

MIKOVINY SÁMUEL FÖLDTUDOMÁNYI DOKTORI ISKOLA
Iskolavezető: DR. LAKATOS ISTVÁN AKADÉMIKUS

**GEOFIZIKAI ADATRENDSZEREK
SZABVÁNYOSÍTÁSA, EGYSÉGES GEOFIZIKAI
LEÍRÓNYELV ÉS ADATBÁZIS NAGYTÖMEGŰ
ADATRENDSZEREK KEZELÉSÉHEZ**

DOKTORI (PHD) ÉRTEKEZÉS

Írta:
SÓRÉS LÁSZLÓ

Tudományos vezető:
DR. TURAI ENDRE
egyetemi docens
a műszaki tudomány kandidátusa

Miskolci Egyetem
Geofizikai Tanszék
Miskolc
2011

Tartalomjegyzék

Abstract	3
Rövidítések.....	4
1. Előzmények.....	6
1.1. A GAIA programrendszer	7
1.2. A GEOMIND projekt.....	9
1.3. A KINGA projekt.....	10
2. Geofizikai adatok metaadat szabványa: a GEOMIND metaadat profil	12
2.1 A geofizikai adatrendszerek sokfélesége	12
2.2. A geofizikai adatrendszerek hierarchiája	14
2.3. A földrajzi adatokra vonatkozó ISO 19115 metaadat szabvány ismertetése	19
2.3.1. Metaadat entitás halmaz (Metadata entity set).....	22
2.3.2. Forrásadatok azonosítása (Identification)	23
2.3.3. Korlátozások (Constraints).....	24
2.3.4. Térbeli leírás (Spatial Representation).....	25
2.3.5. Vonatkoztatási rendszer (Reference System).....	25
2.3.6. Terjesztés (Distribution).....	26
2.3.7. Adatminőség (Data quality)	26
2.3.8. Metaadat kiterjesztés (Metadata Extention).....	27
2.4. Az ISO 19115 geofizikai kiterjesztése	27
2.4.1. Geofizikai metaadat osztályok definiálása	30
2.4.2. Geofizikai információk a GEOMIND profilban	34
2.4.3. XML implementáció	39
3. Az Általános Geofizikai Adatmodell (GGDM)	40
3.1 Geofizikai mérések általános leírása	43
3.1.1. Paraméterek és paraméter katalógusok	43
3.1.2 A mérés (GG_Measurement)	45
3.1.3. A helyi koordináta rendszer (GG_LocalCRS)	45
3.1.4. Terítési komponensek (ME_LayoutComponent).....	47
3.1.5. Értéktartomány (GG_DomainSet).....	51
3.1.6. Felvételek, adattömbök (GG_Recording, GG_MeasDataArray).....	52
3.2. Geofizikai modellek általános leírása	54
3.2.1. Modell (MO_Model).....	54
3.2.2. Fizikai tulajdonság (MO_Property)	55
3.2.3. Rétegmodell (MO_LayerModel).....	55
3.2.4. Rácsmodell (MO_GridModel)	56
3.2.5. Általános geometriájú modell (MO_GeneralModel)	57
3.3. Geofizikai inverziók általános leírása	58
3.3.1. Inverzió (MO_Inversion)	58
4. Az Általános Geofizikai Adatmodell (GGDM) implementációja.....	61
4.1. XML reprezentáció	61
4.1.1. A GGDM séma definíciós csomag.....	61
4.2. Az általános adatmodellre épülő geofizikai adatbázis (GGDB)	63
4.2.1. XML dokumentumtár	64
4.2.2. (Tisztán) Relációs adatbázis	71
4.2.3. Hibrid adatbázis.....	72
4.2.3.1. XML és a hagyományos relációs adatbázis	72

4.2.3.2. Adatrekordok törlése	77
4.2.3.3. Az informatikai alrendszerek áttekintése	79
4.2.3.4. Saját fejlesztésű adatbáziskezelő eszközök	80
4.2.4. Térinformatika	87
5. Az általános geofizikai adatmodell alkalmazásának lehetőségei	98
5.1. A GAIA rendszer átalakítása	98
5.2. WEB adatszolgáltatások	99
5.3. WEB inverziós szolgáltatások	99
5.4. Geofizika és az INSPIRE irányelvek	100
5.5. GGDM és a GeoSciML	101
Összefoglalás	102
Köszönetnyilvánítás	103
Irodalmi hivatkozások	104
Felhasznált szakirodalom	105
1. Függelék, UML: az egységesített modellező nyelv	107
2. Függelék, XML példák	110
Gravitációs mérési adatok	110
VESZ szondázási adatok	112
Rétegmodell, fúrási rétegsor	116
3. Függelék, XML	118
Mi az XML?	118
XML-re épülő technikák, eszközök	120
Az XML alkalmazása és fejlődési irányai	123

Abstract

Several national and industrial standards exist in geophysics. However, there is no general data model that can be used to describe all kinds of measurement, interpretation data and large geophysical datasets. Well known industrial standards cover the most important geophysical methods used in petroleum industry. Instead of uniform concepts they are based on traditions, and method specific principles. The first system that introduces real data modeling in geoscientific data exchange is the GeoSciML markup language. It uses fundamental geometric concepts to build up entities related to observations, measurements. In its' present state GeoSciML it is not ready for geophysical applications. In my theses I present two data models that are proved to be useful for such purposes. They are the GEOMIND Profile, and the General Geophysical Data Model (GGDM). The first one is the geophysical extension of the ISO 19115 metadata standard that supports the data storage and data transfer of two brand new web portals (GEOMIND: <http://www.geomind.eu>, KINGA: <http://kinga.elgi.hu>). The second one is a data model for measurement, processing and interpretation data that ignores the traditional method specific way of describing geophysical data. These models efficiently reduce the diversity of geophysical information and make it easier to create uniform databases and information systems. By introducing the concepts of „geophysical object”, „geophysical object set” and „report” large data systems (projects, campaigns) and documentation systems (report archives, libraries) can be represented by a set of standard metadata records. Using the XML based GGDM markup language complicated geophysical measurements are easily described by a uniform set of layouts composed by sources and sensors.

Two different implementations of the GGDM based geophysical database are also presented. The „XML depot” is a simple archive of XML records that is associated with a special index file generated by a search engine. It provides moderate flexibility and requires low development resources. The other implementation is the „Hybrid database”. Instead of extracting all XML elements to relational tables that would yield a huge, and probably too slow system, important attributes are kept traditional database fields, while independent deep structures are stored as XML. The result is a relatively simple structure, and a high level of flexibility. This solution requires higher development resources. Due to the open source Geoserver application both systems can easily support GIS functionalities.

The XML technology makes it possible to set up new geophysical web services providing more interoperability and a much higher level of integration between independent systems.

Rövidítések

A dolgozatban használt gyakori rövidítések jelentése:

ANSI	Amerikai szabványügyi szervezet (American National Standard Institute)
CygWin	Linux-szerű programkörnyezet Windows operációs rendszerhez
DIF	Data Interchange Format (ELGI fejlesztésű nyelv VESZ és TEM adatcseréhez)
ELGI	Eötvös Loránd Geofizikai Intézet
EPSG	European Petroleum Survey Group
GNU	A nyílt forráskódú Linux operációs rendszer fejlesztését és terjesztését segítő alapítvány. A név rekurzív betűszó (GNU's Not Unix)
GCC	C programnyelv fordító (GNU C Compiler)
ISO	International Standard Organization (nemzetközi szabványügyi szervezet)
GAIA	Geofizikai Adatok Integrált Adatbáziskezelő Rendszere
GEOMIND	Geophysical Multilingual Internet-Driven Information Service
GeoSciML	Geoscience Markup Language (földtudományi leírónyelv)
GIS	Geographic Information System (Térinformatikai rendszer)
GGDM	General Geophysical Data Model (Általános geofizikai adatmodell)
GML	Geography Markup Language (földrajzi leíró nyelv)
INSPIRE	Infrastructure for Spatial Information in Europe (Európai Téradat Infrastruktúra). térinformatikai infrastruktúra kialakítását célzó EU irányelvek összefoglaló neve
KINGA	Közcélú Internetes Geofizikai Adatszolgáltatás
Linux	Nyílt forráskódú, Unix alapú operációs rendszer
OGA	Országos Geoelektromos Adatbázis
OGC	Open Geospatial Consortium (Nyílt térinformatikai konzorcium)
RDBMS	Relational Database Management System (relációs adatbáziskezelő rendszer)
SQL	Sequential Query Language (szekvenciális kereső nyelv)
Tcl/Tk	Tool Command Language/ Toolkit (Nyílt forráskódú makrónyelv)
TEM	Tranziens Elektromágneses Szondázás
UML	Unified Modelling Language (egységesített modellező nyelv)
Unix	Operációs rendszer
VESZ	Vertikális Elektromos Szondázás

Windows	Operációs rendszer
WMS	Web Map Service (Internet alapú szabványos térképi szolgáltatás)
WFS	Web Feature Service (Internet alapú GIS adatbázis lekérdező szolgáltatás)
XML	Extendible Markup Language (kiterjeszhető leíró nyelv)
XPATH	(XML útvonal kereső nyelv)
XSD	XML Schema Definition (XML séma definíciós nyelv)

1. Előzmények

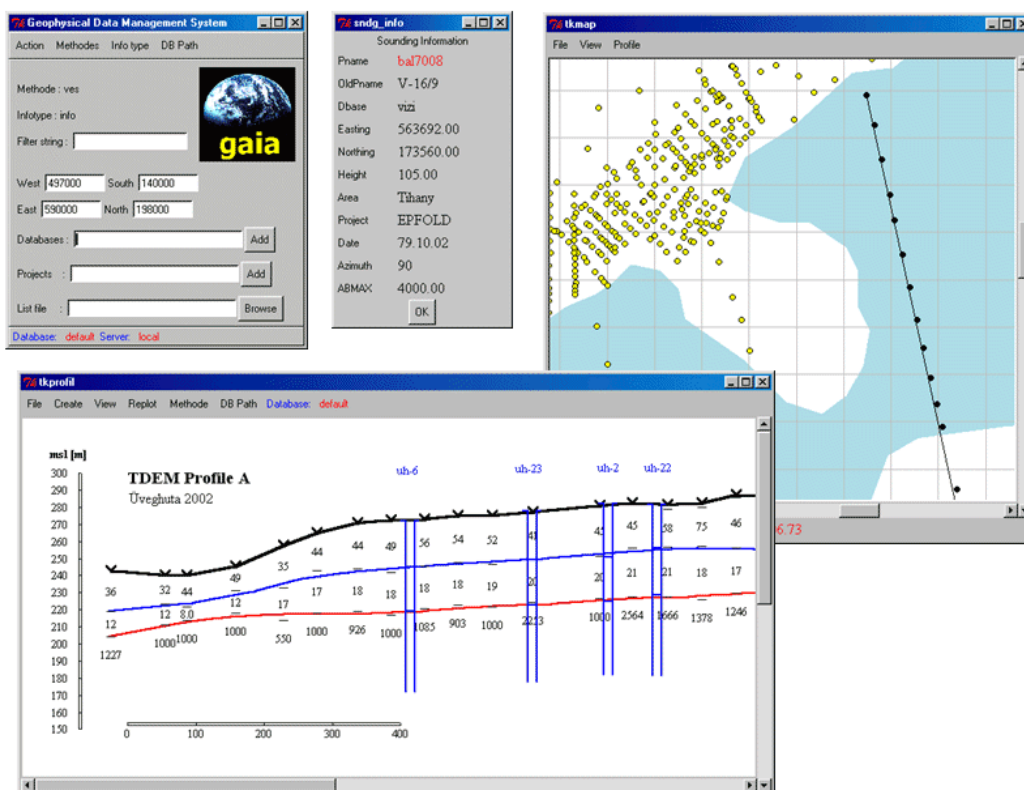
Az Eötvös Loránd Geofizikai Intézet Térképezési Főosztályán 1997-ben kezdtem el foglalkozni a geofizikai adatok egységes kezelését lehetővé tévő rendszerek építésével. Célom a geoelektromos kutatások napi adatfeldolgozási értelmezési, dokumentációs munkáinak megkönnyítése volt. Rugalmas, platform független eszközkészlet kialakítására törekedtem. A leendő rendszer architektúrájának tervezésekor vezérlő elvnek tekintettem azt a filozófiát, amely a komplikált rendszereket alapvető funkciókat ellátó modulok hálózataként képzei el. A modulok önállóan is használhatók, és ugyanakkor rugalmasan egymáshoz kapcsolhatók. Ez az alapvetően Unix-os szemlélet nagyfokú szabadságot engedélyez a professzionális felhasználó számára, egyben lehetőséget ad arra, hogy az átlagos felhasználó számára egyszerű kiszolgáló felületek mögé bonyolult feldolgozási láncokat építsünk ki.

Először VESZ (Vertikális Elektromos Szondázás) és TEM (tranziens elektromágneses) szondázások tárolása céljából készítettem egy adatbázist, amely saját gyártmányú adatbázis motorra épült. Ez a rendszer parancssorból hívható modulokból állt, amelyek strukturált felépítésű szöveges adatállományokból dolgoztak (DIF formátum). E programok képezték alapját az ELGI-ben 1998-ban bevezetett és folyamatosan fejlesztett GAIA (Geofizikai Adatok Integrált Adatbázisa) adatkezelő rendszernek (Sörös L., 2003). A kialakított eszközkészlet nemcsak a napi mérések adatait fogadta, hanem keretet adott a VESZ és TEM mérésekhez kapcsolódó adatfeldolgozó munkának is. A rendszer integrálta a Prácser Ernő által fejlesztett elektromágneses modellező és inverziós programcsomagot és így a geoelektromos mérések feldolgozásának elsősorú eszközévé vált (Prácser E. 1986, Prácser E. 1992). 1998 és 2008 között folyamatos archiválási tevékenység keretében felépült az Országos Geoelektromos Adatbázis (OGA). A csaknem 40 méternyi kéziratos dokumentáció rendezése, feldolgozása, digitalizálása útján előállt digitális adattár jelenleg több, mint 40000 VESZ és 5000 TEM szondázás adatait tárolja. Az országossá nőtt, többnyire archív adatokra épülő rendszer egységes feldolgozásával elkészült Magyarország 3D geoelektromos modellje, aminek horizontális szeleteit nemzetközi fórumokon, és az ELGI digitális térképtárában publikáltuk. (Sörös L., Prácser E, Gulyás Á, Kiss J., 2004)

1.1. A GAIA programrendszer

A történeti háttér felvázolása érdekében röviden összefoglalom a GAIA programrendszer legfontosabb tulajdonságait. A GAIA háromszintű architektúrával jellemezhető. Az első, legalsó szint az adatok fizikai tárolását, a relációs adatbázis kezelést és a lekérdezéseket végzi. Ez magába foglalja az RDBMS modult, a lekérdező és betöltő programokat. A második szintet a felhasználói funkciókat végző, illetve segítő parancskészlet jelenti. Ezek a programok primitív alkalmazások, amelyek a fentről jövő bonyolultabb kéréseket részekre bontják, és az alsó szintnek továbbítják, majd az eredményeket a felettük lévő moduloknak visszaküldik. A harmadik szint a grafikus felhasználói programok csoportja (**2. ábra**). Moduljai a második szint parancskészletének felhasználásához nyújtanak kényelmes, felhasználóbarát felületeket. (**1. ábra**)

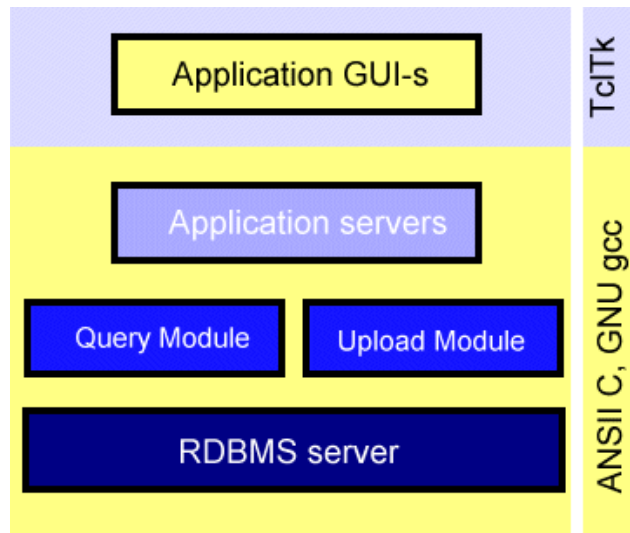
A GAIA-I fejlesztésénél elsődleges szempont volt a teljes platform függetlenség kialakítása és fenntartása. A rendszer alkotó elemei Unix/Linux és Windows operációs rendszereken lényegében egyformán működnek, és a fejlesztés szempontjából is egységesnek tekinthetők. Ez a rendszerkomponensek körültekintő összeválogatásának és főleg a GNU fejlesztőeszközök használatának köszönhető. Az alsó két szint megvalósításában ez az ANSI C programnyelv valamint a GNU GCC fordító használatát jelenti. A GNU fejlesztői környezet Windows alatti működtetését a CygWin rendszer teszi lehetővé. A grafikus fejlesztői környezet kialakításához a Tcl/Tk nyelvet választottam, amely egységes platform független működést és megjelenést biztosít. Az ily módon kialakított GAIA-I rugalmas, ugyanakkor kellően szilárd keretet biztosít az egyre nagyobb igénybevételnek kitett rendszernek.



1. ábra A GAIA kezelő felületei

Adatkezelő-, térképi-, és szelvénytérképező modul

A GAIA adatmodell hiányosságai a VESZ-GP és VESZ-TEM adatok együttes inverziójával folytatott kísérletek során váltak először nyilvánvalóvá. A hagyományos szemlélet, amely az inverziós adatokat a mérésekhez kapcsolja együttes inverziók esetén használhatatlannak bizonyult. Ekkor kezdtem el dolgozni a geofizikai mérési adatok általánosabb leírását lehetővé tevő adatmodellen. A nyílt forráskódú adatbáziskezelő rendszerek és az internetes technológiák fejlődése ráirányította a figyelmet a szabványos technológiák (SQL, XML) alkalmazásában rejlő előnyökre. A szabványosítási törekvéseknek további lökést adott, hogy 2006-ban az ELGI csatlakozott a GEOMIND Konzorciumhoz, amely 9 európai nemzet geológiai szolgálatainak és geofizikai intézményeinek összefogásával egy nemzetközi geofizikai metaadatbázis és internetes portál létrehozását tűzte ki célul. A GAIA rendszer jelenleg még változatlan formában működik, de intenzív fejlesztő munka folyik az adatbázis rendszert kiváltó aléptímeny kialakítására.

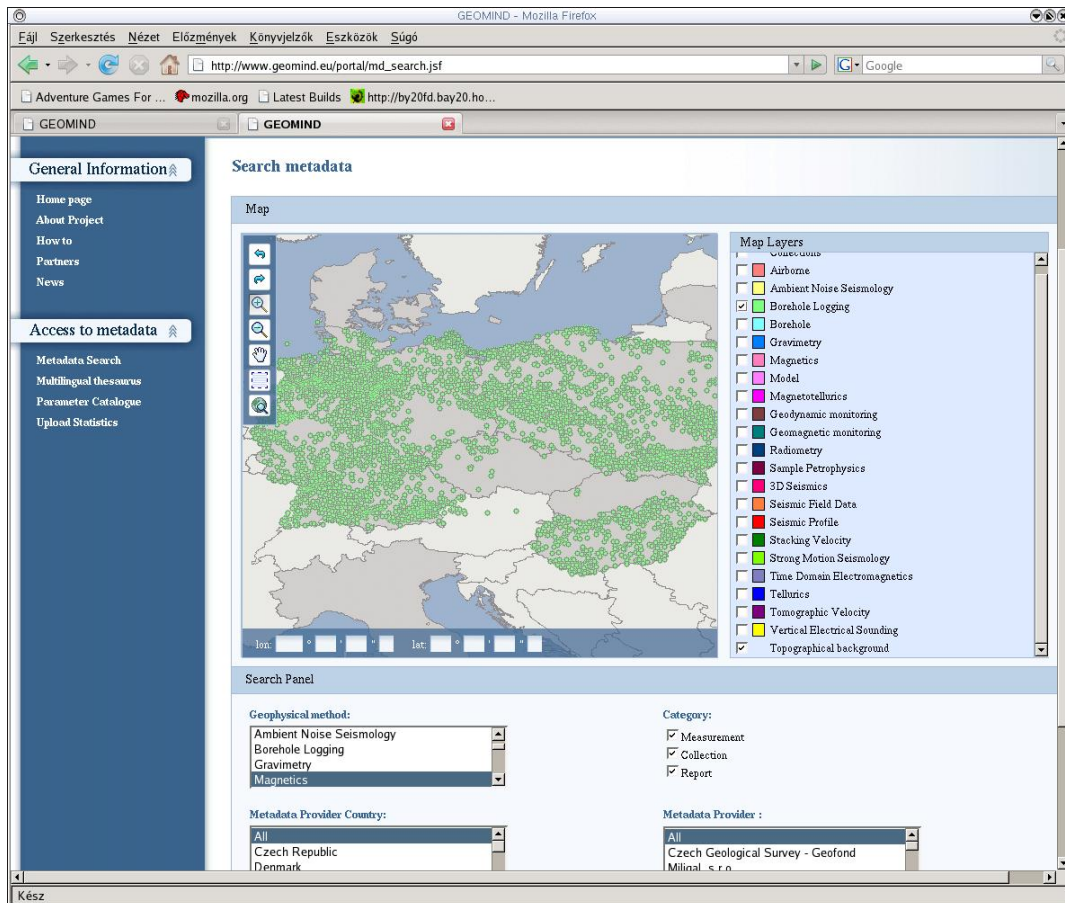


2. ábra GAIA program architektúra.

A rendszer három szintre bontható: A relációs adatbázis kezelőt, feltöltő és letöltő modulokat tartalmazó adatkezelő rendszer, alkalmazás szerverek és grafikus felhasználói felületek.

1.2. A GEOMIND projekt

A projekt az Európai Unió eContent $plus$ programjának támogatásával valósult meg 2006 és 2008 között. A munkában 9 uniós tagállam (Csehország, Görögország, Dánia, Lengyelország, Litvánia, Németország, Magyarország, Olaszország, Szlovákia) 10 különböző intézménye vett részt. Az ELGI két munkacsomag vezetésével fontos szakmai szerepet kapott. A digitális geofizikai adatokra vonatkozó szabványok specifikációját („Specifications of standards for digital geophysical content”) végző WP6-os munkacsomagot az ELGI képviselésében magam vezettem. Feladatomban az ISO 19115 metaadat szabvány adaptációjának előkészítése és geofizikai kiterjesztésének koordinációja volt. A fő hangsúly a GEOMIND Metaadat Profil adatmodelljének definiálásán volt, de az ELGI-ben szerzett tapasztalatok lehetővé tették, hogy megkísérleljem a geofizikai adatok általános leírását is. Megszületett a Geofizikai Adatok Általános Adatmodelljének első verziója (General Geophysical Data Model, GGDM) (Sörös L. et. al. 2007a-c). A projekt 2008 augusztusában zárult. A többnyelvű portál GIS térképi kereséseket és szabványos metaadat attribútumokra épülő kereséseket tesz lehetővé (3. ábra). Az adatbázisból több százezer geofizikai mérésről kaphatunk ingyenesen információkat. A portál a <http://www.geomind.eu> címen érhető el.

