- GeoServer (2013). *GeoServer*. <http://geoserver.org/display/GEOS/What+is+Geoserver> (12. 8. 2013).
- Intergraph (2013a). *Geospatial Intergraph*. <http://geospatial.intergraph.com/2013-release.aspx> (12. 8. 2013).
- Intergraph (2013b). *Intergraph History*. http://www.intergraph.com/about_us/history_2000s.aspx (12. 8. 2013).
- Kimball, R., Caserta, J. (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. New York, Wiley Publishing, 161–170.
- Kralj, P. (2010). *Realizacija sistema poslovnega obveščanja v CPK d. d. Magistrsko delo*. Ljubljana, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko.
- MapAlerter (2013). *Aplikacija MapAlerter*. <http://www.mapalerter.com/> (27. 2. 2013).
- Marti, R. (2011). *Data Warehouse ETL – Processing*. <http://archive.systems.ethz.ch/www.systems.ethz.ch/education/fs11/data-wareh/slides/04-1%20DW%20Data%20Warehouse%20-%20ETL.pdf> (7. 9. 2013).
- Murray, D. (2008). *The growing Need for Spatial ETL*. *Geoinformatics*. <http://www.geoinformatics.com/blog/online-articles/the-growing-need-for-spatial-etl> (26. 11. 2012).
- Ribičič, M., Vidrih, R., Godec, M. (2000). *Seizmološki in geotehnični pogoji gradnje v zgornjem Posočju*. *Geologija* 43(1), 115–143.
- Safe Software (2013). *Solutions for applications*. <http://www.safe.com/solutions/for-applications/> (12. 8. 2013).
- Safe Software (2012). *Spatial ETL*. <http://www.safe.com/> (11. 6. 2012).
- Safe Software (2010). *Spatial ETL. Making Data Accessible*. *Whitepaper*.
- Spatial Data Integrator (2013). *Open source spatial ETL*. <http://www.docstoc.com/docs/27391255/Spatial-Data-Integrator-powered-by> (23. 2. 2013).
- Talend (2013). *Spatial extension for Talend*. <http://talend-spatial.github.io/> (12. 8. 2013).
- The FME Evangelist (2013). *Using FME to Drive a Location-Based Email / Cell Phone Notification Service: Brendan Cunningham, Kilkenny County Council*. <http://evangelism.safe.com/fmeevangelist34/> (9. 3. 2013).
- Standard SIST EN ISO 19113:2002 GI – *kakovostna načela (Quality principles)*.
- Standard SIST EN ISO 19114:2003 GI – *postopki za ocenjevanje kakovosti (Quality evaluation procedures)*.
- Standard SIST EN ISO 19119:2005 GI – *servisi (Services)*.
- Šuštersič, K. (2012). *Vzpostavitev prostorskega podatkovnega skladišča na primeru občine Bovec za potres 2004*. *Diplomsko delo*. Ljubljana, Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo, 58 str.

Wikipedia (2013). Interoperability. <http://en.wikipedia.org/wiki/Interoperability> (14. 8. 2013).

Žefran, E. (2009). Zagotavljanje kakovosti informacij pri podatkovnem skladiščenju. Diplomsko delo. Ljubljana, Univerza v Ljubljani, Ekonomska fakulteta (samozaložba E. Žefran), str. 5.

Östman, A., Abugessaisa, I., Tanyilli, S., He, X., El-Mekawy, M. (2009). GeoTest: A Testing Environment for Swedish Geodata. Gävle. Švedska.

Prispelo v objavo: 10. april 2013

Sprejeto: 18. september 2013

Katja Šušteršič, univ.dipl.inž.geod.

Blom Sweden AB, Hammarbacken 6B, 191 49 Sollentuna, Sweden

E-pošta: katja.sustersic@blomasa.com

doc. dr. Tomaž Podobnikar, univ. dipl. inž. geod.

Univerza v Ljubljani, FGG - Oddelek za geodezijo, Jamova 2, SI-1000 Ljubljana

E-pošta: tomaz.podobnikar@fgg.uni-lj.si