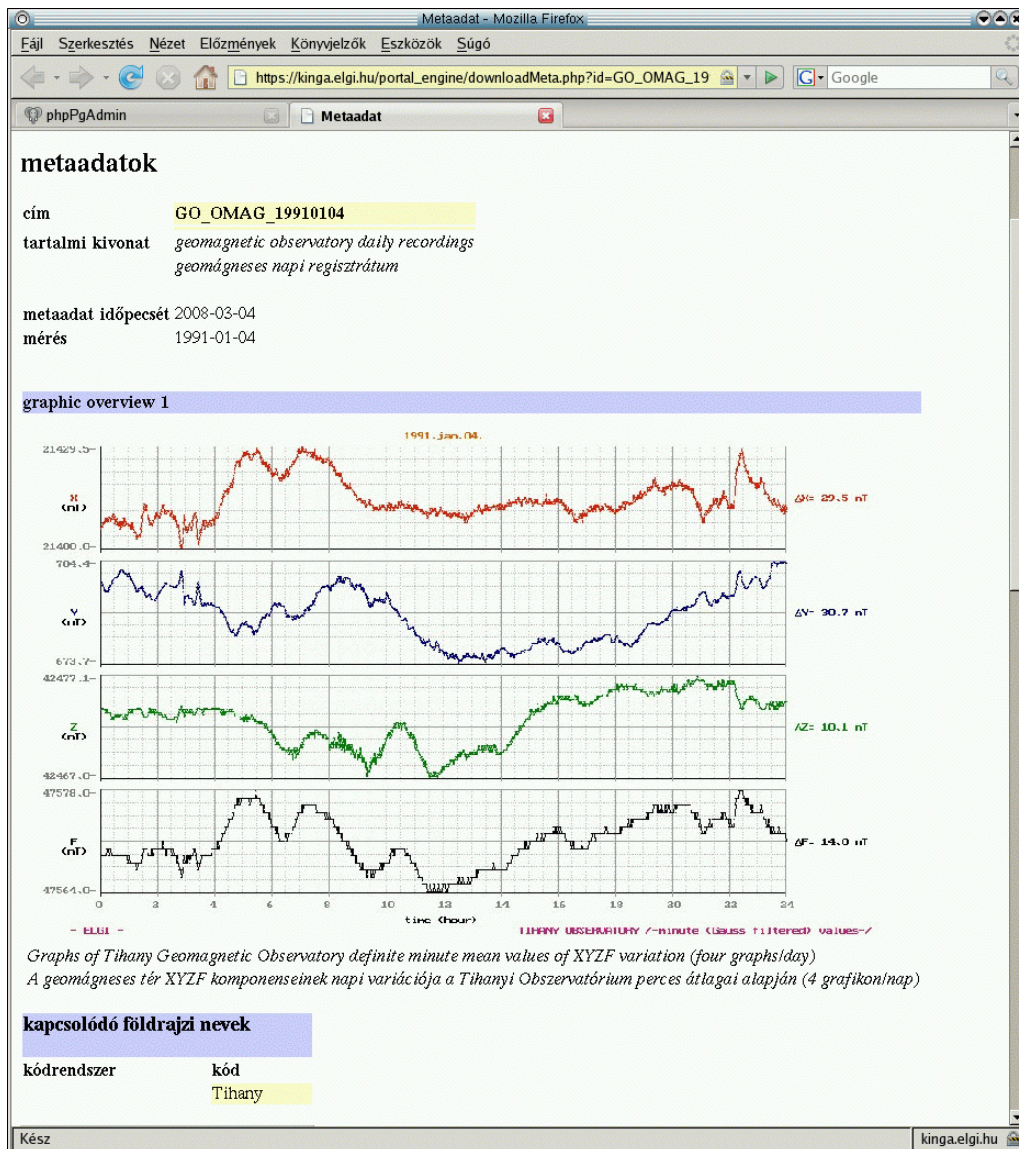


3. ábra A GEOMIND portál GIS kereső modulja lyukgeofizikai mérések területi elterjedését mutatja Közép-Európában. (nem teljes)

1.3. A KINGA projekt

A GEOMIND-dal csaknem párhuzamosan futott a GVOP támogatással megvalósított hazai projekt a KINGA (Közcélu INternetes Geofizikai Adatszolgáltatás, <http://kinga.elgi.hu>). A KINGA egyelőre kizárólag ELGI adatokra épülő információs rendszer, ami sok egyéb mellett a GEOMIND-hoz hasonló funkciókat valósít meg. A portál a felhasználó számára többféle keresési lehetőséget biztosít. A keresések eredménye itt is helyszínrajz, vagy a geofizikai mérésekre vonatkozó legfontosabb kísérő információkat tartalmazó lap, kivonatos képi információval (4. ábra). A technikai megoldásokban eltérő rendszer alapja adatszerkezet tekintetében a GEOMIND profilt implementáló XML sémacsomag egyszerűsített változata. A KINGA projektben feladatom a rendszer alapkonceptiójának kialakítása, az adatkonverziós programok elkészítése, a konverziós és adatfeltöltési munkák koordinálása volt. A projekt az

ELGI vezetésével, az InterComp Kft. részvételével zajlott. Technikailag a portál gerincét az InterComp Kft. által fejlesztett SmartPortál nevű alkalmazás szerver alkotja, erre épül rá a KINGA rendszert megvalósító program csomag.



4. ábra A tihanyi obszervatórium mágneses regisztrátumának metaadatai a KINGA portálon

Adat kontra metaadat

Mind a GEOMIND, mind pedig a KINGA alapvetően geofizikai metaadatokra épült. A metaadatok használata lehetővé teszi az adatrendszerek kielégítő nyilvántartását, a mérési, feldolgozási folyamatok dokumentálását. Fontos kapaszkodókat biztosítanak az adatkereső

mechanizmusok kiépítéséhez, a geofizikai információk átláthatóvá tételéhez. A metaadatok elsődleges fontosságúak az internetes információs rendszerekben, a közcélú adatok nyilvánosságának biztosításában. A geofizikus számára, aki a mérési adatokkal dolgozik, mindez szükséges rossz, és kellemetlen adminisztráció, mindaddig, amíg adatait, eredményeit nem akarja megosztani másokkal. Ha azonban az adatsere hasznos, vagy szükséges, a metaadatok használatának előnyei hamar nyilvánvalóvá válnak. Fontos, hogy a metaadatok rendszere a geofizika számára optimalizálva legyen, és elkészüljenek azok az eszközök, amelyek megkönnyítve az adatkezelők munkáját a metaadatok jó részét az alapadatokból automatikusan kigyűjtik, a szükséges kézi szerkesztéshez pedig kényelmes felületeket biztosítanak.

A mérési és feldolgozott adatok egységes rendszerbe foglalása a metaadat szerkezet helyes kialakításánál is nagyobb feladat. A GGDM Általános Geofizikai Adatmodell megalkotásával erre tettem kísérletet. Amint azt a későbbiekben bemutatom, a „geofizikai módszerek” hagyományos koncepciójának általánosabb értelmezésére és a mérési, feldolgozási adatok absztraktabb szemléletére van hozzá szükség. A nagytömegű adatrendszerek kezelésére vonatkozó elképzeléseim e két projekt során rengeteget formálódtak. Dolgozatom nagyrészt az elmúlt néhány év során kialakított GEOMIND Profilt, és a GGDM adatmodell azóta továbbfejlesztett változatát írja le.

2. Geofizikai adatok metaadat szabványa: a GEOMIND metaadat profil

2.1 A geofizikai adatrendszerek sokfélesége

A geofizikai adatrendszerek jellemző tulajdonsága a polimorfizmus. A mérési adatoktól az eredménytérképekig vagy 3D-4D animációkig igen széles a spektrum. Mivel a továbbiakban a különböző adattípusok egységes rendszerbe foglalásáról lesz szó, célszerű a teljesség igénye nélkül áttekinteni a geofizikai információk megjelenésének formáit és pontosítani néhány velük kapcsolatos fogalmat. Az alábbi osztályozás célja nem a geofizikai adattípusok egzakt

rendszerezése, hanem annak érzékeltetése, hogy egy átfogó igényű adatmodell csakis a geofizikában használt hagyományos fogalmak általánosításának sorozatán keresztül valósítható meg. Információ jellege alapján megkülönböztetünk mérési, feldolgozott és modell adatokat.

Mérési adatok: A mérési adatnak nevezünk minden olyan adatot, ami a műszeres észlelés közvetlen eredményeként jelenik meg.

Feldolgozott adatok: A mérési adatok felhasználásával automatikusan, vagy emberi beavatkozással nyert adatok. A mérési adat és a feldolgozott adat fogalma nem mindig választható szét, mivel a legtöbb mérési adat már bizonyos mértékű analóg vagy digitális feldolgozás eredménye. Közös bennük, hogy általában a vizsgált fizikai teret jellemző értékek, vagy azok egyszerű transzformáltjai.

Modell adatok: Az első két kategóriától élesen elváló fogalom. A modell adatok analóg, vagy numerikus modellezés, illetve inverzió útján keletkeznek. A vizsgált térrész geometriáját és valamely fizikai tulajdonságát jellemző értékek. Hagyományosan dimenzionalitásuk szerint (1D, 2D, 3D, 4D) osztályozhatók. Legjellemzőbb példák: a vízszintesen rétegzett homogén félteret leíró rétegmodell, a véges differenciás módszerekkel előállított rácsmo­dell, és a véges elemes modellezéssel kapott hálózatok.

A modell adatokkal együtt említendő az inverziós folyamatot jellemző technikai adatok, amelyek a számítási folyamatot és az eredmények megbízhatóságát jellemzik.

A mérési elrendezés, vagy térbeli megjelenítés alapján beszélhetünk pontszerű, szondázási, szelvény menti, területi és térfogati adatrendszerekről.

Pontszerű adatok: A mérési adat egyetlen térbeli ponthoz kapcsolható.

Szondázási adatok: A mérési adatok egy kitüntetett térbeli ponthoz köthetők, de az információ tartalom egy képzeletbeli (mélység, vagy idő) tengely mentén oszlik el.

Szelvény menti adatrendszer: Vonal mentén elosztott pontszerű adatok, vagy szondázások összessége.

Területi adatrendszer: Felszíni, vagy felszín alatti 2D ponthálózathoz kapcsolódó pontszerű adatok, szondázások összessége.

Térfogati adatrendszer: 3D ponthálózathoz kapcsolódó pontszerű adatok összessége.

Mindezeknek az időbeli kiterjesztésével idősorokat, vagy monitoring adatrendszereket kapunk. A két fő kategória elemeinek keverésével legtöbbször értelmes kombinációk állnak elő, és ezekkel az adatrendszerekkel a gyakorlatban találkozhatunk is.

A mérési és értelmezési adatok születésük helyének, idejének, módjának megfelelően kampányokba csoportosíthatók, a kampányok pedig projektek részét képezik.

Kampány: Munkaszervezési szempontból egységet alkotó, területileg és időben összetartozó mérések, értelmezési adatok, inverziós modellek és a hozzájuk kapcsolódó információk összessége.

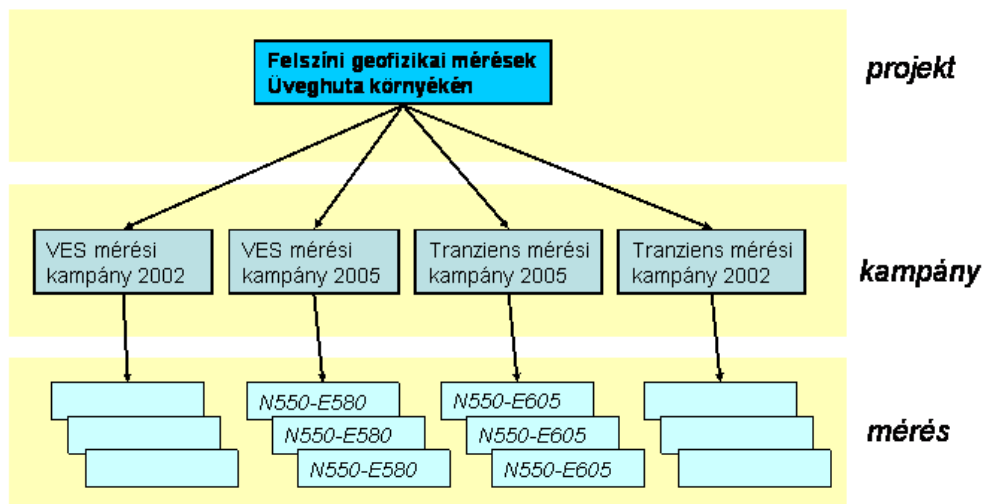
Projekt: Nagyobb területet, vagy hosszabb időtartamot átfogó mérési vagy értelmezési kampányok összessége.

Ha az alkalmazott geofizikai módszerek alapján további felosztást végzünk, az adattípusok áttekinthetetlen rendszeréhez jutunk. A geofizikai adatbázisok közötti átjárhatatlanság legfőbb oka e formagazdagság. A továbbiakban megmutatom, hogy a GEOMIND modell miképpen járul hozzá e kaotikus sokféleség rendezéséhez.

2.2. A geofizikai adatrendszerek hierarchiája

A geofizikai adatok rendszerezése számos módon lehetséges. Én a komplex geofizikai kutatások adatrendszereit hatékonyan kezelő hierarchikus leírást alakítottam ki. A nagytömegű adatrendszerek nyilvántartásához, informatikai rendszerekben való tárolásához célszerű az adatok fá struktúrába rendezése. Az átláthatóság fenntartása és a keresések meggyorsítása érdekében érdemes a geofizikai információs rendszereket az adatok keletkezésének körülményeihez igazítani, és azok természetes strukturáltságát követni. A gyakorlatban elég általánosan alkalmazható a mérések kampányokba és projektekbe szervezése, nagyjából annak megfelelően, ahogy a kutatási projektek ténylegesen felépülnek.

A geofizikai adatok hierarchiájának alján maguk a mérések találhatóak. Minden mérés valamilyen mérési kampány során keletkezik, és mivel ez a kapcsolat általában egyértelmű, célszerű minden mérést egyetlen szülőhöz, az azt létrehozó kampányhoz kapcsolni. Archiv adatoknál gyakran előfordul, hogy a mérési kampány már nem rekonstruálható. Ilyen esetben a szülő objektum az az archiválási kampány lehet, amely a régi méréseket egy adott információs rendszer számára elérhetővé teszi. A lényeg, hogy találjunk (definiáljunk) egy, és csakis egy objektumot, amely a mérési adat szülőjeként használható. Mi a helyzet a feldolgozás és értelmezés során keletkezett adatokkal? A mérési adatok korrekciójával, transzformációjával keletkezett adatok egyértelműen magukhoz a mérésekhez tartoznak, ezekkel igazából nincs probléma. A gond a mérési adatok felhasználásával keletkező modellekkel van. A modellek és mérések között nem feltétlenül egy-egy értelmű kapcsolat áll fenn (említettem már az együttes inverziók okozta problémát). Egy méréshez több inverziós megoldás is létezhet, és egy inverzió tetszőleges mennyiségű adatra épülhet. Célszerű tehát a modelleket a mérésekkel azonos hierarchia szinten álló önálló objektumokként kezelni. Ebben az esetben könnyű megtalálni a modellek szülőjét, ami nem más, mint a feldolgozási kampány, amelyben a modellek megszületnek. A mérési, archiválási és feldolgozási kampányok projektekbe szerveződnek, és a projekteket magasabb szintű, időben, vagy térben elkülönülő projektek fogják össze. Vegyünk egy konkrét példát: Az Üveghutai csökkentett aktivitású radioaktív hulladék tároló előkészítési munkái több éven keresztül zajlottak. A felszíni geofizikai munkák során számos geofizikai módszer bevetésre került. A kampányokban több ezer geofizikai mérés történt. Ezek mérési időszak és geofizikai módszer szerint egyértelműen elkülönültek. A mérési kampányokat egy csúcsojektum az üveghutai mérési projekt fogja össze. (5. *ábra*) Ez az egyszerű felosztás nagymértékben megkönnyíti a nagy mennyiségű és szerteágazó mérési anyag áttekintését.

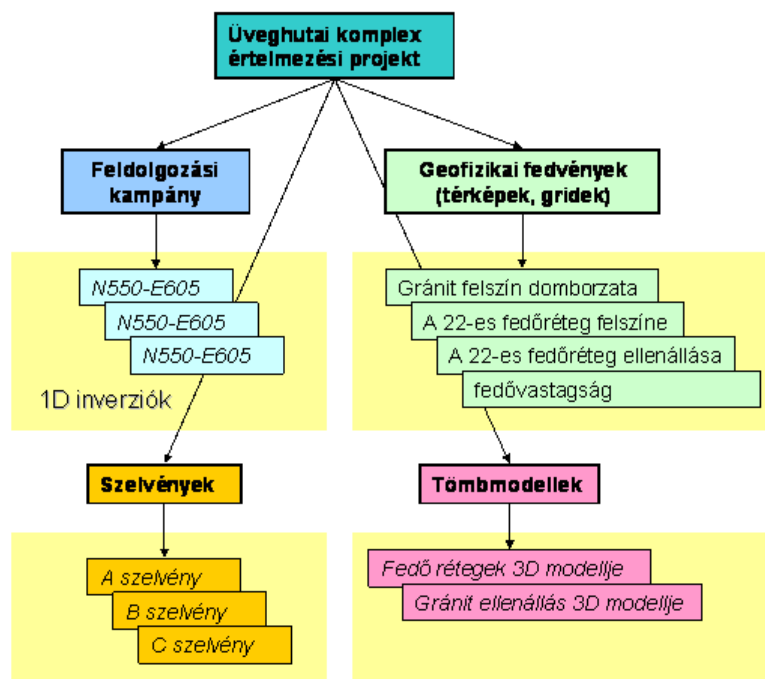


5. ábra Az üveghutai mérési projekt felosztása hierarchia szintekre. A mérés, kampány, projekt felosztás megfelel a mérési adatok és a kutatás természetes tagozódásának

A hierarchikus felosztásban eddig kétféle építőelemet használtunk: egyedi objektumokat és objektumcsoportokat. Ezek között szigorú szülő-gyermek kapcsolat áll fenn. Egy objektumcsoport lehet a szülője más objektumcsoportoknak is, ilyenkor aggregációról beszélünk.

Az adatrendszerek nem mindig sorolhatók be egyértelműen projektekbe illetve kampányokba. Amennyiben túl erőltetett, nem szükséges ezeket a kategóriákat használni. Általános célú objektumcsoportok is definiálhatók, amelyek szerepe nem más, mint objektumok halmazát szülőként összefogni. A geofizikai adatrendszerekben nem csak szülő-gyermek kapcsolatok léteznek. Pl. egy geofizikai szelvény objektumok gyűjteménye, de minden eleme tartozhat más szelvényhez is. Ezeknek az N-N kapcsolatoknak a leírásához használjuk a másodlagos objektumcsoportokat. A másodlagos jelző a szülő, vagyis elsődleges objektumcsoportoktól való eltérést hangsúlyozza. Minden objektum lehet több másodlagos csoportnak is tagja. A szelvények, rácsok (grid) és térképek szintén az objektumcsoport kategóriába sorolhatók. Első hallásra ez talán meglepő, de ha megpróbáljuk ábrázolni az egyedi mérések és térképek egyed – reláció viszonyát, akkor könnyen belátható, hogy minden térkép, (vagy rács) több objektummal áll kapcsolatban, még akkor is, ha kapcsolatuk rejtve marad egy adott informatikai alkalmazásban. Viszont nyilvánvaló, hogy egy térkép a vele kapcsolatban lévő objektumoknak (mérési, vagy modell adatoknak) nem szülője. Ezek tehát másodlagos csoportokként viselkednek. Egy másik példa segít megvilágítani a különböző objektumcsoport típusok szerepét. (6. ábra) A példaként bemutatott üveghutai komplex

értelmezési projekt magába foglal egy feldolgozási kampányt, és három általános objektumcsoportot: a szelvények, 2D és 3D rácsok csoportját. A feldolgozási kampány gyermek objektumai az 1D réteg modellek, A geofizikai szelvények objektumcsoportjának gyermekei az üveghutai szelvények, a fedvények objektumcsoportja tartalmazza a térképeket, grid-eket, a tömbmodellek csoportja pedig a fedő rétegek és a gránit test tömbmodelljét. A szelvények másodlagos csoportok, mivel tagjai nem kizárólagosan hozzá tartoznak. Minden rétegmodell szerepelhet több szelvényben is. A 2D, 3D grid-ek, térképek az objektumcsoportok egy különleges típusát képviselik, amelynek neve geofizikai fedvény. A fedvény térinformatikai fogalom, pontos definícióját az ISO 19123 szabvány tartalmazza.



6. ábra Feldolgozott geofizikai adatok hierarchiája. A projekt egy feldolgozási kampány és három elsődleges objektumcsoport szülője. Ezek további (másodlagos) objektumcsoportokat tartalmaznak.

Ennek értelmében olyan függvényként értelmezhető, amely téridőbeli kiterjedésének minden pontjához egy, vagy több függvényértéket rendel. Lehet diszkrét, folytonos, és a hordozó geometriától függően pont-, vonal-, felület-, vagy térbeli fedvény. A geofizikai paramétertérképek jellemzően folytonos felületi fedvények. A geofizikai fedvények másodlagos csoportokként viselkednek. Minden egyes rétegmodellről (1D inverzió) elvileg

tudni lehet, hogy szerepel-e az adott szelvény, térkép, grid elkészítésénél, azaz, tagja-e az azokat leíró másodlagos csoportnak. (Ha máshol nem, a grid-ek előállításához használt adatfájlokban ezek az információk dokumentálva vannak.)

A geofizikai objektumok és objektumcsoportok közös jellemzője, hogy elválaszthatatlanul kapcsolódnak a hagyományos értelemben vett geofizikai adatokhoz, a műszerekhez, feldolgozási módszerekhez, geofizikai paraméterekhez. Van a geofizikai információknak egy nagy csoportja, amiről eddig nem esett szó: ez a dokumentáció. A jelentések, szakcikkek, ábrák, nyomtatott, vagy digitális szövegek az információ sajátos formáját képviselik, és határozottan elkülönülnek az eddig tárgyalt adattípusoktól. Ezek nem sorolhatók az objektumok és objektumcsoportok közé, mivel a geofizikai adatokhoz fűződő viszonyuk jóval áttételesebb. A geofizikai adatoktól függetlenül, önmagukban is megállnak. Persze, ha kell, a jelentések és mérések összekapcsolásának nincs semmi akadálya, sőt adott esetben nagyon hasznos lehet. Eddig három alapvető kategóriát sikerült elkülönítenünk. Nevezzük ezeket a továbbiakban – alkalmazkodva a később bemutatandó szabványhoz – hierarchia szinteknek.

hierarchia szint	típus	rang
geofizikai objektum	VESZ	-
	TDEM	-
	MT	-

geofizikai objektumcsoport	campaign (kampány)	elsődleges
	project (projekt)	elsődleges
	objectSet (objektumcsoport)	elsődleges
	geophCoverage (geofizikai fedvény)	másodlagos
	objectGroup (másodlagos csoport)	másodlagos
riport	map (térkép)	-
	profile (szelvény)	-
	sounding (szondázás)	-
	boreholLog (karotázs szelvény)	-
	text (szöveg)	-
riportkomponens	ua. mint a riport	-

1. táblázat Geofizikai adatok hierarchiaszintjei és típusai

A három fő hierarchia szint tehát: geofizikai objektum, geofizikai objektumcsoport és dokumentáció, vagy a GEOMIND terminológia szerint riport. Az egyes hierarchia szintek típusait és rangját az **1. táblázat** tartalmazza. Amint látható a geofizikai objektumok típusai átfedésben vannak a hagyományos „geofizikai módszer” fogalmával. A felsorolás persze nem teljes. A riportkomponens, mint önálló hierarchia szint mutatja, hogy a dokumentáció is hierarchikus rendszerbe szervezhető. Az itt közölt felosztás, bár nyilvánvalónak tűnik, egyáltalán nem triviális felismerés, hanem egy hosszú, kristályosodási folyamat eredménye, melynek célja egy rendkívül heterogén adatkört lefedő adatmodell felállítása volt.

A kérdés az, hogy miként képezhetők le ezek a kapcsolatok egy geofizikai információs rendszerben? A hagyományos értelemben vett mérési adatokhoz ezeknek az információknak vajmi kevés köze van, itt tipikus metaadatokkal állunk szemben. A metaadatok legfrappánsabb definíciója: „adat az adatról”. Ez a legtágabb értelemben igaz. Minden információ, ami az adatokkal kapcsolatban tartalmi, formai, technikai vonatkozásában érdekes – tehát nem csak az elérhetőség, vagy a keletkezés helye, ideje – metaadat. Az OGC (Open Geospatial Consortium) által elsődlegesen támogatott szabvány a földrajzi adatokra vonatkozó ISO 19115 a metaadatokat egy rendkívül szerteágazó, és sokoldalúan használható struktúrában definiálja. Mivel földrajzi adatok leírására készült, jól alkalmazható a térinformatikai adatrendszerek jellemzésére. A geofizikai adatrendszerek bizonyos értelemben földrajzi adatoknak tekinthetők, bár, ahogy minden geofizikus tudja, ennél sokkal többről van szó. Ez szükségessé teszi az alapszabvány kiterjesztését, vagyis az adott tematikus csoportra (geofizikai adatrendszerek) való optimalizálást. A továbbiakban bemutatom, az ISO 19115 szabványt és a GEOMIND metaadat profilt, ami a tapasztalatok szerint igen jól használható a geofizikai adatrendszerek széleskörű leírására.

2.3. A földrajzi adatokra vonatkozó ISO 19115 metaadat szabvány ismertetése

A geofizikában nem szokás eltúlozni a metaadatok használatát. Általában az ún. header adatok megadásában kimerül a geofizikus erre irányuló tevékenysége. Bár a nagy ipari szabványok külön helyet biztosítanak a metaadatok (fejléc) számára, a gyakorlatban ezek az adatterületek sokszor üresen maradnak. A metaadatok használatának fontossága és elterjedtsége az analóg és digitális könyvtári rendszerekben közismert. A rokon szakmákban

először a térinformatikai rendszerekkel dolgozók kényszerültek rá a metaadatok használatára. A piaci GIS alkalmazások számos nemzetközi metaadat szabványt támogatnak. Az internetes kereső szolgáltatások fejlődésével a metaadatok jelentősége hamar nőni kezdett. Mára a világhálót feltérképező robotok és az élet minden területére kiterjedő web szolgáltatások rohamos terjedésével a metaadatok szerepe az információk elérésében kulcsfontosságúvá vált. Az Európai Unió térinformatikai közműhálózatának kiépítését célzó irányelvek (INSPIRE) hamarosan törvényben szabályozzák, és kötelezővé teszik a közcélú adatokat szolgáltató intézmények számára a metaadatok szabványos használatát. Bár a geofizika külön megemlítve nem szerepel az INSPIRE által megadott téradat témák között, a 2. számú melléklet B pontjában felsorolt földtannak tágabb értelemben véve nyilvánvalóan része. A nemzetközi metaadat szabványok közül az OGC által is támogatott ISO 19115 a legkiforrottabb, és legígéretesebb.

A szabvány teljes körű megismerése önmagában sem kis feladat, de itt a cél egy olyan szakterületen történő alkalmazás volt, ahol ez korábban még nem történt meg. A kiterjesztés módját, a lehetőségeket maga a szabvány szigorúan korlátozza. A kiterjesztés első számú szabálya: a meglévő elemek használata mindig preferált. Új metaadat elemek bevezetése, csak akkor indokolt, ha az adott információ a már meglévő elemek használatával nem írható le. Első feladat tehát a szabvány értelmezése, vagyis a „törvény szellemének” megismerése volt. Ennek a megismerési folyamatnak a végeredményeként kialakult egy viszonyrendszer, ami az ISO 19115 által definiált általános fogalmakat a geofizikai adatrendszerekre történő alkalmazás során használt konkrét adatokra képezi le. A szabvány alábbi ismertetésében ezekre a kapcsolatokra külön kitérek. Az opcionális elemek használata általában hierarchiaszint függő. Alapelv, hogy minden információt a lehető legmagasabb, de még releváns szinten kell megadni. Egy objektumcsoportra definiált metaadat elem minden gyermekre is érvényes, amennyiben alacsonyabb szinten ezt nem írjuk felül.

Célszerű a metaadatokra úgy gondolni, mint egy információcsomagra, amely a ki, mit, mikor, hol, miért, hogyan, kitől, mennyiért kérdésekre adott válaszokat tartalmazza. Amennyiben az egységesen feltett kérdésekre mindenki egységes formában (és igazat) válaszol, úgy bármilyen káosz is van az adott területen, legalább egységes áttekintést nyerhetünk és lerövidítjük az adatszolgáltató és a felhasználó közötti utat. A metaadatok használatának a dokumentálás mellett legfontosabb célja a forrásadatok keresésének megkönnyítése. Szabványos adatmezőkkel, kontrollált adattartalom mellett a modern adatbázis rendszerek

rendkívül hatékonyan képesek megtalálni több millió rekord közül is azt, amire szükségünk van. A földrajzi adatok tárolásával és GIS rendszerek használatával térbeli kereséseket is végezhetünk.

Az ISO19115-öt 2003-ban bocsátotta ki az ISO szervezet TC 211-es technikai társasága. Az eredeti dokumentum UML (Unified Modeling Language) diagramok és un. adat szótárak (data dictionary) formájában írja le a metaadat modell logikai szerkezetét. Az eredeti dokumentum a metaadat implementációjára vonatkozóan nem tartalmaz megkötéseket. Az ISO19115-öt 2005-ben követte az ISO19139, ami az eredeti szabvány egzakt technikai specifikációja XSD séma definíciós (XML Schema Definition) csomag formájában. Ez meghatározza metaadok XML-ben (eXtensible Markup Language) történő leírásának módját és lehetővé teszi a fájlok automatikus validációját, azaz, a szabványnak való megfelelés programok által történő szigorú ellenőrzését. A csomag az általános metaadat elemeken kívül számos beépülő, főleg GIS-hez kapcsolódó szabvány (ISO19107, -108, -109, -110, -111, -112, -116, 117, -119) és a GML (Geography Markup Language, ISO19136) definícióját is integrálja.

A földrajzi metaadatok szabványa nagyon általános, és meglehetősen bonyolult. Több száz elemet és mély, beágyazott struktúrákat tartalmaz. Tartalmilag 12 témakört ölel fel, amelyeket a **2. táblázat** tartalmaz.

Témakör angol neve	Témakör magyar neve
Metadata entity set	Metaadat entitás halmaz
Identification	Forrásadatok azonosítása
Constraints	Korlátozások
Data quality	Adatminőség
Maintenance	Karbantartás
Spatial representation	Térbeli leírás
Reference system	Koordináta rendszer
Content information	Adattartalom leírása
Portrayal catalog	Portré katalógus
Distribution	Terjesztés
Metadata extension	Metaadat kiterjesztés
Application schema	Alkalmazás séma

2. táblázat Témakörök az ISO19115 metaadat szabványban

Szerencsére a szabvánnyal való kompatibilitáshoz nincs szükség mindig mindenre. A metaadat elemek megadása lehet kötelező, opcionális (nem kötelező) és feltételekhez kötött. A szabvány definiál egy „kemény magot” (core elements), amely a szükséges minimumot tartalmazza, de még ebben is található opcionális elemek. Az alábbiakban sorra vesszem a szabvány magjának legfontosabb részeit. Mivel hivatalos magyar fordítás még nincs, az angol nyelvű megnevezéseket is használom.

2.3.1. Metaadat entitás halmaz (Metadata entity set)

Keretobjektumként magába foglalja az összes többi elemet, és a metaadat rekordok legfontosabb jellemzőiről is ad tájékoztatást. Ilyenek a fájlazonosító kód, nyelv, karakterkészlet, időpecsét stb. Itt kapott helyet a metaadatokat szolgáltató felelős intézmény/személy elérhetősége és a forrásadatokat által használt koordináta rendszerek leírása. Az utóbbinál lehetőség van a vetületi rendszerek teljes definiálására, de elegendő a hivatalos EPSG kód (European Petroleum Survey Group) megadása is.

Geofizikai alkalmazás:

A metaadatok azonosítására a fájlazonosító kód (fileIdentifier) szolgál. Az adatszolgáltató nevével együtt univerzális egyedi azonosítóként szolgál. Célszerűen a hivatkozott forrásadat azonosító kódjából képzett karaktersorozat, amely lehetőleg nem tartalmaz fájlnevekben tiltott, és ékezetes karaktereket. A metaadatok alapnyelve a GEOMIND szabványban mindig angol, de az ISO 19139 implementáció lehetővé teszi a többnyelvű szövegek használatát. A karakterkészlet mindig UTF-8, így magyar, görög, cirill, vagy bármilyen karakterkészlet is használható.

Jellemző hierarchia szint:

geofizikai objektum, geofizikai objektumcsoport, riport

2.3.2. Forrásadatok azonosítása (Identification)

Hivatkozás (citation)

A forrásadatok hivatkozására egy komplett struktúra (citation) áll rendelkezésre. Ez tartalmazza a forrás címét, alcímét, a dokumentummal kapcsolatos felelős intézmények/személyek felsorolását (szerző, tulajdonos, karbantartó stb), azonosítókat, dátumokat (létrehozás, publikáció, revízió dátuma), könyvtári kiadványok esetében a ISBN és ISSN kódokat.

Tartalmi összefoglaló (abstract)

Lehetőség van tartalmi összefoglalóban (abstract) megadni a forrással kapcsolatos legfontosabb közlendőket is.

Grafikus átnézeti kép (graphicOverview)

Mellékelhető egy grafikus átnézeti kép (graphicOverview) amely lehet képfájl, vagy beillesztett GML objektum.

Kulcsszavak (keywords)

Fontos szerepe van a kulcsszavaknak, amelyeket thesaurusokból kell kiválasztani, és a kulcsszavak mellett szerepeltetni kell a témakört és a szótárra való hivatkozást is. Ez biztosítja a kulcsszavak alapján történő keresések egyértelműségét.

Kiterjedés (extent)

Kötelező elem a forrás horizontális kiterjedését megadó struktúra (extent). Használhatunk földrajzi megnevezéseket, vagy kódokat, (település nevek, EOV térképlap szám) illetve megadhatjuk az adatrendszer körvonalát egy polygon formájában. Kötelező szerepeltetni az adatrendszert magába foglaló koordináta- négyszöget (geographic bounding box). Adatrendszerünk időbeli és vertikális kiterjedését is leírhatjuk, ha ennek van értelme. (pl. monitoring adatok, vagy mélységhez kapcsolható fizikai paraméterek esetén)

Geofizikai alkalmazás:

A forrás hivatkozások kötelező minimumát egy cím és legalább egy dátum jelenti. A dátumok megadásánál (citation/date) célszerű legalább a mérés, feldolgozás idejét megadni.

Geofizikai objektumok esetén a hivatkozási cím (citation/title) azonos a fájlazonosító kóddal. Objektumcsoportok, geofizikai fedvények, vagy jelentések esetén ez az adott forrás valódi címének felel meg. Igen hasznos az egyéb azonosítók (citation/identifier) megadása. Amennyiben egy mérésnek több nyilvántartási rendszerben több azonosítója van, ezeket a névtér megjelölésével dokumentálhatjuk.

Az absztraktban (citation/abstract) lehetőség van több nyelven leírni a forrással kapcsolatos legfontosabb tudnivalókat. Ezek a mezők jellemzően a szöveges keresések céltárgyai. A grafikus átnézeti kép megadása opcionális. Általában kis felbontású grafikát jelent, ami a terjesztő érdekeit még nem sérti.

Időbeli kiterjedésen általában a mérés időtartamát értjük, míg a vertikális kiterjedés az adatok behatolási mélységére vonatkozik. Jelentések esetén a térbeli és időbeli kiterjedés a kutatási területtel és időtartamával azonos.

Jellemző hierarchia szint:

geofizikai objektum, geofizikai objektumcsoport, riport

2.3.3. Korlátozások (Constraints)

Az adatok hozzáféréseivel és felhasználásával kapcsolatos korlátozások megadására szolgáló szerkezet. Lehetőség van a korlátozások szabad szöveges megfogalmazására, és a forrás adatokra vonatkozó megszorítások kódszavas jellemzésére. (copyright, patent, licence...)

Geofizikai alkalmazás:

Az opcionális elem elhagyása azt jelenti, hogy az adatok a terjesztő által megadott általános szabályokon túl korlátozások nélkül elérhetők és használhatók. Nem nyilvános adatok esetén a korlátozás kódját kell használni.

Jellemző hierarchia szint:

Általában a magasabb szintű geofizikai objektumcsoport, riport

2.3.4. Térbeli leírás (Spatial Representation)

Meg lehet adni a térbeli reprezentáció típusát (vektor, grid) és a térbeli felbontóképességet.

A térbeli leírás vektoros vagy grid adatrendszerek jellemzésére szolgál. Az első esetben a forrás adatok térképi ábrázolásakor használt elemek típusát és számát, második esetben a gridek méretét, felbontását orientációját adhatjuk meg.

Geofizikai alkalmazás:

A térbeli reprezentáció egy adott GIS implementációban a forrás adatok térképi megjelenésére vonatkozik. Pl. egy VESZ mérés vektor reprezentációjú, és ponttal ábrázoljuk (a szondázás vonatkoztatási pontja), egy szeizmikus szelvény szintén vektoros, és egy vonal (lineString) jeleníti meg. Magyarország Bouguer-anomália térképe grid reprezentációjú. Megadható a rács és a cellák mérete is.

A térbeli felbontóképesség a helyszínrajz felbontóképességét jelenti, amit a méretarányt kifejező integer értéként adunk meg. Pl., ha a kézi digitalizálás hibája 10 m körüli, akkor a felbontás 10000. (a kézi digitalizálás hibája kb. 1 mm)

Jellemző hierarchia szint:

geofizikai objektum, geofizikai objektumcsoport

2.3.5. Vonatkoztatási rendszer (Reference System)

A forrásadatokat leíró koordinátarendszer megadására szolgál. Szabványos koordinátarendszerek esetén elegendő az EPSG kódot megadni, egyébként a térképi vetület egzakt meghatározására szolgáló bonyolult szerkezet használandó.

Geofizikai alkalmazás:

Az EPSG kód alkalmazása kötelező. A WGS-84 koordináta rendszer EPSG kódja EPSG:4326, az EOVS rendszeré EPSG:23700

Jellemző hierarchia szint:

Minden hierarchia szinten kötelező

2.3.6. Terjesztés (Distribution)

A terjesztéssel kapcsolatos adatok rendkívül fontosak, hiszen a metaadatok egyik fő célja épp a terjesztés megkönnyítése. Egy forrás objektumhoz tetszőleges számú terjesztő tartozhat, és mindnek megvannak a saját terjesztői opciói. Ezek tartalmazzák az elérhetőségeket, a média típusát, formátumát, az online hozzáférés módját, valamint az árat, megrendeléssel kapcsolatos instrukciókat, a hozzáférhetőség várható idejét stb. A GEOMIND profil a terjesztési adatok leírására egyszerűsített szerkezetet alkalmaz.

Jellemző hierarchia szint:

Általában a magasabb szintű geofizikai objektumcsoport, riport

2.3.7. Adatminőség (Data quality)

Az adatok minőségének és a feldolgozási lépéseknek a leírására két szerkezet, a report (jelentés) és a lineage (feldolgozási lépések) áll rendelkezésre. A report főleg geodéziai adatok minőségi kontrolljához használható, előre definiált minőségi kategóriákkal dolgozik. A lineage inkább a feldolgozási lépések felsorolására alkalmas. A GEOMIND profilba csak a lineage elem került bele, ami egy minőségre vonatkozó kijelentést, és az ehhez tartozó adatfeldolgozási lépések felsorolását tartalmazza a felelős intézmények/személyek, az időpontok, és a feldolgozási lépés céljának megjelölésével. (Ki, mikor, miért, mit csinált? Mi ennek az eredménye?)

Geofizikai alkalmazás:

A lineage lehetővé teszi a teljes mérési és feldolgozási folyamat dokumentálását, pl. egy gravitációs térkép elkészítésének, vagy egy szeizmikus szelvény feldolgozásának lépéseit.

Jellemző hierarchia szint:

geofizikai objektumcsoport, ahol az információ még minden gyermekre igaz.

2.3.8. Metaadat kiterjesztés (Metadata Extention)

Az alapszabványtól eltérő elemeket dokumentálni kell. Ennek módját a szabvány egy külön erre szolgáló elem használatával szabályozza. Meg kell adni az új elem nevét, rövidítését, definícióját, számosságát, adattípusát, ha komplex szerkezetéről van szó, akkor az osztály típusát, és a szülő elem nevét. Megadható a kiterjesztés indoka, és a felelős intézmény, vagy személy is.

Geofizikai alkalmazás:

A GEOMIND profil leírása az ISO-19139 metaadat implementációnak megfelelő formában az Interneten elérhető. (GEOMIND Consortium, 2007a) Az XML dokumentum minden alapszabványtól eltérő részletet a szabvány által megadott formában közöl.

2.4. Az ISO 19115 geofizikai kiterjesztése

Miért van szükség a szabvány kiterjesztésére?

Geofizikai paramétertérképeink leírására elvileg minden változtatás nélkül használhatnánk az alapszabványt. Ebben a térinformatikában jól bevált gyakorlat szerint járhatnánk el, a más szakterületen dolgozó térképkészítő kollégákhoz hasonlóan. De mi a helyzet a többi geofizikai adatformával? A bevezetőben felvázolt sokszínűséggel a hátunk mögött nehezen megválaszolható kérdésekbe ütközünk. Rögtön az első: hogyan értelmezzük a szabványban definiált hierarchia szinteket? A metaadatok érvényességi körére vonatkozó kódlista (MD_ScopeCode) az attribútumtól a szolgáltatásokig széles skálán mozog, de nehéz megmondani, melyik felel meg a geofizikai mérés, inverziós modell, vagy szeizmikus szelvény fogalmának. A **3. táblázat** értékei közül leginkább azt tudjuk megmondani, hogy melyik biztosan nem. Közelebb visz a megoldáshoz a következő gondolatmenet: Az alapállás szerint egy térinformatikai rendszerbe kell illesztenünk a geofizikai adatainkat. Képzeljük el, hogy a geofizikai mérés, inverziós modell, vagy szeizmikus szelvény ebben hogyan jelenne meg!

MD_ScopeCode
attribute
attributeType
collectionHardware
collectionSession
dataset
series
nonGeographicDataset
dimensionGroup
feature
featureType
propertyType
fieldSession
software
service
model
tile

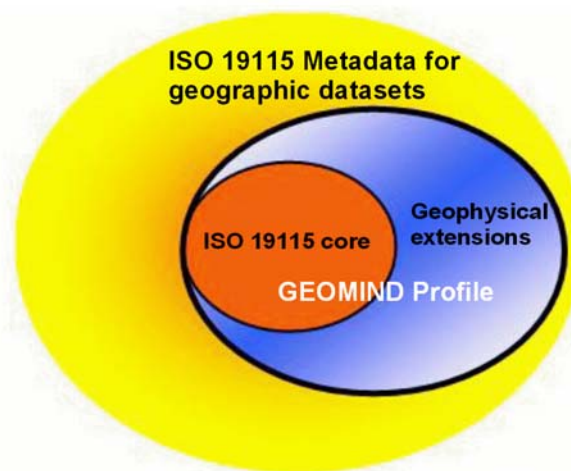
3. táblázat Az ISO 19115 metaadatok lehetséges hierarchia szintjei

A listában van néhány klasszikus GIS elem, mint pl. attribute (tulajdonság), feature (alakzat), dataset (adatszett). Térinformatikai rendszerünkben a geofizikai mérések helyhez kötött geometriai elemekként (pontok, vonalak, vagy polygonok) láthatók a digitális térképen. Minden egyes elem kiválasztható, és a hozzá kapcsolódó adatbázis tartalom lekérdezhető. Ebben az elképzelt példában a térkép, ami a mérések adott halmazát mutatja, nem más, mint maga a „dataset”. Az ISO 19115 metaadatok hierarchia szintje egyébként alapértelmezésben ez. Minden mérés „feature”, és a lekérdezhető adatbázis tartalom attribútumok halmaza. Ez egy jól áttekinthető helyzetnek tűnik. De mi van akkor, ha a méréseinket részletesebben akarjuk láttatni? Pl. a szeizmikus szelvény minden egyes geofonját ábrázoljuk, és minden geofonhoz lekérdezhetjük az adott szeizmikus csatorna adatait. Ebben az esetben van egy térképi rétegünk, egy „dataset series”, ami több szeizmikus vonalat, „dataset”-et ábrázol. A „feature” a geofon, az attribútum pedig a csatorna specifikus adat. Vegyünk egy harmadik példát, ahol a térkép geofizikai projektek körvonalait ábrázolja. Az egyes polygonokat kiválasztva információkat kaphatunk az adott projekt idejéről, témájáról, felelőseiről. Jelen esetben a projekt felel meg a „feature” hierarchia szintnek, szemben az előző példával, ahol „dataset series” volt. Hol az igazság? Nincs igazság, helyesebben mindig van egy optimális megoldás, amit az adott implementáció célja határoz meg. A későbbiekben látni fogjuk, hogy a hierarchia szintek kódlistájának kibővítésével használható kompromisszumhoz jutunk. A már megismert objektum, objektumcsoport, riport kategóriákról van szó. A geofizikai

adatformákat hierarchia szintekhez köthetjük, és egyben rögzítjük őket a „dataset”-hez képest. Ebben a rendszerben a geofizikai objektum lesz a „feature”, az objektumcsoport a „dataset” az objektumcsoportok magasabb szintű aggregációja pedig a „dataset series”. Az angol terminológia erősebben sugallja a párhuzamot. (geophObjectSet – dataset). A GEOMIND profil alapvető célja a geofizikai mérések elérhetőségének javítása. A mérés, vagyis a geofizikai objektum a rendszer alapköve, ezért indokolt a térinformatikai alapelemhez, a „feature”-höz horgonyozni. Ennyi bevezető után lássuk, milyen kiterjesztéseket tartalmaz a GEOMIND profil.

A kiterjesztés módja

Az ISO19115 meghatározza a szabvány kiterjesztésének a módját, lehetővé téve, hogy speciális közösségek a saját szakmai igényükhöz jobban alkalmazkodó metaadat leírásokat használhassanak. A GEOMIND profil a geofizikus társadalom igényeihez próbál alkalmazkodni. Az egyszerűsége törekedve a kötelező minimumból, a szabvány magjából indul ki. (7. ábra) Az igen kiterjedt szabvány számos olyan elemet foglal magába, amelyek a geofizika szempontjából kevésbé érdekesek.



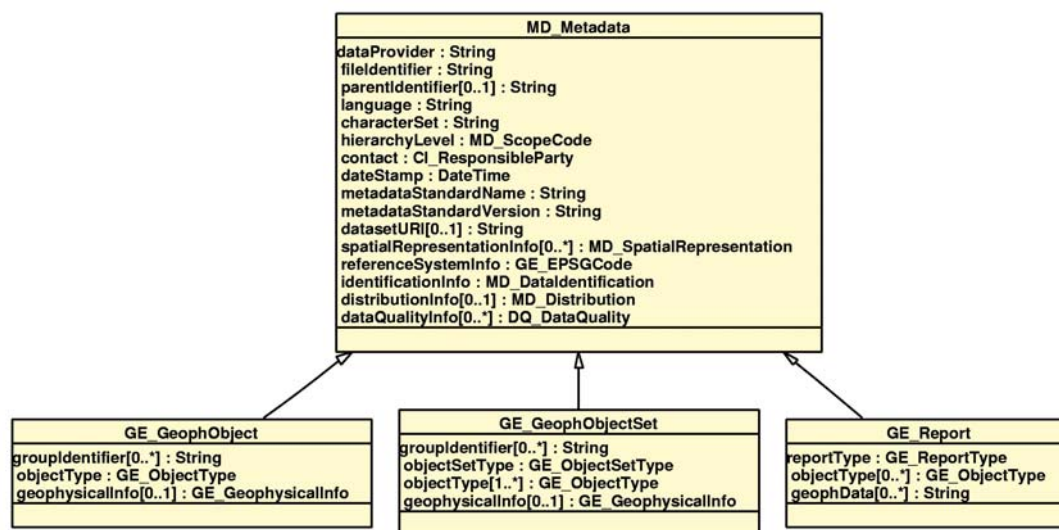
7. ábra. A GEOMIND profil az alapszabvány magját és a geofizikai kiterjesztéseket tartalmazza.

Ezeknek a használata a geofizikai metaadat rekordok méretét feleslegesen növelné. Ugyanakkor számos olyan információ hiányzik az ISO 19115-ből, amelyek a geofizikai adatok szempontjából fontosak. Ezen információk leírására a GEOMIND profil külön

elemeket vezet be, és a kiterjesztés szabályait figyelembe véve a következő változtatásokkal él:

- A meglévő struktúrák új metaadat elemekkel való bővítése
- Új témakör (section) bevezetése „geofizikai információk” (geophysicalInfo) néven.
- Új kódlistákat definiálása, a meglévők új elemeket bővítése.
- Bizonyos meglévő kötelezettségek szigorítása

A legszembevetőbb változtatás a gyökér elem (MD_Metadata) kiterjesztése három általam definiált származtatott osztállyal.(8. *ábra*) Ezek a GeophObject (geofizikai objektum), a GeophObjectSet (geofizikai objektumcsoport) és a Report (jelentés). A gyökér elemből származtatott osztályok (specified classes) rendelkeznek az MD_Metadata minden tulajdonságával és ezen felül még további elemekkel is. Minden valós geofizikai adat metaadat szintű leírása e három típus egyikével történhet.



8. ábra A GEOMIND profil osztályainak származtatása az ISO 19115 gyökér eleméből.

2.4.1. Geofizikai metaadat osztályok definiálása

Az alábbiakban bemutatom az ISO szabvány geofizikai kiterjesztését alkotó új metaadat osztályokat.

GE_GeophObject (GO)

A geofizikai objektum absztrakció, amely mérések, vagy modellek leírására szolgál. A mérés fogalma módszerenként változó, és különböző komplexitással jellemezhető hierarchia szinteket jelenthet. (Egy gravitációs mérési pont és egy szeizmikus szelvény komplexitása nagyon eltérő.) Megállapodás kérdése, hogy egy adott módszer esetében mit tekintünk mérésnek. Általánosságban azt mondhatjuk, hogy mérés a mérési tevékenység során előálló adatok halmaza, amely a kutatás természetes alapegységét képezi. A mérési adatok leírásával a később bemutatandó Általános Geofizikai Adatmodell (GGDM) foglalkozik. A GEOMIND által támogatott geofizikai objektumok típusait a GE_ObjectTypeCode nevű kódlista (**4. táblázat**) sorolja fel, ami lényegében geofizikai módszerek bővíthető listája. (Mivel az eltérő objektum típusokhoz némileg eltérő metaadat szerkezet tartozik, ezért a bővítés az XML séma változtatását vonja maga után).

Az inverziós modell is geofizikai objektum. Geometriával és kapcsolt fizikai paraméterek együttesével jellemezhető entitás, amit modellezés, vagy inverzió során egy, vagy több mérésből kapunk. A modellek részletes leírását szintén az Általános Geofizikai Adatmodell (GGDM) tárgyalja.

GE_ObjectTypeCode	Geophysical object types (geophysical methods)
VES	Vertical Electric Sounding
TDEM	Time Domain Electromagnetic sounding
MT	Magnetotelluric sounding
tellurics	Telluric soundings
gravimetry	Gravimetry (single station)
magnetometry	Magnetometry (single station)
radiometry	Radiometry (radioactive emission intensity measurements) (single station)
airborneComplex	Complex airborne measurement
obsMagnetometry	Observatorial Mag recordings
obsExtensometry	Observatorial recordings of Earth deformations
seismicFieldData	2D Seismic field data
seismicProfile	2D Seismic profile (processed)
stackingVelocity	Stacking velocity data (from seismic processing)
tomoVelocity	Tomographic velocity data
boreholeLog	Borehole logging
ambientNoiseSeismology	recordings of ambient seismic noise
strongMotionSeismology	recordings of seismological vibrations

rockSamplePetrophysics	petrophysical measurement data on rock samples
borehole	borehole data
model	geophysical model

4. táblázat A GEOMIND profilban definiált objektum típusok. (A bővíthető lista a 2008-as állapotot tükrözi)

GE_GeophObjectSet (GOS)

A geofizikai objektumok gyakran alkotnak természetes csoportokat. A geofizikai objektumcsoportok típusai (a GE_ObjectSetTypeCode lista elemei) a következők: repository (adattár), project (projekt), campaign (kampány), objectSet (objektum csoport), objectGroup (másodlagos csoport), geophCoverage (geofizikai fedvény). A repository magas szintű osztályozatlan objektum halmaza, vagy objektumcsoportok aggregátumát jelenti. A project adminisztratív egység, amely egy, vagy több mérési kampány adatait fogja össze. A kampány egy adott mérési módszerrel végrehajtott, egymáshoz kapcsolódó mérések homogén halmaza. Az objectSet nem meghatározott módon összekapcsolt objektumok halmaza. Az objectGroup mérések alkalmi, vagy permanens, másodlagos csoportja. Minden objektum gyermeke egy és csakis egy elsődleges csoportnak, ami általában egy mérési kampány (campaign), vagy más objektum csoport (objectSet). Ezzel szemben a másodlagos csoportok sohasem szülők. Egy geofizikai objektum több ilyenhez is tartozhat. A geofizikai fedvény szintén az objektumcsoport hierarchia szinthez tartozik. Lényegében egy speciális másodlagos csoportként fogható fel, amely geofizikai objektumok adott halmazából készült. Paraméter térképek grid, vagy vektor típusú alapadatai (pl. digitalizált szintvonalak) tartoznak ebbe a kategóriába.

GE_Report (REP)

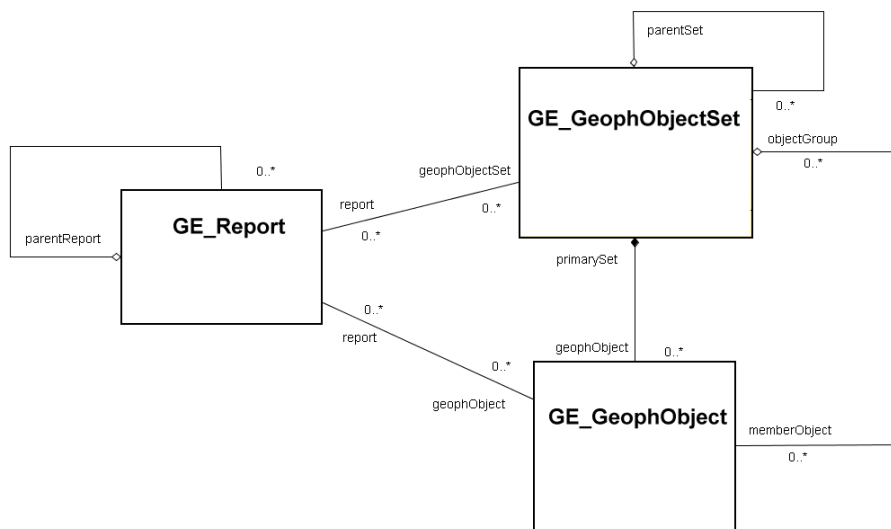
A report tetszőleges dokumentációs elem, vagy elemek aggregációja (ld. 1. függelék), ami egy geofizikai objektumhoz, vagy objektumcsoporthoz kapcsolható. Ennek a típusai (a GE_ReportTypeCode lista elemei): map (térkép), profile (szelvény), sounding (szondázás), text (szöveg), report (jelentés). Minden report elem tartalmaz egy listát, amely a kapcsolódó geofizikai adatokra mutat.

A legtöbb geofizikai adat-, vagy dokumentációs rendszer az itt bemutatott három osztály egyedeivel és azok kapcsolatrendszerével modellezhető. A GEOMIND profil nemcsak a

leendő portál internetes metaadatbázisának, hanem bármely lokális metaadatbázisnak alapja lehet.

Kapcsolatok az objektumok között

A modell rugalmassága a lehetséges kapcsolatok sokféleségében rejlik. A három fő osztály segítségével igen bonyolult adatrendszerek is modellezhetők. A mérés – kampány – projekt jellegű hierarchiák leírásán túl lehetőség van komplex dokumentációs rendszerek geofizikai adatokhoz kapcsolására. A fő osztályok közötti kapcsolatokat, és a kapcsolatokban résztvevő osztályok szerepkörét (role) a **9. ábra** mutatja be.

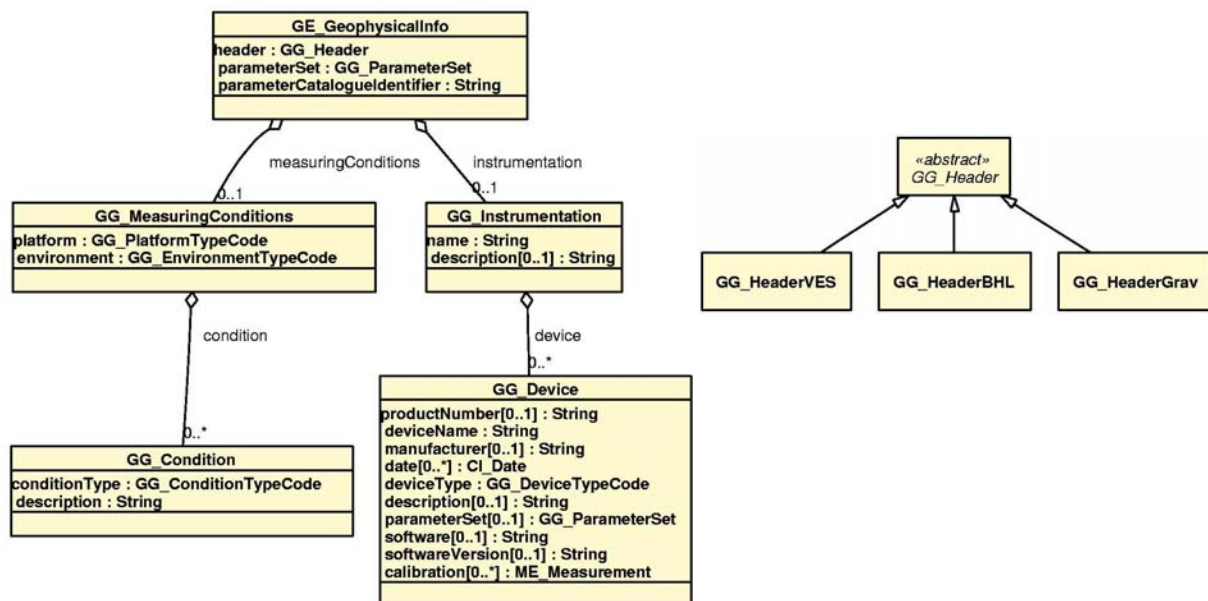


9. ábra A GEOMIND származtatott osztályok kapcsolatrendszere

- Az objektumok kompozíciója (ld. 1. függelék) alkotja az elsődleges csoportokat (primarySet). Pl. mérés – kampány kapcsolat
- Az Objektumok aggregációja alkotja a másodlagos csoportokat (objectGroup) pl. mérés –szelvény kapcsolat
- Az objektumcsoportok aggregációja hierarchikus rendszereket képez. Pl. kampány – projekt kapcsolat
- A jelentések és a forrás objektumok, illetve objektum csoportok összekapcsolhatók
- A jelentések komponensei hierarchikus rendszereket képeznek. Pl. részjelentések, ábrák, térképek összekapcsolása komplex jelentésekké.

2.4.2. Geofizikai információk a GEOMIND profilban

A szabvány kiterjesztés külön szakaszt szán a geofizika-specifikus információk leírásának. Ide kerül számos olyan metaadat, amely hagyományos leírásokban fejléc (header) adatként szerepel. A geofizikai információkat tároló (GE_GeophysicalInfo) szerkezetben lehetőség van megadni a műszerezettségre, mérési körülményekre vonatkozó információkat, a mérés fontosabb technikai paramétereit (header paraméterek). Szükség lehet a felhasználó által definiált saját paraméterek megadására. Ez a GG_ParameterSet szerkezettel történhet. A paraméterek azonosíthatósága miatt kötelező megadni a definíciókat tartalmazó paraméter katalógus azonosítóját. A geofizikai információkat tartalmazó elem UML diagramját a **10. ábra** mutatja.



10. ábra Geofizikai információk a GEOMIND profilban.

A mérési körülmények, műszerezettség, és header paraméterek megadására is lehetőség van.

GG_MeasuringConditions (mérési körülmények)

Az adatok, a feldolgozás, vagy értelmezés szempontjából fontos mérési körülményeket sorolja fel. Megadható a mérés platformja (légi-, földi mérés), és földrajzi helyzete (onshore, offshore). A lista a kondíció típus nevét, és egy szabad szöveges leírást tartalmaz. (topográfiai viszonyok, időjárási viszonyok, talaj kondíció stb)

GG_Instrumentation (műszerezettség)

A műszer együttes megnevezését, és rövid leírását, valamint a méréshez használt eszközök részletes listáját tartalmazza. Az eszközök leírásánál megadhatjuk a gyártási számot, gyártót, típust, gyártási-, vásárlási dátumot, a műszer szoftver nevét, és a kalibrációs mérések azonosítóit. A szerkezet részletes műszer katalógus kialakítására is alkalmas.

GG_Header (fejléc)

A fejléc adatok a GEOMIND modellben olyan metaadatok, amelyek nem férnek bele az eredeti ISO19115 szabványba. Az operátor, feldolgozó, mérés helye, ideje stb., vagyis a hagyományosan fejléc adatnak tekintett információk tipikus ISO konform metaadatok. Ezzel szemben a header rövid technikai információkat tartalmazó attribútum tábla, amely segíti a felhasználót abban, hogy az adatokról műszaki értelemben helyes képet kapjon. A fejléc szerkezete mérési módszerenként különböző. A GG_Header-ből származtatott osztályok tartalmazzák a módszerfüggő fejléc paramétereket. Természetesen ezekből nem csak három van, de az ábrán nem fért el több. A header táblák és a méréseket reprezentáló geometriai elemek használatával egyszerű és hatékony térinformatikai alrendszerek építhetők ki. Ezek az adatok könnyedén kiolvashatók a metaadat rekordokból és hagyományos GIS alkalmazásokba exportálhatók. A nyílt forráskódú Geoserver rendszer segítségével akár WMS és WFS szolgáltatásba is bekapcsolhatók.

GG_ParameterSet (paraméterkészlet)

Az adatrendszer szempontjából fontos technikai paraméterek és azok értékének listája. (pl. egy Bouguer-anomália térkép esetében a korrekciós sűrűség) A paraméter megadása név – érték párral történik. A név azonosítja a hivatkozott paraméterkatalógusban szereplő típusdefiníciót.

GG_ParameterCatalogueIdentifier (paraméterkatalógus azonosító)

Minden paraméter, amely egy GEOMIND metaadat rekordban szerepel megtalálható a hivatkozott paraméterkatalógusban. A paraméterkatalógus típusdefiníciókat tartalmaz,

amelyben a paraméterek neve, mértékegysége, adattípusa, alapértelmezés szerinti és/vagy lehetséges értékei szerepelnek. A magyarázó leírás mellett a típusdefiníció egy parameterSet-et is tartalmazhat, ami lehetővé teszi paraméterfüggő paraméterek definiálását. A szabvány kiterjesztéshez kapcsolódóan kidolgozásra került egy GEOMIND paraméterkatalógus, amely számos, különböző mérési módszernél használt paraméterhez szolgáltat definíciókat. A katalógus a meglévő nemzetközi szabványok figyelembevételével készült. Mivel a GEOMIND paraméterkatalógus használata nem zárja ki saját fejlesztésű paraméterkatalógusok használatát, a metadat leírásokban ezekre hivatkozni kell. Ezt a célt szolgálja a paraméterCatalogueIdentifier elem.

Egy minta állomány

Ezzel áttekintettem a GEOMIND profil legfontosabb elemeit, a metaadat alapszabványt és a geofizikai adatrendszerek leírására szolgáló kiterjesztést. A jobb érthetőség érdekében az alábbiakban példaként közlöm egy tranziens mérési kampány adatait tartalmazó objektumcsomag metaadat leírását. A tagolás a beágyazott struktúrákat szemlélteti. A metaadat elemek neve normál, az adatok dőlt betűs szedésben láthatók. A példa az áttekinthetőség kedvéért az ISO19115 szerkezeteket leegyszerűsítve mutatja.

A 2005-ös üveghutai tranziens mérési kampány GEOMIND metaadat leírása:

```
MD_Metadata
  dataProvider   http://www.elgi.hu
  fileIdentifier GOS_TDEM_uh2005
  language      en
  characterSet   utf8
  hierarchyLevel geophObjectSet
  contact
    individualName „Sőrés László”
    organisationName ELGI
    role           custodian
  dateStamp      2007.09.17
  metadataStandardName ISO19115-Geomind
  metadataStandardVersion 2007
  referenceSystemInfo EPSG:23700
  identificationInfo
    citation
      title „TDEM soundings at Uveghuta in 2005”
      date
        dateType measurementStart
        date 2005-07-15
```

```

date
  dateType  measurementEnd
  date      2005-09-30
identifier
  authority  ELGI
  code       GOS_TDEM_uh2005
abstract    „The dataset contains TDEM CIL soundings over the northwestern part of the
uveghuta grid measured in 2005. The measurements are extensions to the previous grid. The aim
of measurements was to investigate the Carboniferous granite surface and the structure of the
loose sediment cover. The measuring campaign is related to the Bataapati nuclear waste
disposal site assessment project“
status      completed
pointOfContact
  individualName  „Pém József“
  organisationName  ELGI
  role            operator
pointOfContact
  individualName  „Sóres László“
  organisationName  ELGI
  role            partyChief
descriptiveKeywords
  thesaurusName
  citation
    title „GEOMIND Thesaurus“
  keyword  TDEM, sounding, resistivity, water, basement, electromagnetic, inversion,
radioactive waste
resourceConstraints  restricted
spatialRepresentationType  vector
spatialResolution  10000
language  hu, en
characterSet  utf8
topicCategory  geoscientificInformation
EX_Extent
  description  „Northwestern part of the Uveghuta TDEM grid“
EX_GeographicExtent
  EX_GeographicBoundingBox
    westBoundLongitude  615000
    eastBoundLongitude  616400
    southBoundLatitude  95100
    northBoundLatitude  96500
  EX_GeographicDescription
    authority
    code  Bataapati
temporalElement
  startTime  2005-07-15
  endTime    2005-09-30
verticalElement
  minimum  0
  maximum  80
  verticalDatum  DEPTH
distributionInfo

```

```

distributor
  distributorContact
    organisationName  MBFH
    role  distributor
  distributionOption
    distributionOrderProcess
      fees
      instruction „Please, initiate the Geomind ordering procedure“
    distributionFormat
      name  Geomind_xml
      version  1.0
      decompressionTechnique  tar.zip
    distributionTransferOption
      online  http://www.geomind.hu/dataroom/GOS_TDEM_uh2005
dataQualityInfo
  statement  „standard preprocessing was completed“
  processingStep
    description  „averaging of individual records“
  processingStep
    description  „apparent resistivity calculation“
  processingStep
    description  „manual masking of noisy channels“
  processingStep
    description  „1D Marquardt inversion“
objectSetType  campaign
objectType  TDEM
geophysicalInfo
  measuringConditions
    platform  groundBased
    environment  onShore
    condition
      conditionType  terrainConditions
      description  „rugged terrain“
    condition
      conditionType  soilConditions
      description  „upper 30cm very dry“
instrumentation
  name  ELGI_TDEM-01
  description  „Digital PROTEM receiver with TEM47 transmitter“
  device
    deviceName  Prot57_1
    manufacturer  Geonics
    deviceType  TDEM_Receiver
    software  prot57.exe
    version  2.0
  device
    deviceName  TEM47_1
    manufacturer  Geonics
    deviceType  TDEM_Transmitter
parameterSet
parameterCatalogueIdentifier  PCAT_TDEM_GEOMIND

```

Még néhány szó a hierarchiáról

A metaadat rekordok lehetnek igen egyszerűek és igen terjedősek. Mivel a szabvány előír egy kötelező minimumot, a legkisebb méret adott. A másik vélet: ha kihasználjuk a struktúra kínáta lehetőségeket, és sok információt közlünk, akkor a metaadat rekordok mérete túlságosan nagy is lehet. Ugyanakkor nem szükséges minden információt minden szinten tárolni. Pl. egy adatrendszer tulajdonosát, vagy terjesztőjét elegendő egy helyen, mondjuk a projekt metaadat rekordjában megadni. Ilyenkor az alkalmazásnak gondoskodnia kell arról, hogy alacsonyabb hierarchia szintekről is megtaláljuk a keresett információt. Ha ez a mechanizmus nem áll rendelkezésre, akkor az adott információt kénytelenek vagyunk minden szinten redundánsan tárolni.

2.4.3. XML implementáció

A ISO 19115 szabvány a metaadatok szerkezetének logikai modelljét írja le. A konkrét implementációra vonatkozóan csak ajánlásokkal és általános alapelvekkel szolgál. Az adatszerkezetek konkrét megvalósítását az ISO 19139 séma definíciós csomag (schema definition package) határozza meg.

Az XSD sémákról később részletesen szó lesz. Előljáróban csak annyit, hogy a séma egy speciális XML fájl (xsd fájlnev kiterjesztéssel), amely más XML fájlok szerkezetét írja le. Használatával előre definiálható egy adatfájl szerkezete, a struktúrát felépítő adatok típusa, száma, formátuma. Az XML fájlok a séma ellenében érvényesíthetők (validation), ami az adatfájl szerkezetének automatizált ellenőrzését jelenti.

A GEOMIND adatmodell implementációja az ISO 19139 megfelelő módosításával történt. A séma definíciós csomag az Internetről szabadon letölthető (GEOMIND Consortium 2007b)

A séma ellenében érvényesített XML formátumú metaadat rekordok százezerrel találhatók a <http://www.geomind.eu> címen elérhető portálon, ami szemléletesen mutatja be az adatmodell alkalmazásában rejlő lehetőségeket.

3. Az Általános Geofizikai Adatmodell (GGDM)

Történeti áttekintés

Az Általános Geofizikai Adatmodell fejlesztése a GEOMIND projekt során 2006-ban kezdődött. A GEOMIND profillal együtt készült, azzal a céllal, hogy az adatszolgáltatás egyik támogatott szabvány formátuma legyen. A neve GEOMIND részletes adatmodell (GEOMIND Detailed Data Model) volt. (Sörös L. et al. 2007a) Az eredeti adatmodell elkészítésében személyes konzultációk erejéig a GEUS, a GGA, és az ITG szakértői is segítségemre voltak. Az modellhez tartozó XSD csomag elkészítése a GEUS-szal közösen végzett munka volt. Mivel a portált kizárólag metaadatok kezelésére terveztük az adatok és metaadatok teljes mértékben szeparálva lettek. Az adatmodell komplett dokumentációja példákkal együtt a projektjelentésben megtalálható. (Sörös L. et al. 2007b-c) A GEOMIND projekt zárása után az adatmodellen tovább dolgoztam. A céloom egy önmagában is megálló teljes rendszer kialakítása volt, ezért számos változtatásra volt szükség. Amint azt bemutattam, a GEOMIND metaadat modellben a geofizikai osztályok az ISO 19115 MD_Metadata elemből származnak. Ez a portál céljait tekintve indokolt megoldás volt, bár azt a képzetet sugallta, hogy a metaadatok magukba foglalják a geofizikai méréseket. A látszólagos ellentmondás könnyen feloldható: A GEOMIND adatmodellben a GE_GeophObject, GE_GeophObjectSet, GE_Report metaadat alosztályok, amelyeket a GE csomag (package) tartalmaz. Az új GGDM adatmodellben az objektum, objektumcsoport és riport osztályok (HG_GeophObject, HG_GeophObjectSet, HG_Report) a hierarchia csúcsára kerültek, és mint befoglaló elemek, a mérési és modell adatokat, valamint a metaadatok alrendszerét is tartalmazzák. A GGDM adatmodellben a metaadat leírás a GEOMIND adatkonverziók során szerzett tapasztalatoknak megfelelően némileg leegyszerűsített, de továbbra is pontosan követi az ISO 19115 előírásait.

Modell és valóság, az adatmodellezés

A valós világ adatainak és a közöttük levő kapcsolatok elvont kategóriáinak absztrakt ábrázolását adatmodellnek nevezzük. Az adatmodell tervezése során a leképezendő rendszert elemzésnek vetjük alá, és meghatározzuk a tárolandó adatok körét. Természetesen figyelembe

kell venni azokat a felhasználói elvárásokat, amelyeket a leendő adatbázis kiszolgálni hivatott. A GGDM tervezésénél az adatok dokumentálhatósága, az adatfeldolgozás, inverziók támogatása, térinformatikai integrálhatóság egyaránt fontos szempont, emiatt az adatkör igen kiterjedt. A rendszer logikai tervezésénél a tárolandó adatok halmazát szorosan összetartozó egységekbe szervezzük és megállapítjuk ezek egymáshoz viszonyított kapcsolatait. Az adatmodellek történeti fejlődése során számos leírási mód alakult ki, amelyek az adatbázis-kezelő rendszerek alapvető működési elveihez igazodtak. A hierarchikus és hálós adatbázis-kezelő rendszerek után a relációs adatbázis-kezelők döntő szerephez jutottak. A jövő várományosai az objektumorientált adatbázis-kezelők, de jelenleg még a relációs adatbázis-kezelők a terep. Az ilyen rendszerek leírásának hagyományos módja az Egyed/Kapcsolat (Entity/Relation) modell használata. A GGDM konkrét implementációját leíró részben ezekről részletesebben szó esik, azonban az adatmodell ismertetését a magasabb szintű reprezentációt jelentő UML osztálydiagramokkal folytatom. Mégis ide kívánczik néhány szó az Egyed/Kapcsolat modell tervezés és optimalizáció alapelveiről. A relációs adatmodell táblázatokból épül fel. (Minden lekérdezés eredménye egy táblázat, ezeket nevezzük relációnak.) Minden egyedhez tartozik egy adattábla, melynek sorai az egyes példányokat tárolják. Az oszlopok elnevezései az egyedek tulajdonságait (attribútumok) jelölik. Az adatbázis tervezés legfontosabb célja, hogy az egyedeket és ezek kapcsolatait úgy definiáljuk, hogy az adattárolás redundanciáit, ezzel az adatkezelés integritását megbontó esetleges hibás működést minimálisra csökkentsék. Ennek matematikailag egzakt módja az adattáblák normálása. Az adatmodellezés elmélete 5 egymásra épülő normálformát ismer. Ezek a redundancia kizárásának különböző szintjeit jelentik. A harmadik normálformára hozás a legtöbb esetben biztosítja az optimális adatszerkezetet. A probléma lényege az egyedek és a hozzájuk tartozó attribútumok körének megfelelő definiálása. A modellezett problémakör beható ismerete és nagy gyakorlat szükséges az Egyed/Kapcsolat rendszer megfelelő kialakításához. Általában igaz, hogy ha az egyedekben környezetüktől természetesen elkülönülő objektumok és valós kapcsolataik tükröződnek, akkor az őket reprezentáló táblák további normalizálás nélkül is jó eséllyel kielégítik a kívánt normálformát. Ugyanakkor bizonyos szintű redundancia a hatékony működés érdekében megengedett. Természetesen az alkalmazásnak gondoskodnia kell a redundáns szerkezetek korrekt kezeléséről. A GGDM adatmodellben az egyedek kialakítása a terepi mérések, az adatfeldolgozás, és az inverziók során zajló folyamatokból és az ezekhez kapcsolódó logikai struktúrákból következik. A tervezés során nagy segítséget jelentett számomra a terepi mérésekhez, adatkezeléshez és értelmezési munkához kapcsolódó 25 éves geofizikusi tapasztalatom. Az alább tárgyalt UML

diagramok osztályai az E/K modell egyedeinek, illetve az ezekből felépített komplex struktúráknak felelnek meg.

A geofizikai mérések és modellek leírása a GGDM adatmodellben meglehetősen absztrakt, ennek ellenére viszonylag pontosan követi azt, ami a terepi mérés során történik. A legtöbb mérés eléggé hasonló módon, nagyjából a következő lépéseket követi:

1. válogassuk össze a méréshez szükséges felszerelést
2. menjünk ki a terepre
3. állítsuk össze a mérési elrendezést (elektródák, érzékelők, kábelek...)
4. állítsuk be a technikai paramétereket (adóáram, erősítés...)
5. adatgyűjtés
6. állítsuk össze a következő mérési elrendezést és ismételjük a 4–6 pontok lépéseit, amíg a mérés kész nincs.
7. menjünk a következő mérési pontra és ismételjük a 3–7 pontok lépéseit, amíg az összes mérést el nem végeztük.

A mérési módszertől függően ez a procedúra lehet egyszerű, mint egy mágneses észlelés proton-magnetométerrel, vagy bonyolult, mint egy szeizmikus vonal hetekig tartó lemérése. Ha megvizsgáljuk a mérés folyamatát, elemi lépésekre bonthatjuk, és minden lépéshez fizikai, vagy logikai elemeket, egyedeket (entitásokat) rendelhetünk. Ezen egyedek és kapcsolataik összessége az Általános Geofizikai Adatmodell. A fenti lépéssort a GGDM terminológiáját használva a következőképpen írhatjuk le:

1. válogassuk ki a *forrás* és *érzékelő* eszközöket (source, sensor)
2. határozzuk meg a *lokális koordináta rendszerünk* középpontját, tengelyirányait, magasságát
3. állítsuk be a *terítést* (layout) a *terítés komponenseinek* (layout component) pozíciójának, irányítottságának rögzítésével
4. adjunk értékeket a komponenseket jellemző *paramétereknek*
5. rendeljünk a terítés komponenseihez *mérési tartományokat* és *érték sorozatokat*
6. állítsuk be a következő terítést és ismételjük a 4–6 pontok lépéseit, amíg a mérés kész nincs.
7. váltsunk koordináta rendszert és ismételjük a 3–7 pontok lépéseit, amíg szükséges

A dőltbetűvel szedett fogalmak a GGDM adatmodell alapvető elemeit képviselik. Ezek lehetnek fizikai (forrás, érzékelő), vagy logikai egységek (lokális koordináta rendszer, terítés), de mindig saját funkcióval és jól meghatározható kapcsolatokkal bírnak.

3.1 Geofizikai mérések általános leírása

A „geofizikai módszer” fogalmának elvetése, a layout, layoutComponent, domainSet, rangeSet fogalmak bevezetése

A geofizikai mérések leírásában kiemelkedő szerepet játszanak a paraméterek. A metaadat modellel kapcsolatban már esett szó a paraméter katalógusokról. A geofizikai mérések és modellek részletes tárgyalását célszerű ezekkel kezdeni.

3.1.1. Paraméterek és paraméter katalógusok

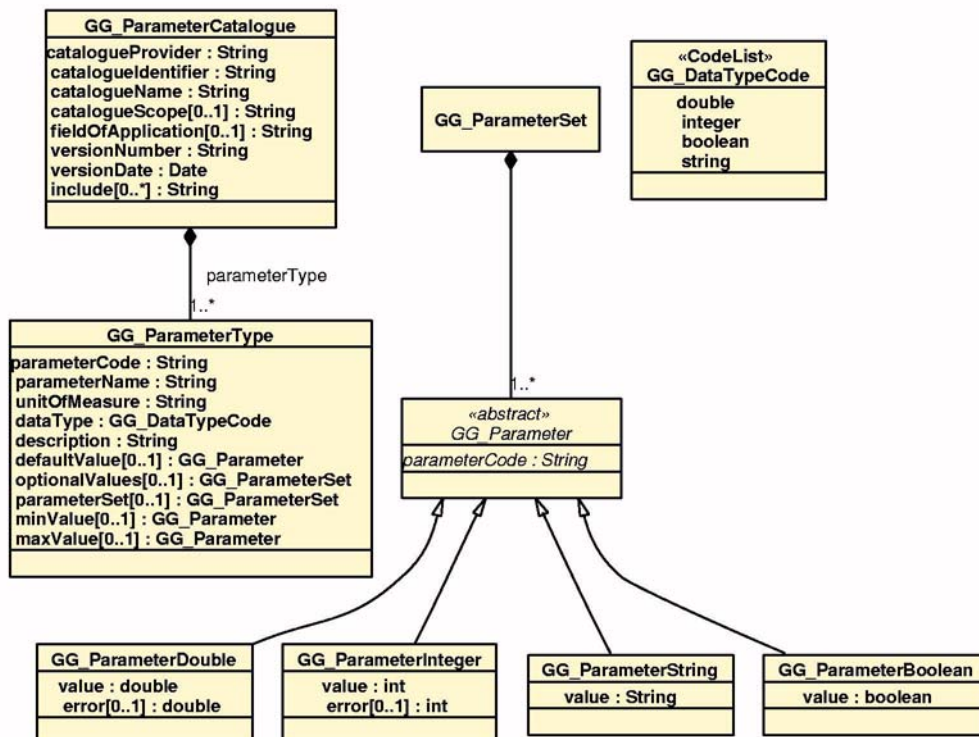
A paraméterek valójában név-érték párok, amelyeket leginkább technikai paraméterek és fizikai mennyiségek tárolására használunk. A név (vagyis kód) azonosítja a paramétert, az érték numerikus, szöveges, vagy logikai változó. A paraméterek használatának egységesítése gyakorlatilag megoldhatatlan feladat. A kódok, mértékegységek sokasága és az évtizedek során rögzült eltérő szokások miatt ez még a nagy kőolajipari szabványok esetében sem sikerült igazán. A probléma kezelésére a GGDM a paraméter katalógusokat használja. Ezek nyilvános paraméter definíciós gyűjtemények, amelyek leírják, hogy egy adott kóddal jelölt paraméter hogyan értelmezendő. **(11. ábra)** A paraméter típus definíciók kötelezően tartalmazzák a paraméterek kódját, általánosan ismert nevét, mértékegységét, leírását. Ha van értelme, megadható az alapértelmezés szerinti érték és a megengedett minimum, maximum.

parameterCode	GRAV_BOUG
parameterName	bouguerAnomalyGeomind
unitOfMeasure	mGal
dataType	double
description	Geomind Bouguer anomaly calculated by the proposed standard procedure [Hinze et al.2005] Applied datums corrections and density are listed in parameterSet.

defaultValue	
optionalValues	
parameterSet	DTM_H=WGS84 DTM_V=GRS80 DTM_GR=IGSN71 DENS=2670.0 EQ_CORR_NRM=Somigiana EQ_CORR_HGT=SecondOrderFormula EQ_CORR_BG=SphericalCap PROC_REF=PROC_TC

5. táblázat A GRAV_BOUG paraméter definíciója a GEOMIND gravitációs paraméterkatalógusban.

Minden paraméter definíció tartalmazhat két speciális paraméterkészletet (optionalValues, parameterSet). Az egyiket opcionális értékek megadására (dinamikus kódlisták definiálása), a másikat beágyazott paraméterek tárolására használhatjuk. Erre jó példa a GRAV_BOUG paraméter definíciója a GEOMIND gravitációs katalógusából. (5. táblázat) A paraméterkészletben megadhatjuk, hogy egy adott származtatott paraméter milyen más paraméter felhasználásával kap értéket (pl. DENS=2670.0, vagyis 2670 kg/m³ sűrűséggel számolt Bouguer érték).



11. ábra A GG_ParameterSet és a GG_ParameterCatalogue osztály UML diagramja

A paramétereknek négy típusa van, melyek az absztrakt GG_parameter osztályból származnak, Ez egyetlen kötelező attribútumot, tartalmaz, a paraméter kódját. Az érték típusának megfelelően lehet dupla pontos, vagy egész szám, szöveges, vagy logikai érték. Az érték megadása kötelező. A numerikus paramétereknél megadható a hiba is, ami alapértelmezésben a szórással azonos. Amennyiben nem, a típus definíciójánál a hiba értelmezése is megadandó. A paraméterkód azonosítja a paraméter típusát a dokumentumhoz kapcsolódó metaadat rekordban hivatkozott katalógusban. Paraméter katalógus bejegyzésre példa a 2. függelék „gravitációs mérési adatok” című részében található.

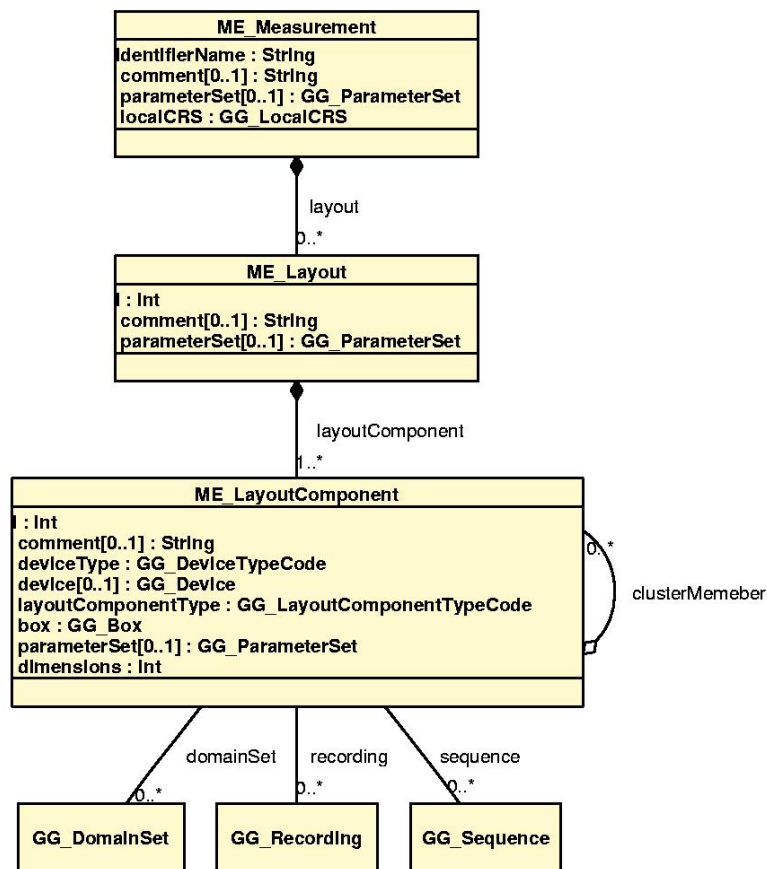
3.1.2 A mérés (GG_Measurement)

A GGDM adatmodell a mérést terítések (layout) kompozíciójaként írja le (**12. ábra**). A terítés magába foglalja az összes elemet (entitás, egyed) ami az adott mérési elrendezéshez hozzátartozik, beleértve a mérési geometriát, a felhasznált eszközöket, és a gyűjtött adatokat. A mérésen belül a terítések száma nincs limitálva, és lehet zéró is. Ilyenkor az adatokat az ME_Measurement osztály részeként megadott paraméterkészletben tároljuk. Olyan esetekben, amikor a mérési elrendezés részletes dokumentálása szükségtelen, a mérés helyének és a mért paraméter értékeknek a tárolása elegendő (gravitációs, mágneses mérések). A másik véglet: szeizmikus méréseknél az adatrendszer terítések sokaságából áll össze, és minden terítés akár több ezer szenzort is tartalmazhat. Ezeket a terítéseket egy közös, lokális koordináta rendszerben adjuk meg. Ennek definiálására való a GG_LocalCRS osztály.

3.1.3. A helyi koordináta rendszer (GG_LocalCRS)

Sok esetben a terepen egy alkalmasan kiválasztott helyi koordináta rendszerben rögzítjük a terítések komponenseinek helyzetét. Pl. egyszerűbb, ha egy vonalmenti mérés esetén csak távolságokat tárolunk, pontos koordináta transzformációval számítandó koordináta értékek helyett. Természetesen a helyi koordináta rendszer origóját, tengelyirányait, magasságát rögzítenünk kell. Erre szolgál a GG_LocalCRS szerkezet. (**13. ábra**) Az origót több globális koordináta rendszerben is megadhatjuk, használhatunk derékszögű, vagy gömbi koordináta

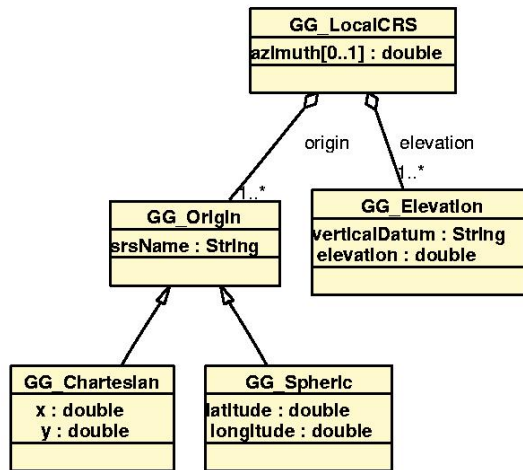
rendszert. Nincs akadálya tehát, hogy méréseink helyét mind EOV, mind WGS84 rendszerben tároljuk. A használt globális rendszer EPSG kódját viszont kötelező megadni.



12. ábra Az ME_Measurement osztály UML diagramja

Az European Petroleum Survey Group az OGC által is támogatott nyilvános adatbázisban tárolja a különböző országokban használt vetületi rendszereket definiáló adatokat. Az EPSG kódra való hivatkozás ezért elegendő bármilyen rendszer használata esetén a pontos lokalizáláshoz. A lokális rendszer origójának magassága szintén megadható, több függőleges vonatkoztatási rendszerben is. A vertikális dátum kódjának megjelölése kötelező, ehhez egy kódlista áll rendelkezésre.

Persze nincs akadálya, hogy a lokális koordináta rendszer egybe essen a globális vetületi rendszerrel. Ilyenkor a mérési geometriáját valódi koordinátákkal adjuk meg.



13. ábra A GG_LocalCRS osztály UML diagramja

A helyi koordináta rendszer megadására példák a 2. függelékben találhatók.

3.1.4. Terítési komponensek (ME_LayoutComponent)

A terítést (layout) terítési komponensek (layout component) együtteseként definiáljuk. Kompozícióról van szó, tehát, ha egy terítést törölünk, akkor a komponenseit is törölni kell az adatbázisból. Két fajtájuk van, forrás (source) és szenzor (sensor). Logikai megfelelői a terepen használt fizikai eszközöknek. Az eszköz lehetséges típusait a deviceType nevű bővíthető kódlista tartalmazza (6. táblázat). Elemei között vannak egyszerű fizikai eszközök, mint pl. elektromos dipól (electricDipol), vagy bonyolult érzékelők, mint pl. magnetométer (magnetometer). Az értékek kapcsolóként használhatók az adatkezelő alkalmazásokban, melyek meghatározzák, hogy az adott komponens miként kezelendő.

GG_DeviceTypeCode	domain value	description
currentElectrode	1	electrode to drive electric current
potentialElectrode	2	electrode to measure electric potential
currentDipole	3	dipole to drive electric current
electricDipole	4	dipole to measure voltage
inductionCoil	5	coil to measure magnetic field time variations
rectangularCurrentLoop	6	loop to generate EM field
circularCurrentLoop	7	loop to generate EM field
magnetometer	8	device to measure magnetic field
vectorMagnetometer	9	device to measure 3 magnetic field component

extensometer	10	device to measure dilatation
gravimeter	11	device to measure gravity field
spectrometer	12	device to measure nuclear emission intensity
geophone	13	seismic sensor (velocity meter)
hydrophone	14	seismic sensor (accelerometer)
seismometer	15	sensor to measure ground motion
TDEM_Receiver	16	equipment to measure TDEM data
TDEM_Transmitter	17	equipment to generate TDEM source field
resistivityMeter	18	device to measure electric resistivity
combinedloggingTool	19	Combined tool with several devices
acousticProbe	20	Logging tool for acoustic measurements
neutronProbe	21	Logging tool for neutron porosity
resistivityProbe	22	Logging tool for resistivity measurements
virtualDevice	23	virtual device related to calculated parameters

6. táblázat A terítési komponensekhez tartozó eszköz típusok (*deviceType*) listája

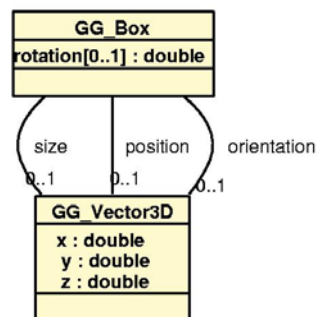
A terítési komponens (layout component) által logikailag reprezentált valóságos eszközt a GG_Device osztály írja le. Ez az elem a metaadat kiterjesztés tárgyalásakor a műszerezettség leírásában már szerepelt. (ld. **10. ábra**) Itt az opcionális device[] tömb az adott eszköz(ök)re mutató pointer. Az eszköz lehet virtuális (virtualDevice), ami nem meghatározott, vagy elképzelt eszközt jelent. Ez lehetővé teszi a számított paraméterek mérési adatokkal együtt történő tárolását, vagy a kompaktabb leírást. Pl., ha egy VESZ szondázást a mérés valódi menetének megfelelően, részletesen dokumentálni akarunk, akkor minden AB-MN kombinációt külön terítésként (layout) kell tárolnunk (hiszen az is!). Ilyenkor minden terítés két komponensből, egy forrás (AB) és egy érzékelő elektromos dipólusból (MN) áll. Egy kompaktabb leírásban az azonos MN-hez tartozó AB sorozatokat összevonhatjuk. A mérésben csak annyi terítés lesz, ahány MN, és terítésenként 1-1 komponens, amit egy virtuális eszközhöz kapcsolunk. (Egy ilyen képzeletbeli eszköz egy menetben lemérné a teljes AB sorozatot)

Bizonyos módszereknél nincs szükség forrásra, mert természetes fizikai tereket mérnek (gravitációs, mágneses mérések, MT, szeizmológia). Ellenben nehéz elképzelni olyan mérést, amelyben nincsenek szenzorok. Emiatt egy terítésnek (layout) legalább egy komponens (layout component) tartalmaznia kell.

A kódolás hatékonysága érdekében a terítési komponensek klaszterekbe szervezhetők. Ez esetben a klaszter szülő komponense által tárolt adatok a gyermekekre is vonatkoznak. Ezt jelöli az UML diagramon látható „clusterMember” reflektív aggregáció. Egy triviális példa a

klaszterek használatára: MT mérésnél a három indukciós tekercs (layout component) szinkronban működik. A három tekercs egyetlen klasztert alkot, ami egyetlen idősorral jellemezhető. Az egyes tekercseknél elegendő a mért értékeket tárolni, az idősor a klaszter szülőből kiolvasható. Egy kevésbé magától értetődő eset: Sokelektrodás mérésnél az egyes elektródák szerepe a mérési konfiguráció kapcsolgatása során folyamatosan változik. A mérési adatok a kiválasztott elektródákból álló dipólusokhoz kapcsolódnak. Ha a terítéseket dipólokkal írjuk le, minden esetben meg kell adni azok pozícióját, és méretét. Ez esetben nem tudjuk kihasználni, hogy valójában ugyanazokkal az elektródákkal dolgozunk. Ha a dipólusokat klaszterként adjunk meg, amelynek tagjai az egyes elektródák, akkor elegendő azok pozícióját egyszer definiálni, a továbbiakban pedig csak hivatkozni kell rájuk. Nagy adatrendszerknél ez jelentős megtakarítást jelent.

A terítési komponensek geometriáját a GG_Box osztály hivatott tárolni (14. ábra). Ez lényegében az adott eszközt befoglaló 3 dimenziós hasáb definiálását jelenti a méret (size), pozíció (position), és irányítottság (orientation) megadásával. A koordináták a lokális koordináta rendszerben értendők.



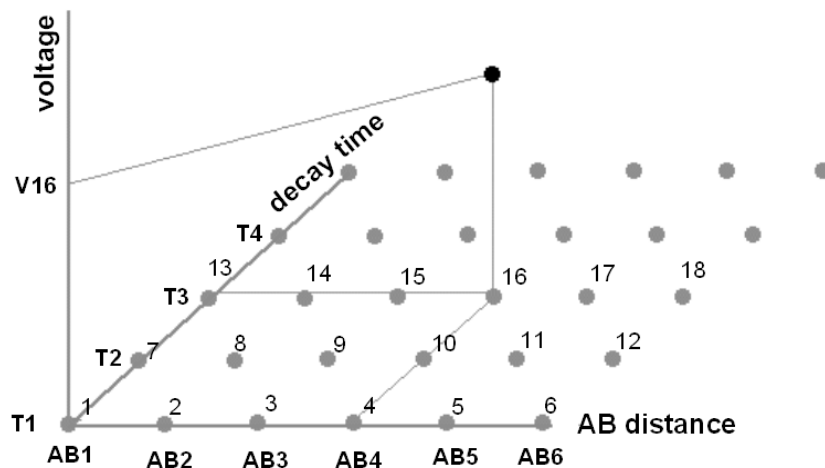
14. ábra A GG_Box osztály UML diagramja

Az eszköz típusa meghatározza, hogy az adott geometria miként kezelendő. Pl., elektromos dipólus esetén a méret egyetlen dimenzióra (x) korlátozódik, ami a dipólus hosszát jelenti ($y=0, z=0$). A dipól tengely irányát az orientáció, mint 3D egységvektor adja meg. Az opcionális rotáció paraméter az eszköz x irányú tengelye körüli elforgatást jelenti. Ennek szerepe pl., egy lyukkamera esetében lehet.

A terítési komponensek (layout component) mért és feldolgozott adatokat egyaránt tartalmazhatnak. A mérési adatok tárolása a GGDM adatmodellben hasonló a GML fedvényekhez (multipoint coverage) (ISO/TC 211, 2004). Az értéktartomány (domain set) és értékkészlet (range set) koncepciójára épül, nagyjából a hagyományos matematikai értelmezésnek megfelelően. Az értéktartomány (domain set) dimenzióként egy-egy vektorban egy pontsorozatot definiál az n dimenziós paraméter térben. Az értékkészlet (range set) pedig egyetlen vektor, ami ezekre a pontokra vetített értékeket adja meg. A dimenziószámra és az egyes dimenziókhoz kötődő paraméterekre nézve nincs megkötés, lehet tér, idő, frekvencia, energia stb. Viszont az értéktartományt (domain set) és az értékkészletet (range set) leíró adatvektoroknak harmonizálniuk kell. Egy egyszerű példa 1 dimenziós adatrendszerre: tranziens EM szondázásoknál a mérési adatokat időben lecsengő feszültség sorozat jelenti. Az értéktartomány (domain set) mint 1D vektor definiálja a mintavételi időket, az értékkészlet (range set) pedig a mérési adatokat. 2 dimenziós terítési komponensre példa lehetne egy digitális képrögzítő eszköz, 3 dimenzióra pedig egy video kamera. Bár lehetséges, nem igazán lenne célszerű videó filmeket ily módon tárolni. Egy életközeli 2 dimenziós példa: időtartománybeli GP mérések Schlumberger elrendezésben. Terítésenként (közös MN) egy-egy virtuális komponensben tárolhatjuk az adatokat. Az értéktartomány (domain set) 2 vektorral egy képzeletbeli síkban fekvő pontsört reprezentál (**15. ábra**) Az egyik tengelyen az AB elektróda távolságok sorozatát (AB distance), a másikon az idősorok mintavételi időit (decay time) ábrázoljuk. A mérési adatokat az értékkészlet egy rendezett tömbben tárolja, ami szinkronban van az értéktartomány vektoraival. Az értékkészlet tömbjének feltöltése az egyes értéktartományok sorszama szerinti rendben történik. A példában az első értéktartomány az „AB distance”, tehát a feszültség adatok tömbje így néz ki:

$$V[] = \{V1(AB1,T1), V2(AB2,T1), V3(AB3,T1), V4(AB2,T1), V5(AB3,T1), V6(AB6,T1), \\ V7(AB1,T2), V8(AB2,T2), V9(AB3,T2), V10(AB2,T1), V11(AB3,T1), V12(AB6,T2), \\ V13(AB1,T3), V14(AB2,T3), V15(AB3,T3), V16(AB2,T3), V17(AB3,T3), V18(AB6,T3) \dots\}$$

Az értéktartomány további finomítása lehetséges a hozzáadott szekvenciákkal (sequence). Ezek olyan vektorok, amelyek az értéktartomány leírását pontosítják. Pl., az előző esetről maradva, tárolják a mintavételi időkhöz rendelt kapuszélességeket.



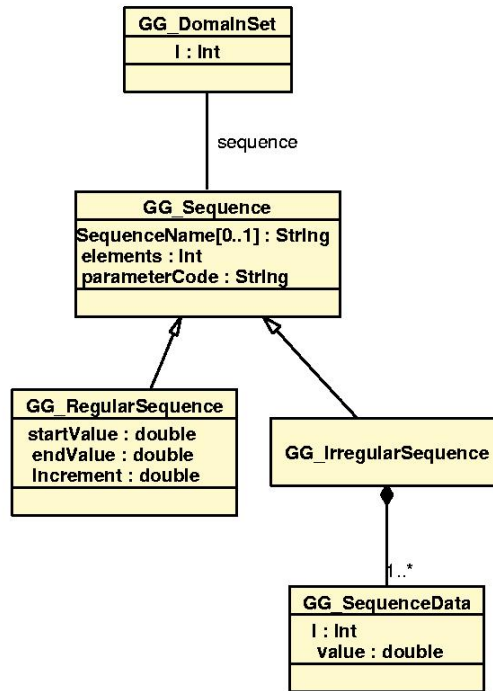
15. ábra Az értékkészlet vektor indexelése 2D virtuális eszköz esetén.

Az értékkészlet elemei a sorszámoknak megfelelően rendelhetők hozzá az értéktartomány pontjaihoz. Domain set 1: AB távolságok (AB1, AB2, AB3 ...), domain set 2: mintavételi idők (T1, T2, T3 ...) A mérési adatokat az értékkészlet tömbje tárolja (V1, V2, V3,...)

A paraméterkészletek (parameter set) többszintű használata lehetővé teszi, hogy különböző hierarchia szinteken rugalmasan definiáljunk releváns paramétereket. A terítési komponenseknél alkalmazott paraméterkészletben tárolhatunk különböző eszköz-specifikus paramétereket, mint pl., adóáram, erősítés stb.

3.1.5. Értéktartomány (GG_DomainSet)

Az értéktartomány (domain set) rögzített értékek sorozata (sequence), ami a paramétertér egy adott dimenziójához tartozik. (16. ábra) A szekvencia lehet szabályos (regular) és szabálytalan (irregular). A szabályos szekvencia egyenközű mintavételt jelent, amely kezdő és végértékkel, valamint a növekménnyel adható meg. A szabálytalan szekvenciákat (sequence) index-szel ellátott érték kompozíciói adják. Az értéktartományhoz tartozó fizikai paraméter hozzárendelése egy kóddal (parameterCode) történik. Minden szekvenciát egyedi névvel azonosíthatunk. Gyakran használt szekvenciák esetén ez lehetővé teszi a névvel történő hivatkozást. Minden értéktartományhoz (domain set) tartozik egy sorszám, ami többdimenziós adathalmaznál az értékkészlettel történő szinkronizáláshoz szükséges.



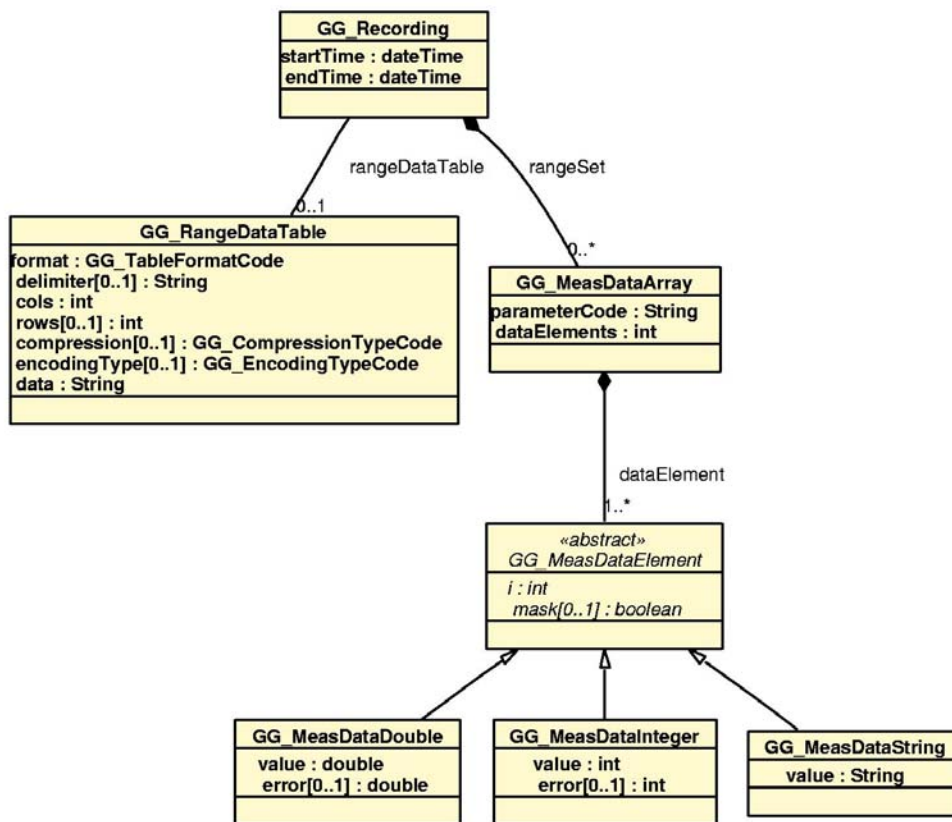
16. ábra A GG_DomainSet osztály UML diagramja

3.1.6. Felvételek, adattömbök (GG_Recording, GG_MeasDataArray)

A terítési komponensek (layout component) adatait az értékkészletek (range set) tárolják. Az összetartozó adatsorokat felvételekbe (recording) rendezzük. Minden felvétel tetszőleges számú, paraméterenként elkülönülő értékkészletet tartalmazhat. Ez lehetővé teszi többféle összetartozó, mért és származtatott adat együttes tárolását. Minden értékkészlet a GG_MeasDataArray osztály egy példánya, ami a tárolt paraméterek kódját, és az értéktömböt tartalmazza. Ez az absztrakt GG_MeasDataElement osztályból származtatott dupla pontos (GG_MeasDataDouble), integer (GG_MeasDataInteger), vagy szöveges (GG_MeasDataString) adatelemek kompozíciója. Az adatelem sorszáma az értéktartománnyal való szinkronizálás miatt fontos. A maszk a hibás adatok ki-be kapcsolására való. A numerikus adatokhoz a hiba értéke is eltárolható. A hiba alatt alapértelmezés szerint itt is a szórás értendő. Ha nem, azt a paraméter katalógusban dokumentálni kell.

Ez a szerkezet kellően rugalmas, de kissé terjengősnek mondható. Nagy adattömbök tárolására jóval takarékosabb a táblázatok használata. Ehhez egy alternatív megoldást biztosít

a GG_RangeDataTable osztály. Ez megadott elválasztó karakterekkel készített szöveges táblázatok formájában tárolja az adattömböt. Konvenció szerint minden oszlop egy paraméter értékészletének (range set) felel meg, és az első sor kötelezően a paraméter kódot tartalmazza. Szövegesen kódolt, tömörített formátumot is használhatunk, de meg kell adni a kódolás és a kompresszió típusát, hogy az adatkezelő alkalmazás dekódolni tudja az adatrendszert. A szerkezet használata különösen XML fájlokban indokolt.



17. ábra A GG_Recording osztály UML diagramja

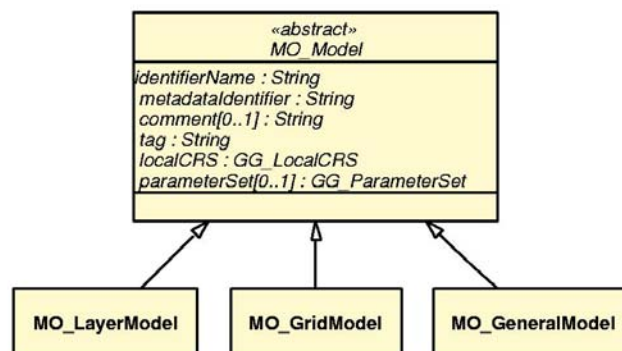
Az itt felvázolt adatmodell számos geofizikai módszerre minden gond nélkül alkalmazható. A rugalmasságát a modell komponenseinek absztrakt jellege adja. A méréseket felépítő entitások egy kirakós játék részeként egymásba, egymás mellé illeszthetők. Az elképzelhető legegyszerűbb módszertől a legbonyolultabbig egyaránt alkalmazható a geofizikai mérési

rendszerek leírására és a rögzített adatok tárolására. (38. ábra) A terítési komponensek használatára példa a 2. függelék „VESZ szondázási adatok” című részében található.

3.2. Geofizikai modellek általános leírása

3.2.1. Modell (MO_Model)

Modell alatt a numerikus modellezés, vagy inverzió eredményét értjük. A GGDM-ben a modellt a méréstől elkülönítve kezeljük. A mérés és modell kapcsolata sokrétű, maga az inverzió kapcsolja össze őket. A modell, mint a valóságot tükröző absztrakció, a fizikai paraméterek térbeli eloszlását leegyszerűsítve mutatja meg. Három típusát különítjük el: rétegmodell (layer model), rácsmodell (grid model) és általános geometriájú modell (general model) (18. ábra). Rétegmodell a hagyományos 1D modellnek felel meg. A rácsmodellek merőleges oldalfalú cellákból épülnek fel. Dimenzionalitásuk nincs korlátozva. Lehetnek 1, 2, 3, 4, vagy több dimenziósak. (Az 1D rácsmodell a rétegmodell ekvivalens reprezentációja) Az általános geometriájú modellek tetszőleges alakú testekből, tartományokból épülnek fel. A geometria leírására a ISO 19107 szabványnak megfelelő GML nyelv (Geography Markup Language) szolgál. A modellek geometriai elemeihez tetszőleges számú fizikai tulajdonságot rendelhetünk.



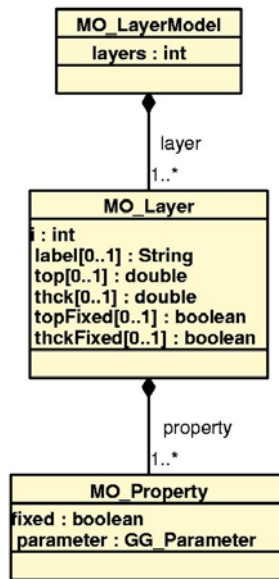
18. ábra Az MO_Model osztály UML diagramja

3.2.2. Fizikai tulajdonság (MO_Property)

Az összes származtatott modelltípusnál a fizikai tulajdonságokat egy különleges osztály, az MO_Property írja le (**19. ábra**). Ez az inverzió révén meghatározott fizikai paraméterből és egy logikai kapcsolóból áll, ami megmutatja, hogy az inverzió során az adott paraméter értéke rögzített volt (igaz), vagy sem (hamis). Minden réteghez, cellához, vagy tetszőleges alakú modell komponenshez korlátlan számú fizikai tulajdonság rendelhető. Ez lehetővé teszi a különböző geofizikai módszerekre épülő együttes inverziók eredményének leírását.

3.2.3. Rétegmodell (MO_LayerModel)

A rétegmodell egymás felett elhelyezkedő, véges vastagságú, horizontálisan végtelen kiterjedésű rétegek leírására szolgál (**19. ábra**). Tetszőleges számú rétegből állhat, amelyeket a hozzájuk kapcsolt fizikai tulajdonságok mellett sorszám, fedő mélység és vastagság jellemez. Minden réteghez tartozik két logikai kapcsoló is, a rögzítések megadására. Az egyik a fedőmélység, a másik a vastagság paraméter rögzítettségére vonatkozik. Érdekességként megjegyzem, hogy a kommersz inverziós programok általában csak a rétegvastagságok rögzítését engedélyezik. Az ELGI-ben kifejlesztett inverziós program ezzel szemben a rétegek mélységének rögzítését is lehetővé teszi. Ez igen hasznos akkor, amikor fúrásokból származó mélységadatokkal kényszerített inverziót hajtunk végre (Prácser, 2008). A fedőmélység, és a vastagság opcionálisak. Az első rétegnél a fedőmélység, az utolsónál a vastagság megadása nem kötelező. Minden réteghez megadható egy címke is (tag), amely annak geológiai szerepére utal, és a rétegsorban elfoglalt helytől független azonosíthatóságot szolgálja. Mivel a GG_Parameter osztály szöveges adatok tárolását is lehetővé teszi, a rétegekhez sztratigráfiai, vagy petrofizikai elnevezések is kapcsolhatók. Ily módon a rétegmodell osztály minden további nélkül használható fúrási adatbázisok kialakítására is. A kőzet- és rétegtani kifejezések a kapcsolt paraméter katalógusban egzakt módon rögzíthetők.



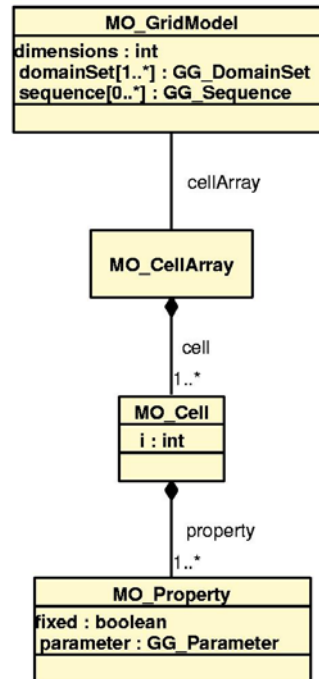
19. ábra Az MO_LayerModel osztály UML diagramja

A rétegmodell használatára példa a 2. függelék „Rétegmodell fúrési rétegsor” című részében található.

3.2.4. Rácsmodell (MO_GridModel)

A rácsmodell tetszőleges számú merőleges falú cella kompozíciója (20. ábra). A rétegmodellhez hasonlóan az MO_Property osztályt használva minden cellához több fizikai tulajdonságot is rendelhetünk, sőt, akár szöveges leírást is. A fizikai paramétereket ugyanúgy paraméter kódok kapcsolják a katalógusban lévő típus definíciókhoz. A többdimenziós rácsok kezelése analóg a többdimenziós terítési komponensekével. A cellák elhelyezkedését értéktartomány tömbök (domain set) adják meg. Minden térbeli, frekvencia-, vagy időtengelyhez tartozik egy-egy értéktartomány tömb. A cellák indexelése az értéktartományokat megadó tömbök sorrendjében történik. A térbeli pozicionálást az értéktartományokhoz tartozó paramétertípus definíciók teszik egyértelművé (középpont, alsó, felső határ koordináta stb). A cellákra vonatkozó egyéb információk (pl. cella méret), ha szükséges, további szekvenciákban adhatók meg. Egy tipikus 3 dimenziós rács a következő értéktartomány tömböket tartalmazza: A cella közepek helye a 3 térbeli tengely mentén (CELL_CENT_X, CELL_CENT_Y, CELL_CENT_Z), és 3 szekvencia a cellák méretével (CELL_SIZE_X, CELL_SIZE_Y, CELL_SIZE_Z). Ez utóbbi teljes térkitöltésnél redundáns

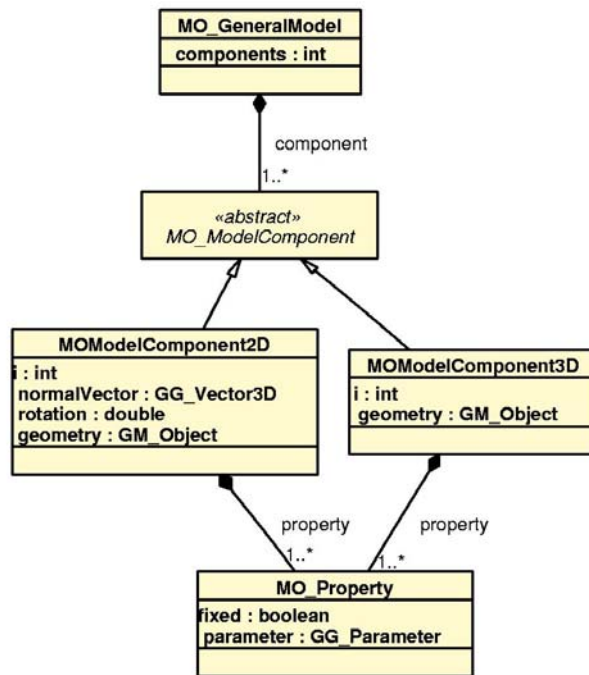
információ, mivel a cella közepekből a cellák mérete adódik. Tranziens modellek esetén a mintavételi idők meghatározásához egy 4. értéktartomány tömb is szükséges.



20. ábra Az MO_GridModel osztály UML diagramja

3.2.5. Általános geometriájú modell (MO_GeneralModel)

Az általános geometriájú modellek egymástól elkülönült, tetszőleges alakú, 2, 3 dimenziós homogén testekből épülnek fel (21. ábra). A rétegekhez, vagy modell cellákhoz hasonlóan fizikai tulajdonságok rendelhetők hozzájuk. E fizikai tulajdonságokkal bíró testeket modell komponenseknek nevezzük. Egy általános modell tetszőleges számú modell komponens kompozíciója. A 2D modellek 3D térben való elhelyezéséhez egy normálvektorra (a modell síkjára merőleges egységvektor) és egy elforgatási szögére is szükség van. Míg egy 2D modell komponenst geometriailag egy egyszerű polygon határoz meg, a 3D testek egzakt leírása meglehetősen komplikált. A modell komponensek geometriáját a GGDM-ben az ISO 19107 térbeli sémára épülő GML leírónyelvvel definiáljuk, ami igen bonyolult, akár időben változó alakú testek leírására is alkalmas. Az általános geometriájú modellek használatára jó példa, amikor gravitációs, vagy mágneses anomáliák kialakulását lehatárolt, szabályos testek hatásával közelítjük.



21. ábra Az MO_GeneralModel osztály UML diagramja

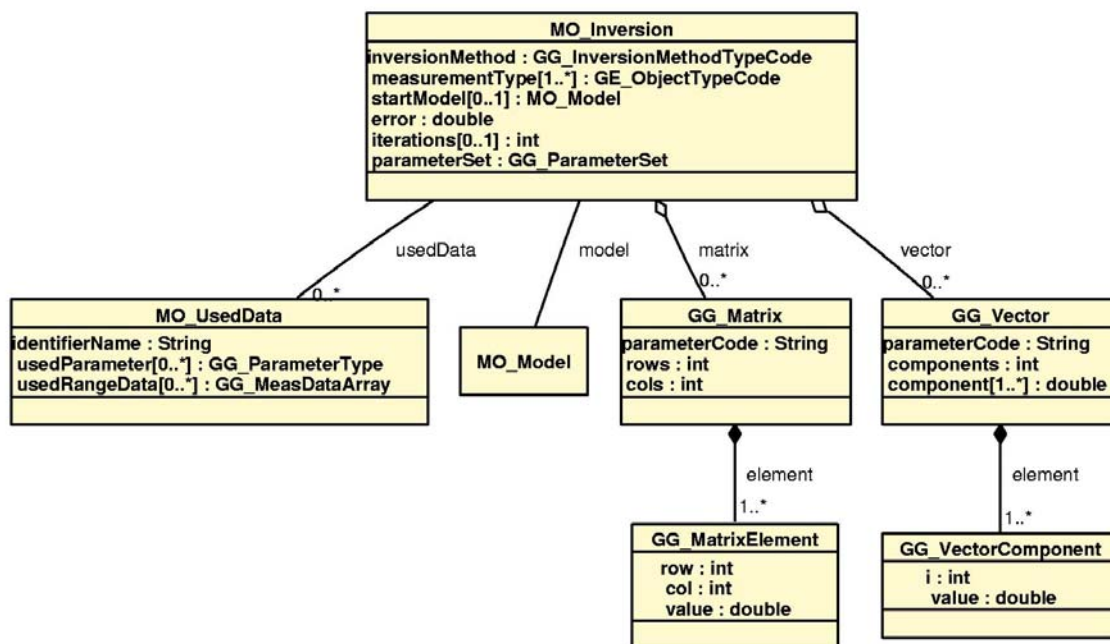
3.3. Geofizikai inverziók általános leírása

Inverzió egy optimalizációs eljárás, amely során a mérések földtani környezetét helyettesítő geofizikai modellt állítunk elő. A kapott modell felhasználásával számolható szintetikus adatok optimálisan illeszkednek a mérési adatrendszerhez. Ebben a megfogalmazásban a mérés, modell és inverzió egymástól határozottan elkülönülő fogalmak, mint ahogy a GGDM adatmodellben is azok. Az inverzió, mint entitás, az optimalizációs eljárás fontosabb technikai paramétereit és a mérésekkel, modellel meglévő kapcsolatait foglalja magába. A mérésekről és modellekről már volt szó, következzenek az inverzió adatmodelljének részletei.

3.3.1. Inverzió (MO_Inversion)

A geofizikában modellek manapság leggyakrabban inverzió eredményeként születnek. A mérések végfelhasználóját általában nem érdeklik az adatfeldolgozás részletei, csak az eredmény modell. A feldolgozást végző geofizikus számára azonban sokszor fontos, hogy az

inverzió folyamatát nyomon követhesse, az alkalmazott technikai paramétereket, részeredményeket, akár utólag is, szemügyre vegye. A legtöbb piaci inverziós szoftver a méréseket és az inverziós paramétereket együtt kezeli. Elég általános, dolog a mérés és inverzió között 1-1 értelmű megfeleltetést feltételezni. Sok esetben azonban egy méréshez több inverzió, és így több modell is tartozhat (1-N kapcsolat). A különböző típusú mérések együttes inverziójánál, hamar nyilvánvalóvá válik, hogy ez kevés. Egy inverzióhoz több mérés is tartozhat, tehát N-N kapcsolatról van szó. A GGDM adatmodellben egy méréshez bármennyi inverziót tárolhatunk (ezek lehetnek ekvivalens modellek, vagy a komplex modellalkotás különböző lépcsőfokai, amelyek tárolásra érdemesek), és egy inverzióban tetszőleges számú mérést használhatunk. A geofizikai inverziók elmélete, és a vele kapcsolatos fogalmak köre túlságosan szerteágazó ahhoz, hogy az adatmodell mindezt teljesen lefedje. Az adatmodellben csak a leismertebb technikák legfontosabb elemei vannak képviselve. Az MO_Inversion osztály (22. ábra) attribútumai között szerepel az inverzió típusa (least squares, simulated annealing, genetic algorithm), az iterációk száma, az illeszkedés hibája. A kiindulási modell tárolására is van mód. Különleges technikai paraméterek tárolására egy külön paraméterkészlet áll rendelkezésre. (A paraméter típusokat a kapcsolt katalógusban definiálni kell)



22. ábra Az MO_Inversion osztály UML diagramja

Az eredmény modell megadása mellett a felhasznált mérési terítések és paraméterek tárolására, vagy a rájuk való hivatkozásra egyaránt van mód (MO_UsedData). Az inverziók során gyakran keletkeznek mátrixok, vektorok, amelyeknek fontos technikai szerepük van, a mérések felbontóképességével, vagy a modell megbízhatóságával kapcsolatosak.

Ezeket a GG_Matrix és GG_Vector osztályok egyedeiben tárolhatjuk. A tárolt tömbadatok jelentése a megszokott módon, a paraméter kódok használatával tehető egyértelművé. A GG_Matrix sor és oszlop indexeket, valamint dupla pontos értékeket tároló mátrix elemek kompozíciója. A GG_Vector ezzel lényegében analóg, leszámítva, hogy komponensei a tömbelemek értékén kívül csak egyetlen indexet tartalmaznak. A struktúra alkalmas a leggyakrabban használt globális optimalizációs módszerek és a gradiens módszer inverziós paramétereinek tárolására. Különleges inverziós eljárás az ún. Általános Sorbafejtés módszere (Generalized Serial Expansion), ahol az inverzió során valamely fizikai paraméter viselkedését leíró függvénysor együtthatóit határozzuk meg. (Gyulai Á., Ormos T. 1998) Ilyenkor az eredmény modell tárolása a függvénysorból számolható modell paraméterekkel hagyományos módon történhet (pl. rétegmodellek sorozatával), a sorfejtés paramétereit pedig, az inverzióhoz kapcsolt vektorokban menthetjük el.

Az általános geofizikai adatmodell használatára XML példák a 2. függelékben és a <http://geomind.elgi.hu> címen található.

4. Az Általános Geofizikai Adatmodell (GGDM) implementációja

4.1. XML reprezentáció

XML az eXtensible Markup Language kifejezés rövidítéséből származik. A nyelv a 90-es évek végén terjedt el. Mára az internetes adatcsere és a strukturált rendszerek leírásának alapvető eszközévé vált. A ráépülő informatikai eszközök és technikák bemutatása kívül esik a dolgozat keretein. Rövid ismertetése és a geo-tudományok területén történő alkalmazásainak felvázolása a 3. számú függelékben található.

4.1.1. A GGDM séma definíciós csomag

A GGDM séma definíciós csomag az adatmodell szabatos technikai specifikációja, amely az UML adatmodell alapján készül és meghatározza, hogy a modell körébe tartozó objektumokat

name	definition	obligation	condition	dataType	max Occurrence	domainValue	parentEntity
HG_GeophObject	Geophysical object			aggregateClass			
dataProvider	URI of data provider	mandatory		class	1	URI	HG_GeophObject
identifierName	name of the object, uniq in the local namespace	mandatory		characterString	1		HG_GeophObject
parentIdentifier	identifier of parent metadata item	optional		characterString	1	varchar[50]	HG_GeophObject
groupIdentifier	identifier of higher level group	optional		characterString	N	varchar[50]	HG_GeophObject
objectType	type of this geophysical object	mandatory		class	1	GE_ObjectTypeCode	HG_GeophObject
geometry	geometry of the geophysical object stored in geometry type database field.	optional		class	1	GM_Object	HG_GeophObject
metadata	metadata about this object	mandatory		association	1	GG_Metadata	HG_GeophObject
archive	distributed archive	optional		association	1	HG_Archive	HG_GeophObject

7. táblázat. A GGDM séma definíciókat leíró táblázat részlete.

hogyan írjuk le XML-ben. A GGDM adatmodell szöveges dokumentációját eredetileg az ISO 19115 kiterjesztéseket leíró elem (MD_ExtendedElementInformation) mintájára készült táblázat tartalmazza. Az adatmodellel kapcsolatos változtatások ebbe a fájlba kerülnek be, és a séma definíciós XSD fájl is ez alapján készül (**7. táblázat**). A táblázat tartalmazza az adatmodell elemeinek nevét, típusát, számosságát, szülőjét, és egy rövid ismertető szöveget. Egy másik táblázatba kerülnek a sémában rögzített kódlisták, mint pl. a geofizikai módszerek nevei, vagy a szöveges fejléc paraméterek lehetséges értékei. Egy sémageneráló program ezekből az adatokból készíti el az XSD fájlt.

Az XSD fájl egy részlete jól szemlélteti a struktúra leírásának módját. Az áttekinthetőség kedvéért a magyarázó szövegeket (annotation) a kód nagy részéből eltávolítottam:

```
<xs:complexType name="HG_GeophObject_Type">
  <xs:annotation>
    <xs:documentation>Geophysical object</xs:documentation>
  </xs:annotation>
  <xs:sequence minOccurs="0">
    <xs:element name="dataProvider" type="xs:anyURI">
      </xs:element>
    <xs:element name="identifierName" type="xs:string">
      </xs:element>
    <xs:element name="parentIdentifier" type="xs:string" minOccurs="0">
      </xs:element>
    <xs:element name="groupIdentifier" type="xs:string" minOccurs="0"
maxOccurs="unbounded">
      </xs:element>
    <xs:element name="objectType" type="GE_ObjectTypeCode_Type">
      </xs:element>
    <xs:element name="geometry" type="gss:GM_Object_PropertyType" minOccurs="0">
      </xs:element>
    <xs:element name="metadata" type="GG_Metadata_Type">
      </xs:element>
    <xs:element name="archive" type="HG_Archive_Type" minOccurs="0">
      </xs:element>
    <xs:choice>
      <xs:element name="ME_Measurement" type="ME_Measurement_Type" minOccurs="0">
        </xs:element>
      <xs:element name="MO_Model" type="MO_Model_Type" minOccurs="0">
        </xs:element>
    </xs:choice>
  </xs:sequence>
  <xs:attribute name="uuid" type="xs:string"/>
</xs:complexType>
```

A geofizikai objektum definícióját egy „complexType” típusú elem tartalmazza. Ezen belül helyezkedik el a „sequence” elem, a beágyazott elemek sorozatával. Az elemek nevének, típusának és számosságának megadására az „element” szolgál. Az elemek típusa lehet egyszerű, és máshol definiált komplex típus. A „choice” lehetőséget ad alternatív szerkezetek definiálására. A példa alapján a geofizikai objektum lehet mérés (ME_Measurement), vagy modell (MO_Model). A definiált típushoz kapcsolható attribútumokat az „attribute” elem adja meg. A kódlistákhoz kapcsolt szövegek ellenőrzésére szolgál az alábbi kódrészlet:

```
<xs:simpleType name="GG_SeisSrcTypeCode_Type">
  <xs:restriction base="xs:string">
    <xs:enumeration value="vibrator"/>
    <xs:enumeration value="explosive"/>
    <xs:enumeration value="hammer"/>
    <xs:enumeration value="airgun"/>
  </xs:restriction>
</xs:simpleType>
```

A szeizmikus forrás típusát leíró paraméter értéke csak az „enumeration” elemek által megadott értékeket veheti fel. Az XSD fájlok egyszerű szintaxisával bármilyen bonyolult szerkezet leírható, és a sémának megfelelő XML adatok formai ellenőrzése (validáció) elvégezhető. A GGDM séma definíciók az Internetről szabadon letölthetők. (Sörös L. 2009b)

4.2. Az általános adatmodellre épülő geofizikai adatbázis (GGDB)

Az XDS séma definíciós csomag (schema definition package) egzakt módon meghatározza az XML formátumban tárolt geofizikai adatok szerkezetét. A sémákban leírtaktól való legkisebb eltérés is érvényesítési (validation) hibát eredményez. A felhasználó biztos lehet benne, hogy az érvényesítésen sikerrel átesett XML fájlok formailag tökéletesen megfelelnek az adatmodellnek. E szöveges állományok azonban elsősorban az adatcsere célját szolgálják, végleges adattárolásra, vagy adatbázisok építésére közvetlenül csak korlátozottan használhatók. Az érvényes XML fájlok halmaza nem tekinthető adatbázisnak, de bizonyos technikákkal megfelelő működés érhető el. A séma definíciós csomagban rögzített logikai szerkezetek bizonyos mértékben meghatározzák a fizikai megvalósítás módját, de a felhasználói igények és a fejlesztői kapacitás függvényében számos megoldás elképzelhető. Ebben a fejezetben azt vizsgálom meg, hogy a GGDM sémával érvényesített XML

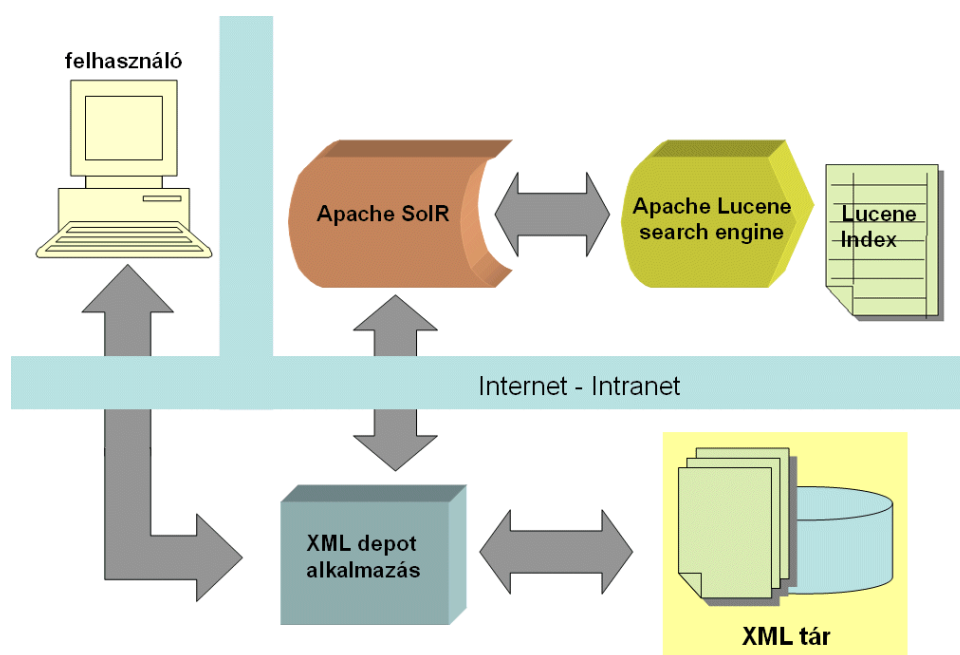
állományokból hogyan építhetők hatékony adatbázisok. Az adatbázisok komplexitását tekintve 3 szintet különíttek el, ennek megfelelően három különböző technikát mutatok be.

1. szint: XML dokumentumtár
2. szint: Hibrid adatbázis
3. szint: Relációs adatbázis

A három szint a megvalósítás oldaláról nézve különböző erőforrás igényű, a teljesítmény tekintetében szintén jelentősen különbözik egymástól.

4.2.1. XML dokumentumtár

Az XML dokumentumtár érvényesített XML fájlok halmazára épülő alkalmazás, amely egy dokumentumkereső motor (Apache Lucene search engine) szolgáltatásainak kihasználásával hatékony hozzáférést biztosít. Az XML fájlok tetszőleges rendszerben tárolhatók, akár hagyományos könyvtárszerkezetben, akár erre a célra kialakított egyszerű relációs adatbázisban. A fájlok gyors megtalálását egy speciális indexállomány teszi lehetővé. A rendszer négy fő komponensből áll (23. ábra):



23. ábra. Az „XML dokumentumtár” alkalmazás működésének logikai vázlata

1. Az XML tár, amely magukat a geofizikai adatokat (XML fájlok) tartalmazza valamilyen rendezett formában.
2. Keresőmotor, amely az XML tár elemeihez való gyors hozzáférés érdekében egy indexállományt hoz létre, és a keresési parancsokra az elérési útvonallal válaszol. Erre a célra jól megfelel a nyílt forráskódú Apache-Lucene keresőmotor (search engine), amelyet „Google” típusú szöveges keresésekre fejlesztettek ki. (Apache Lucene, 2009)
3. Az Lucene-nel való kommunikációt az Apache-Solr kereső szerver biztosítja (Apache Solr, 2009).
4. Az XML dokumentumtár alkalmazás felelős az XML fájlok fogadásáért, az indexelési parancsok kiadásáért, a hozzáférési útvonalak lekérdezéséért, a keresett XML fájlok alkalmas formában való megjelenítéséért.

A felhasználó az XML dokumentumtár alkalmazással kommunikál. Az adattárolás menete a következő: Az érvényesített geofizikai adatokat az alkalmazás elhelyezi az XML tárból. A tárolt fájlból XSL konverzió segítségével előállt indexelési parancsokat a Solr fogadja, és feldolgozás céljából továbbítja a Lucene-nek. A Lucene a tárolandó rekord elérési útvonalára, és a megadott mezőkre elvégzi az indexelést. Adatlekérés során az XML dokumentumtár elküldi a kereső kifejezést a Solr-nak. A Solr a Lucene-nel elvégzetteti a keresést, és a megkapott elérési útvonalat visszaküldi. Az XML dokumentumtár alkalmazás az elérési útvonal birtokában kiolvassa a kért rekordot az XML tárból, majd továbbítja a felhasználónak. A Lucene indexelési stratégiája a lekérdezéseknél nagy fokú rugalmasságot és gyors elérést biztosít. A lekérdezések formáját a Lucene keresőnyelv (Lucene search syntax) szabályozza.

Illusztrációképpen néhány egyszerű példán bemutatom, hogy különösebb programozás nélkül, pusztán néhány Unix parancs összeláncolásával miképpen használható az XML depó elvére épülő adatbázis.

A Solr update nevű modulja HTTP csatornán keresztül az alábbihoz hasonló XML állományokat vár:

```
<?xml version="1.0"?>
<add xmlns:fo="http://www.w3.org/1999/XSL/Format">
  <doc>
    <field name="fileName_t">example/cpt/116.xml</field>
```

```

<field name="id">GO_CPT_BAL-0117</field>
<field name="pid">GOS_CPT_ELGI</field>
<field name="objectType_t">boreholeLog</field>
<field name="measDate_t">1986-05-27</field>
<field name="purpose_t">civil engineering,</field>
</doc>
</add>

```

A szöveg feldolgozása során a Lucene index állomány a megadott mezőnevekhez kapcsolt értékekkel bővül. Ez az XML tárból lévő fájl elérési útvonalát (example/cpt/116.xml) összekapcsolja a fájlban lévő azonosító (id), szülő azonosító (pid), objektumtípus (objectType_t) stb. paraméterekkel. Ilyen indexelő parancsfájlok előállítását könnyen elintézhető egy egyszerű XSL transzformációval:

```
xml tr Solr-ggdb.xsl -s fileName="example/cpt/116.xml" example/cpt/116.xml
```

Az XML program az xmlStarlet nevű csomagban található, amely XML fájlok parancssoros feldolgozására használható igen hatékony eszköz. A tr parancs első paramétere az XSL fájl neve, a második a forrás XML. A Solr-ggdb.xsl fájl tartalma pl. a következő:

```

<?xml version="1.0" encoding="ISO-8859-2"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:fo="http://www.w3.org/1999/XSL/Format">
<xsl:output method="xml" indent="yes"/>
<xsl:param name="fileName"/>
<xsl:template match="/">
  <add>
    <doc>
      <field name="fileName_t"><xsl:value-of select="$fileName"/></field>
      <field name="id"><xsl:value-of select="//identifierName"/></field>
      <field name="pid"><xsl:value-of select="//parentIdentifier"/></field>
      <field name="objectType_t"><xsl:value-of select="//objectType"/></field>
      <field name="measDate_t"><xsl:value-of
select="//date[dateType='measurement']/theDate"/></field>
      <field name="purpose_t"><xsl:value-of select="//purpose/value"/></field>
    </doc>
  </add>
</xsl:template>
</xsl:stylesheet>

```

A transzformáció az xpath kifejezések segítségével egyszerűen kiszűri a megfelelő értékeket a forrás XML fájlból, és elhelyezi a kimeneti állományba. Az így előállt XML fájlt a

parancssoros HTTP kommunikációra használható curl program segítségével továbbítjuk a Solr kereső szervernek a következő módon:

```
curl http://localhost:8938/Solr/update --data-binary @tmp.xml -H 'Content-type:text/xml; charset=utf-8'
```

A curl a localhost 8938-as portján figyelő „update” nevű servletnek elküldi a megadott fájlban (tmp.xml) lévő szöveget, ami az XSL konverzió kimeneteként előállt indexelő állomány. A változtatások véglegesítése céljából ki kell adni a “commit” parancsot is:

```
curl $URL --data-binary '<commit/>' -H 'Content-type:text/xml; charset=utf-8'
```

Több fájl egyidejű küldésére alkalmas az alábbi rövid kis shell script, a post.sh:

```
FILES=$*
URL=http://localhost:8938/Solr/update

for f in $FILES; do
    echo Posting file $f to $URL
    curl $URL --data-binary @$f -H 'Content-type:text/xml; charset=utf-8'
    echo
done

curl $URL --data-binary '<commit/>' -H 'Content-type:text/xml; charset=utf-8'
echo
```

A leírt műveleteket foglalja össze a két soros mkindex.sh script:

```
xml tr Solr-ggdb.xsl -s fileName=$1 $1 > tmp.xml
./post.sh tmp.xml
```

A bemutatott példában az XML tár az „example” nevű könyvtár, ahol VES, TDEM és mérnökgeofizikai szondázási adatok (Cone Penetration Test, CPT) találhatóak. A három könyvtár összes állományának indexelése a következő három paranccsal elvégezhető:

```
ls example/tDEM/*.xml | xargs -n 1 ./mkindex.sh
ls example/VES/*.xml | xargs -n 1 ./mkindex.sh
ls example/CPT/*.xml | xargs -n 1 ./mkindex.sh
```

Az ls parancs által szolgáltatott fájlnev lista minden elemével lefut a mkindex.sh program, elvégezve az indexelést a három könyvtárban lévő összes XML fájlra. Ezek után alkalmasan

választott Lucene kifejezésekkel keresgélhetünk az indexállományban. A kéréseket a select nevű servlet-nek küldjük. Néhány egyszerű példa:

1. Keressük azt a mérést, amelyeket 1985. szeptember 24-én mértek. A válasz megkapjuk, ha a WEB böngészőbe beírjuk a következő címet:

```
http://localhost:8983/Solr/select/?q=measDate_t: 1985-09-24
```

A kérésben a „q=” után szerepel a Lucene kereső kifejezés, melyre a válasz egy XML fájl, ami egy fejléc kíséretében a találati rekord adatait tartalmazza:

```
<?xml version="1.0" encoding="UTF-8"?>
<response>

<lst name="responseHeader">
  <int name="status">0</int>
  <int name="QTime">1</int>
  <lst name="params">
    <str name="indent">on</str>
    <str name="start">0</str>
    <str name="q">measDate_t: 1986-05-27</str>
    <str name="version">2.2</str>
    <str name="rows">10</str>
  </lst>
</lst>
<result name="response" numFound="1" start="0">
  <doc>
    <str name="fileName_t">example/cpt/116.xml</str>
    <str name="id">GO_CPT_BAL-0117</str>
    <str name="measDate_t">1986-05-27</str>
    <str name="objectType_t">boreholeLog</str>
    <str name="pid">GOS_CPT_ELGI</str>
    <int name="popularity">0</int>
    <str name="purpose_t">civil engineering,</str>
    <str name="sku">GO_CPT_BAL-0117</str>
    <date name="timestamp">2009-08-31T16:47:31.452Z</date>
  </doc>
</result>
</response>
```

2. Kérjük a boreholeLog objektumtípusba tartozó fájlok elérési útvonalát:

```
http://localhost:8983/Solr/select?indent=on&version=2.2&q=objectType_t:
boreholeLog&fl=fileName_t
```

A válasz (a fejléc nélkül):

```
<?xml version="1.0" encoding="UTF-8"?>
<response>

<result name="response" numFound="42" start="0">
  <doc>
    <str name="fileName_t">example/cpt/24.xml</str>
  </doc>
  <doc>
    <str name="fileName_t">example/cpt/2.xml</str>
  </doc>
  <doc>
    <str name="fileName_t">example/cpt/81.xml</str>
  </doc>
  ...

  <doc>
    <str name="fileName_t">example/cpt/113.xml</str>
  </doc>
  <doc>
    <str name="fileName_t">example/cpt/114.xml</str>
  </doc>
  <doc>
    <str name="fileName_t">example/cpt/115.xml</str>
  </doc>
  <doc>
    <str name="fileName_t">example/cpt/116.xml</str>
  </doc>
</result>
</response>
```

Amint látható, a mezőnév megadásával (fl=fileName_t) kiszűrhetjük a számunkra nem érdekes információkat. A keresés a curl segítségével parancssorból is elvégezhető. Az előbb már látott xml program sel parancsa segítségével a válasz XML-ből kiirathatjuk a fájlnevek listáját.

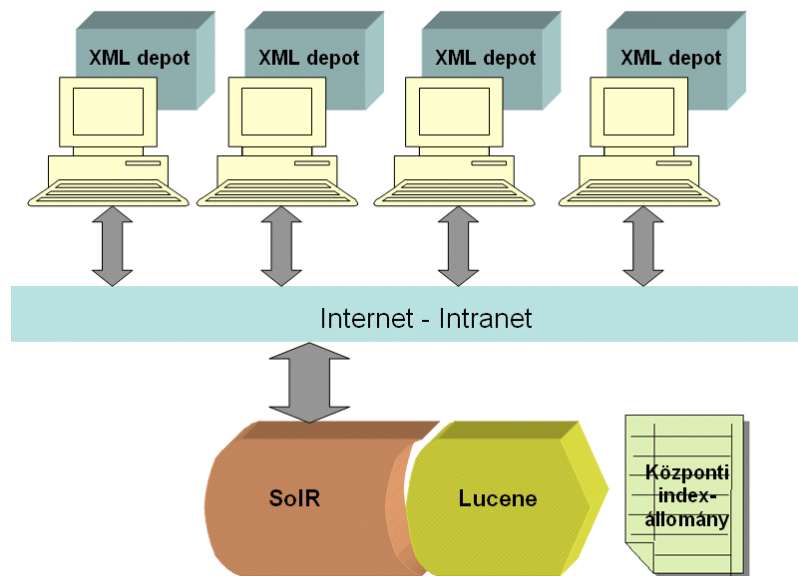
```
curl "http://localhost:8983/solr/select?q=objectType_t:boreholeLog&rows=10&l=fileName_t" | xml
sel -t -m "//str[@name='fileName_t']" -v "." -n
```

az eredmény:

```
example/cpt/24.xml
example/cpt/2.xml
```

example/cpt/81.xml
example/cpt/113.xml
example/cpt/114.xml
example/cpt/115.xml
example/cpt/116.xml

A rendszer nagy előnye, hogy a vezérlő alkalmazást leszámítva nyílt forráskódú, platform független csomagokra épül. Egy viszonylag fejlett XML dokumentumtár alkalmazás megírása is egyszerű feladat, mivel az intelligens kiszolgáló modulok révén szinte kizárólag szövegfeldolgozásra (XSL transzformációk) és HTTP kommunikációra van szükség. A modulok közötti hálózati kapcsolat lehetővé teszi osztott rendszerek kiépítését is. Egy központi Solr-Lucene blokk indexállománya tetszőleges számú független XML dokumentumtár alkalmazás és XML tár kiszolgálására használható (24. ábra).



24. ábra Központi Lucene indexállományra épülő osztott adatbázis

Az XML dokumentumtár elven felépülő rendszerek hátránya, hogy a lekérdezés fájl szintű, vagyis, a bevitelnél egy XML fájlba kódolt információ egyszerre érhető el. Ez viszonylag kis méretű fájlok esetén egyáltalán nem okoz problémát. Nagy adattömeget tartalmazó méréseknél az XML fájlok kezelése egy határon túl nehezkessé válik. A rendszer jól alkalmazható pontszerű mérések és egyszerűbb szondázások (VESZ, TDEM) adatainak

tárolására. Nem javasolt hosszú mérési szelvényekből álló légi geofizikai méréseknél, mélyfúrás-geofizikai vagy, szeizmikus méréseknél.

A leírt XML dokumentumtár rendszerhez hasonló elven működik a KINGA portál adatbázisa. A különbség az, hogy a vezérlő alkalmazás nem a Solr-on keresztül, hanem közvetlenül kapcsolódik a Lucene keresőmotorhoz, az XML fájlok tárolása pedig az InterComp kft. által kifejlesztett, relációs adatbázisra épülő rendszerben, az ún. dokumentum szerveren történik.

4.2.2. (Tisztán) Relációs adatbázis

Az XML technológiát támogató modern fejlesztőrendszerek egyre magasabb szinten támogatják a tervezést és a rendszerépítést. Az XSD séma felhasználásával lehetőség van olyan kód generálására, amely különösebb fejlesztői munka nélkül szolgáltatja az XML fájlok ki és beolvasáshoz, és a sémában leírt szerkezetek memória struktúrákba való leképezéséhez használható eszközöket. Erre alkalmas pl. a SUN Microsystems által fejlesztett JBIND java csomag, amely az XSD séma alapján automatikusan elkészíti az XML adatok beolvasására (marshalling), kiíratására (unmarshalling) és a memóriában tárolt objektumok kezelésére szolgáló kódot.

Hasonló módon lehetséges olyan kódot generálni, amely az XSD sémában leírt struktúrák összes elemét azok kapcsolataival együtt relációs adatbázisba képezi le, és előállítja a táblák előállításához és kezeléséhez szükséges összes programmodult is. Az ily módon generált függvény könyvtárakból bonyolult adatszerkezet esetén is viszonylag gyorsan felépíthető egy stabilan működő rendszer. Példaként a szintén java alapú Hibernate rendszer említhető.

Az ilyen megoldások hátránya, hogy a generált kód nehezen áttekinthető, és ha valahol mégis változtatásra, vagy speciális igények kielégítésére van szükség, az általában aránytalanul sok fejlesztői ráfordítást igényel. Gyakori ellenérv az séma alapján generált rendszerek ellen, hogy az általánosság a hatékonyság rovására megy. Nagy adattömeg esetén a működés az optimalizáció hiánya miatt jelentősen lelassul.

A GGDM adatmodell komplexitása miatt a teljes struktúra relációs adatbázisba való automatikus leképezése feltehetően több száz táblát eredményezne. Egy ekkora rendszer kézbentartása túlságosan komplikált feladatnak tűnik, de mégsem ez a fő ok, amiért

véleményem szerint ez a megoldás nem használható. Az egymásba láncolt relációk hosszú sora oly mértékben lassítja le az adatokhoz való hozzáférést, ami már nem elviselhető.

Strukturális szempontból is abszurdnak tűnik egy olyan implementáció, ahol pl. az adatterjesztésre vonatkozó részletező adatok 10-15 táblába vannak lebontva, miközben egy viszonylag nagy rendszerben is ezekben a tábláknak csupán néhány rekord foglalna helyet. Különösen az ISO 19115 adatmodell esetén gyakoriak a mély struktúrák, amelyek kis egyedszámmal jellemezhetőek.

4.2.3. Hibrid adatbázis

A hibrid adatbázis egyfajta kompromisszumot valósít meg a két szélsőséges megoldás között. Az egyik szélsőség az XML dokumentumtár stratégiája: mindent XML fájlban tárolunk. A másik a tisztán relációs adatbázis, ahol mély XML struktúrák egészét relációs táblákba képezzük le. A GGDM adatmodell használata közben szaporodó tapasztalatok kirajzolják azon elemeknek a körét, amelyeknél az elkülönített tárolás és hozzáférés valóban fontos. A hibrid adatbázis alapgondolata a következő: A gyakran keresett, vagy önállóan használt elemek kerüljenek relációs táblákba, míg a ritkábban használt mély struktúrákat tartsuk együtt XML formátumban. Az XML adatok tárolása persze szintén relációs adatbázisban, XML típusú mezőkben történik. A modern adatbázis-kezelő rendszerek megfelelő eszközöket biztosítanak az XML típusú mezőkben tárolt szövegek feldolgozásához. Pl. a PostgreSQL rendszer SQL parancsokban lehetővé teszi az XPATH kereséseket és az XSL konverziókat is. Ez a megoldás az XML szövegeket tökéletesen integrálja az adatbázisba.

4.2.3.1. XML és a hagyományos relációs adatbázis

A relációs adatbázisok elméletének és gyakorlatának ismertetése nem célom. A téma szakirodalma igen kiterjedt (Halassy B 2002, Timár et al. 1996, Timár et al. 1997). A hibrid adatbázis tervezésekor a legfontosabb kérdés az, hogy az XML fájlokat milyen mértékben szedjük szét. Mint láttuk, a két véglet az XML depó, és a tisztán relációs adatbázis. A kettő között sok lehetséges megoldás van. Ezek létjogosultságát az adott rendszerrel szemben támasztott követelmények szabják meg. A hibrid adatbázis egy lehetséges megvalósítása az

ELGI-ben kifejlesztett kísérleti rendszer. Terveim szerint ez egy általános célú adatbázis, amely nem csak a mérési adatok archiválását, hanem a feldolgozó és értelmező munkát, vagyis a mérési adatok, paraméterek szintjén történő hozzáférést is támogatja.

A rendszer felépítését a **25. ábra** látható egyed-reláció diagram mutatja. A 27 táblából álló adatbázis lefedi a teljes GGDM adatmodellt. Az ábrán minden egyes négyszög egy egyedet (entitást) ábrázol, amely fizikailag egy táblázatnak felel meg. Az első sor az egyed nevét, a többi az egyed tulajdonságait (attribútumait) mutatja. Ezek a táblázat oszlopai. A négyszögeket összekötő vonalak az egyedek közötti kapcsolatokat jelzik, amelyeket az FK-val (foreign key) jelölt idegen kulcsok valósítanak meg. A kapcsolatokkal nem rendelkező egyedek elérése az XML mezőkben tárolt kulcsokon keresztül történik. A táblák az adatmodell főbb objektumaihoz kapcsolódnak a következő rend szerint:

metaadat egyedek: *go* – *geophObject*, *gos* – *geophObjectSet*, *rep* – *report*, *md* – *metadata*, *pcat* – *parameterCatalogue*, *party* – *responsibleParty*, *instr* – *instrument*, *dev* – *device*

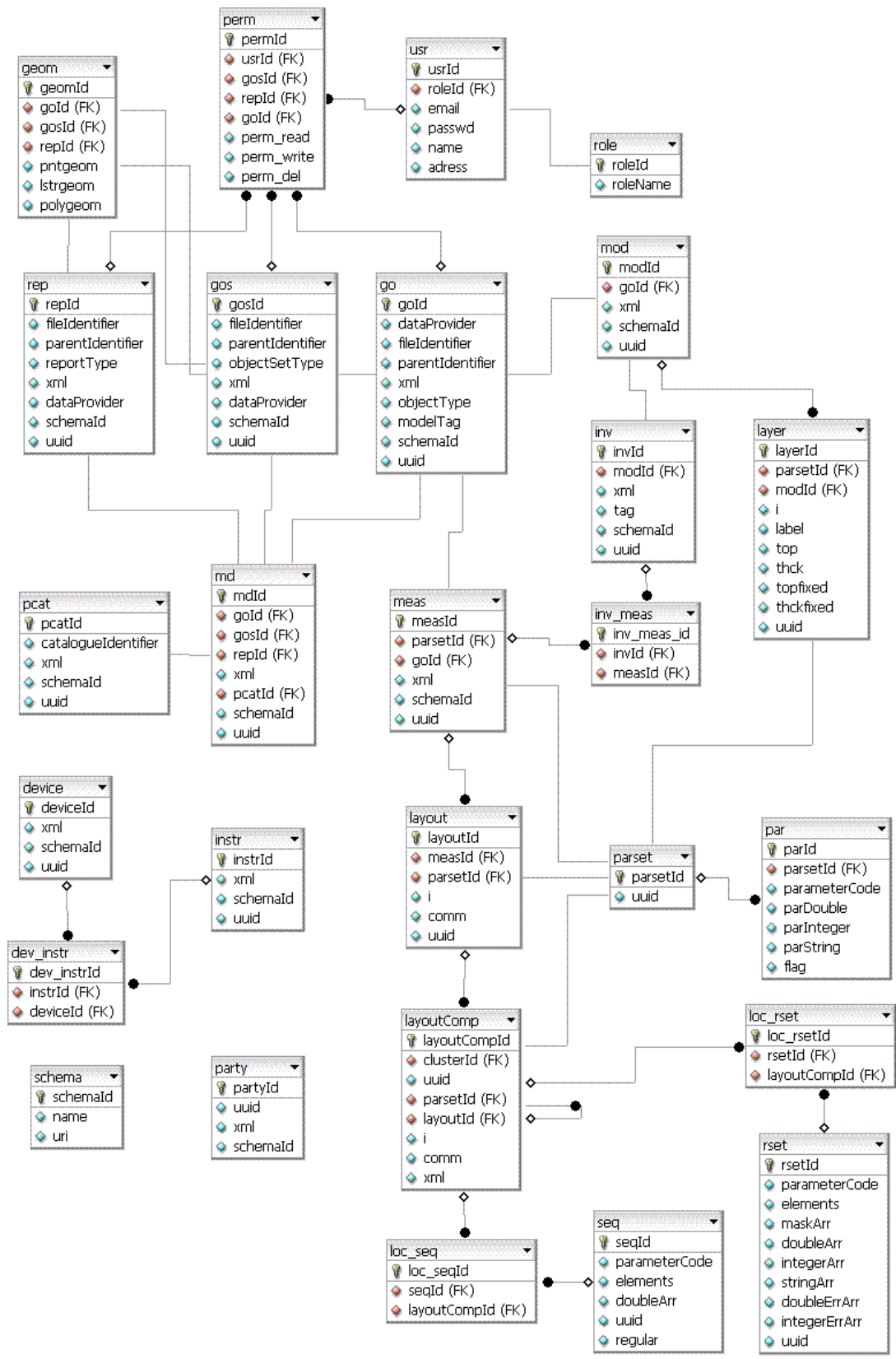
mérési adat és modell egyedek: *meas* – *measurement*, *layoutcomp* – *layoutComponent*, *rset* – *rangeSet*, *seq* – *sequence*, *parset* – *parameterSet*, *par* – *parameter*, *mod* – *model*, *lay* – *layer*, *inv* – *inversion*

A rendszerfunkciók kiszolgálása érdekében további táblákra is szükség van. Térinformatikai keresések támogatása: *geom*

Segéd táblák az N:M kapcsolatokhoz: *inv_meas*, *loc_rset*, *loc_seq*, *dev_instr*

A felhasználó azonosítást segítik a következő táblák: *perm* – *permission*, *usr* – *user*, *role*

Látható, hogy a fő táblák csak a legfontosabb attribútumokat tartalmazzák külön adatbázismezőként. Az adatok jelentős része XML mezőkben van tárolva. A közvetlenül elérhető hagyományos adatmezők, és az XML-be ágyazott, xpath mechanizmussal elérhető mezők hozzáférési ideje jelentősen eltérő. Különösen nagy rekordszámok esetén az xpath alapú keresések megengedhetetlenül lassúak lehetnek.



25. ábra Az ELGI-ben kifejlesztett kísérleti adatbázis egyed-relációs modellje

A sima vonalak 1:1 kapcsolatot, a ponttal és gyémánttal jelöltek 1:N kapcsolatokat jelentenek. (A gyémánt a kapcsolat 1-es, a pont az N oldali vége.)

A gyakran használt adatok indexelt mezőkbe helyezése a hatékonyság miatt nagyon fontos. Ez bizonyos fokú redundanciát jelent, hiszen ugyanaz az adat különböző helyen két példányban van tárolva. Bár ez sérti a relációs adatbázisok elméletének alapelveit, a hatékonyság érdekében tett hasonló kompromisszumok a gyakorlatban előfordulnak. Ez a probléma a hibrid megoldás legnagyobb hátrányának tekinthető. Alapvető fontosságú, hogy az adat integritás érdekében az adatkezelő tranzakcióknak biztosítani kell az XML-ben tárolt, és az indexelt mezők egyezését.

Faktorizáció

A felesleges redundanciák elkerülése céljából a XML-ek nem változatlan formában kerülnek tárolásra, hanem egy speciális átalakítás után, amit faktorizációnak neveztem el. A faktorizáció az XML szerkezet fő elemeire bontását jelenti. Azon elemeket, amelyek külön táblákban lesznek tárolva, kiemeljük a szülő elemből, és a helyükön csak hivatkozásokat hagyunk. A kiemelt XML fragmentum szintén tartalmazhat kiemelendő elemeket, ezért a faktorizációt rekurzív módon kell végrehajtani. Ezután az XML fragmentumokból kiemelt mezőkkel együtt az adatrekordok inzertálhatók az adatbázisba.

A folyamatot szemlélteti a következő példa:

Egy gravitációs mérés, mint geofizikai objektum (geophObject) tartalmaz egy „metadata” és egy „ME_Measurement” elemet, amelyek külön táblába kerülnek. Faktorizáció után a go tábla XML mezője a következő szöveget tartalmazza:

```
<geophObject>
  <dataProvider>http://www.elgi.hu</dataProvider>
  <identifierName>GO_GRAV_13630</identifierName>
  <parentIdentifier>GOS_GRAV_pecs001</parentIdentifier>
  <objectType>gravimetry</objectType>
  <metadata uuid="MD_elgi.hu_GO_GRAV_13630"/>
  <ME_Measurement uuid="ME_elgi.hu_GO_GRAV_13630"/>
</geophObject>
```

Az üres XML elemek a vastag betűvel szedett azonosítókkal (uuid) hivatkoznak az md tábla és a meas tábla megfelelő rekordjaira. A hivatkozott ME_Measurement elem faktorizáció után a következő alakú:

```
<ME_Measurement uuid="ME_elgi.hu_GO_GRAV_13630">
```

```

<identifierName>GO_GRAV_13630</identifierName>
<localCRS>
  <origin>
    <srsName>EPSG:23700</srsName>
    <GG_CartesianCRS>
      <x>565496</x>
      <y>63791.5</y>
    </GG_CartesianCRS>
  </origin>
  <origin>
    <srsName>EPSG:4326</srsName>
    <GG_SphericCRS>
      <latitude>45.91349</latitude>
      <longitude>17.95843</longitude>
    </GG_SphericCRS>
  </origin>
  <elevation>
    <verticalDatum>BalticSea</verticalDatum>
    <elevation>100.64</elevation>
  </elevation>
</localCRS>
<parameterSet uuid="PS_1_elgi.hu_GO_GRAV_13630"/>
</ME_Measurement>

```

A parameterSet elem faktorizációja nem eredményez további XML fragmentumot, annak minden eleme relációs táblákba kerül (parset és par táblák). Az elemet leíró szöveg a következő:

```

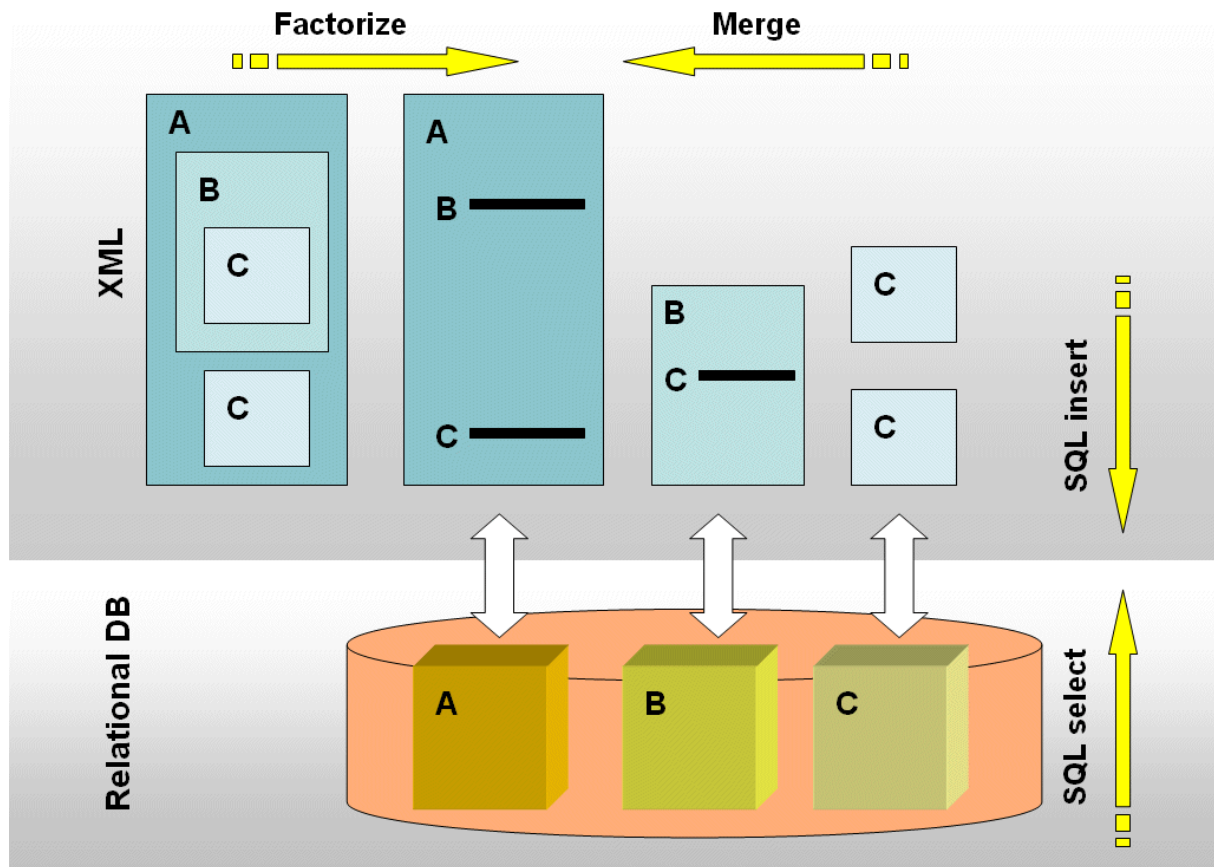
<parameterSet uuid="PS_1_elgi.hu_GO_GRAV_13630">
  <parameter>
    <parameterCode>PNID</parameterCode>
    <GG_ParameterInteger>
      <value>13630</value>
    </GG_ParameterInteger>
  </parameter>
  <parameter>
    <parameterCode>G</parameterCode>
    <GG_ParameterDouble>
      <value>684270.0</value>
    </GG_ParameterDouble>
  </parameter>
  <parameter>
    <parameterCode>K</parameterCode>
    <GG_ParameterDouble>
      <value>2.0</value>
    </GG_ParameterDouble>
  </parameter>
  <parameter>
    <parameterCode>TDOLL</parameterCode>

```

```

    <GG_ParameterString>
        <value>pecs001</value>
    </GG_ParameterString>
</parameter>
</parameterSet>

```



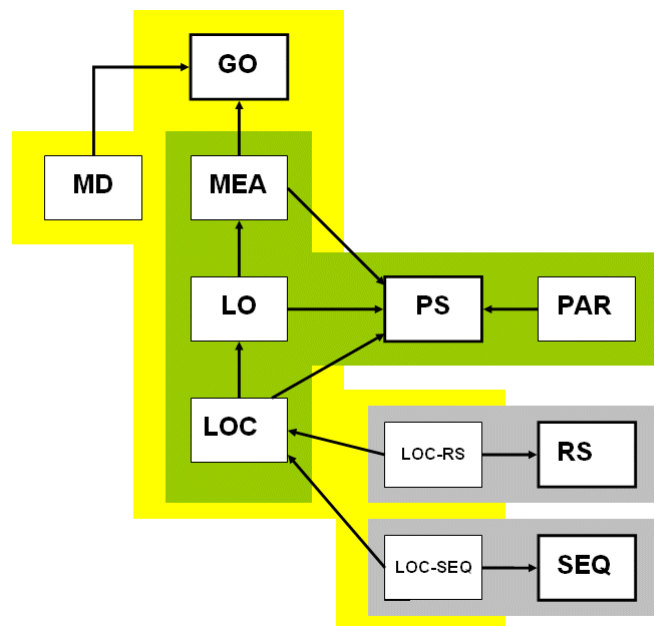
26. ábra XML és SQL műveletek a hibrid adatbázisban

Az xml fájlok relációs táblákban történő tárolása előtt faktorizáció történik. Az xml-el kiolvasásakor ezzel ellentétes folyamatra, összeolvasztásra (merge) van szükség

Az adatok kinyerésekor a faktorizációval szétdarabolt XML-ek összeállítására ellentétes irányú folyamat. Az összeolvasztás (merge) során az adatkezelő a relációs táblákból kiolvasott adatokból XML fragmentumokat készít, majd a szülő objektum uuid kódokat tartalmazó üres elemek helyére beilleszti azokat. A rekurziós láncon visszafelé haladva előáll az eredeti XML szöveg rekonstrukciója.

4.2.3.2. Adatrekordok törlése

Az adatrekordok törlésekor az adatbázis integritásának megőrzése érdekében körültekintően kell eljárni. A gondatlanul végrehajtott törlések során elérhetetlen adatok – adatbázis szemét – maradhatnak hátra. A táblák definiálása során megadhatjuk, hogy a rendszer miként kezelje az egymással kapcsolatban álló táblákra vonatkozó törlési parancsokat. Előírhatjuk pl., hogy egy szülő rekord törlése vonja maga után a gyermek rekordok törlését is. (Ez a CREATE TABLE parancsoknál az ON DELETE CASCADE záradék megadásával érhető el) Ez nagy terhet vesz le az alkalmazás programozó vállaról, de ez még nem elég. Az adattáblák bonyolult kapcsolatrendszere miatt a különböző rekordok törlésének sorrendjére is figyelni kell. A **27. ábra** a kaszkád törlések hatókörét mutatja. Az alapszabály az, hogy egy adott tábla adott rekordjának törlése a rámutató táblák megfelelő rekordjait is eltünteti. Egy geofizikai objektum törlésekor pl., a sárga polygon területén belül lévő táblák megfelelő rekordjai automatikusan törlődnek. Ha ezt megtennénk, a parset (PS), rset (RS) és seq (SEQ) táblák hivatkozott rekordjai szülő nélkül maradnának. Soha többé nem tudnánk megmondani, hogy az árván maradt paraméterek, értéktartományok, értékészletek melyik objektumhoz tartoztak. (orphan records) Ennek elkerülése érdekében a következő módon kell eljárni:



27. ábra Az idegen kulcsok és a kaszkád törlések hatóköre

A vastag kerettel jelölt táblázatok rekordjainak törlése a színezett sávba eső táblák rekordjainak törlését vonja maga után. A rekordok helyes törlési sorrendje: RS, SEQ, PS, GO

Először töröljük a kiszemelt geofizikai objektumhoz tartozó rset (RS) és seq (SEQ) rekordokat, ami magával vonja a kék mezőkbe eső köztes táblák (LOC_RS, LOC_SEQ) rekordjainak törését. Ezután töröljük a pset (PS) rekordokat, ami egyben a zöld mezőbe eső meas (MEA), layout (LO), és layoutcomp (LOC) rekordok törlését is jelenti. Végül töröljük a geofizikai objektum (GO) rekordot, ami magával viszi a metaadat rekordot is. Ezzel a módszerrel a törölt objektum után nem marad hátra adatbázis szemét.

4.2.3.3. Az informatikai alrendszerek áttekintése

Az ELGI-ben kifejlesztett GGDM alapú hibrid adatbázisrendszer kizárólag nyílt forráskódú, professzionális szoftverre és saját fejlesztésű java programokra épül. Legfontosabb eleme a postgresQL adatbázis kezelő és a térinformatikai támogatást biztosító postGIS függvénykönyvtár. A postGIS használata lehetővé teszi az adatbázis téradatainak OGC szabványos WMS és WFS szolgáltatásokba történő integrálását. Ezt a funkciót a Geoserver látja el, amely remekül együttműködik a postgresQL rendszerrel. Az adatkezelést és a felhasználói funkciókat magába foglaló keretrendszert saját fejlesztésű java programok adják.

postgresQL

A postgresQL megelőzve a korábbi élvonalas MySql-t mára a szabad szoftver világ legnépszerűbb adatbáziskezelő rendszere lett. Szolgáltatásai és támogatottsága vetekszik a profi világ vezető rendszerével az ORACLE-lel. Támogatja az adatbázis tranzakciókat, tárolt procedúrákat, a geometriai és XML adatmezők magas szintű kezelését, és a legmodernebb SQL szabványnak megfelelő funkciókat.

postGIS

A postGIS a postgresQL rendszerhez kapcsolódva komplex térinformatikai funkciókat valósít meg. Valójában egy C nyelven írt függvénykönyvtár, amelynek elemei a postgresQL rendszerből SQL szinten elérhetők.

Geoserver

A Geoserver egy WEB-es térinformatikai műveletek támogatására kifejlesztett, java szervletekből felépülő csomag. A Geoserver projekt fő célja az OGC szabványos WEB szolgáltatások magas szintű támogatása. Segítségével különböző adatforrások, többek közt shape fájlok, MySql és postgresQL adatbázis táblák integrálhatók a rendszer által nyújtott Web Map Service (WMS) és Web Feature Service (WFS) szolgáltatásokba. Ezek az Internetes téradat infrastruktúra gerincét képezik, és többek közt az INSPIRE fejlesztések homlokterében állnak.

Java

A java a SUN Microsystems védőszárnyai alatt kibontakozó, széles körben használt objektumorientált programnyelv. A platform független rendszerek, így a WEB alkalmazások fejlesztésének igen hatékony eszköze, amely támogatja a legmodernebb informatikai technológiákat. A java csomagok révén a számítástechnika legkülönbözőbb területeihez kapcsolódó magas szintű függvény könyvtárak állnak a fejlesztők rendelkezésére. Ezek közül számunkra az adatbázis kezelést segítő JDBC (Java Database Connectivity) csomag, az XML támogatást megvalósító DOM, SAX, és a WEB alkalmazásokat támogató Java™ Platform, Enterprise Edition (Java EE) rendszer a legfontosabbak. Az ELGI geofizikai adatbázisrendszere a szintén SUN által támogatott NetBeans fejlesztő rendszer használatával készül.

4.2.3.4. Saját fejlesztésű adatbáziskezelő eszközök

A hibrid GGDM adatbázisból minden információ kinyerhető a PostgreSQL szabványnak megfelelő SQL lekérdezésekkel. Ehhez a szöveges psql kliens program, vagy a phpPgAdmin nevű WEB-es alkalmazás egyaránt használható. A hibrid adatbázis szerkezetét és az SQL-t kevésbé ismerő felhasználók segítésére elkészítettem néhány alkalmazást, amely megkönnyíti az adatbázis használatát. Az itt bemutatott eszközök alapvető funkciókat látnak el, de puritán megjelenésük nagyfokú hatékonyságot takar. Az eszközök fejlesztése folyamatos. Az itt bemutatott kép egy pillanatnyi állapotot tükröz, amely a tesztelés közben felmerülő problémák megoldása és a felmerülő igények kielégítése közben változik. A bemutatott

programok az ELGI belső hálózatán érhetőek el, az adatokhoz való hozzáférés külső felhasználók számára biztonsági okokból korlátozott. Három eszközt mutatok be, melyekből az első egy parancssoros lekérdező alkalmazás, a másik kettő pedig WEB böngészőn keresztül használható eszköz.

ggdbx

A parancssoros alkalmazás nevét a GAIA rendszer hasonló alappilléreiről, a dbx-ről kapta. C nyelven írt elődjétől eltérően ez egy java program. Ereje UNIX-os környezetben nyilvánul meg igazán. Az awk és xmlStarlet parancsokkal összekapcsolva a gyakorlott felhasználó kezében igen hatásos eszköz. Argumentumai parancsokból és opciókból állnak. A parancsok a következők:

md metaadat kiíratás
dd adat kiíratás
geom geometria kiíratás WKT (wellknown text) formátumban
l azonosító lista kiíratás (alapértelmezés)

Az opciók a találati halmaz meghatározására és a kiíratás formájára vonatkozó paramétereket adják meg. Az opciók megadásának alakja opt=paraméter. A következő paraméterek definiálására van lehetőség:

clt Osztály típusa (class type). Az értéktartomány a GGDM modell osztályainak felel meg. A kódok a szokásos GGDM rövidítések. (GOS|GO|REP|ME|LO|LOC| ...)
hl Hierarchia szint (hierarchy level). Az értéktartomány azonos a clt paraméterekkel.
ot Objektum típus (object type). Az értéktartomány a GGDM sémában definiált kódlistával egyezik. (VES|TDEM|MT|gravimetry|magnetometry|boreholeLog ...)
ost Objektumcsoport típusa (object type) Az értéktartomány a GGDM sémában definiált kódlistával egyezik. (campaign|project|objectSet|objectGroup|geophCoverage ...)
id Azonosító kód, (fileIdentifier)
pid Szülő azonosító (parentIdentifier)
gid Csoport azonosító (groupIdentifier)
lst azonosító lista fájl neve
frm kimeneti formátum (xml|html)

xsl xsl fájl neve. A megadott fájlal xsl konverzió történik. (frm=xml beállítás elvárt)

A programban rejlő lehetőségeket és a használatot néhány gyakorlati példán keresztül mutatom meg.

1. Projekt típusú objektumcsoportok kilistázása:

```
ggdbx clt=gos ost=project
```

```
PRJ_vemend2008  
PRJ_ARBN_1987-91  
PRJ_INV_uhuta2006  
...
```

Parancs argumentum nincs megadva. Alapértelmezésben a program azonosító listát nyomtat. Az eredmény a projekt típusú objektumcsoportok listája.

2. Keressük egy megadott projekt alá tartozó kampányokat:

```
ggdbx clt=gos pid=PRJ_vemend2008
```

```
GOS_MOD_vemend  
GOS_TDEM_vemend
```

3. Keressük a megadott kampányhoz tartozó méréseket:

```
ggdbx clt=go pid=GOS_TDEM_vemend
```

```
GO_TDEM_vmnd-1  
GO_TDEM_vmnd-10  
GO_TDEM_vmnd-11  
GO_TDEM_vmnd-12  
GO_TDEM_vmnd-13  
GO_TDEM_vmnd-14  
...
```

4. Adott azonosítójú objektum metaadatainak kiírása:

```
ggdbx md clt=go id=GO_TDEM_vmnd-20 frm=xml
```

```
<itemCollection>
```

```

<item id="GO_TDEM_vmnd-20">
<geophObject xmlns:gml="http://www.opengis.net/gml" uuid="elgi.hu-GO_TDEM_vmnd-20">
<dataProvider>http://www.elgi.hu</dataProvider>
<identifierName>GO_TDEM_vmnd-20</identifierName>
<parentIdentifier>GOS_TDEM_vemend</parentIdentifier>
<objectType>TDEM</objectType>
<metadata uuid="MD_elgi.hu-GO_TDEM_vmnd-20">
<hierarchyLevel>geophObject</hierarchyLevel>
<language>EN</language>
<characterSet>utf8</characterSet>
<metadataContact uuid="custodian=sores@elgi.hu">
...

```

A példa az md parancs és az frm opció használatát mutatja. A kiíratás xml formátumban történik.

5. Adott azonosítójú objektum terítési komponenseinek azonosító listája:

```
ggdbx clt=loc id=GO_TDEM_vmnd-20
```

```

GO_TDEM_vmnd-20_0_1
GO_TDEM_vmnd-20_0_2
GO_TDEM_vmnd-20_1_1
GO_TDEM_vmnd-20_1_2
GO_TDEM_vmnd-20_2_1
GO_TDEM_vmnd-20_2_2
GO_TDEM_vmnd-20_3_1
GO_TDEM_vmnd-20_3_2

```

6. Adott azonosítójú objektum sorszámmal meghatározott terítési komponensének kiíratása:

```
ggdbx dd clt=loc id=GO_TDEM_vmnd-20 i=1 j=1 frm=xml
```

```

<itemCollection>
<item id="GO_TDEM_vmnd-20">
<layoutComponent uuid="GO_TDEM_vmnd-20_0_1">
<i>1</i>
<deviceType>rectangularCurrentLoop</deviceType>
<layoutComponentType>source</layoutComponentType>
<box>
<size>
<x>150.00</x>
<y>150.00</y></size>
<orientation>
<z>1.0</z>
</orientation>
</box>
...

```

A példa a dd parancs használatát mutatja. Egy adott mérés adott terítési komponensét a terítés sorszámaival (i=1) és a terítésen belül a komponens sorszámaival (j=1) adhatjuk meg.

7. Azonosító listák használata I.

```
ggdbx clt=ps hl=gos lst=id.lst
```

```
PS_0_elgi.hu_CMP_ARBN-1  
PS_0_elgi.hu_CMP_ARBN-42  
PS_0_elgi.hu_CMP_ARBN-43  
PS_0_elgi.hu_CMP_ARBN-44  
PS_0_elgi.hu_CMP_ARBN-45  
PS_0_elgi.hu_CMP_ARBN-46  
...
```

Az id.lst nevű fájl légeofizikai kampányok azonosító listáját tartalmazza. Ez a parancs kiírja a listában szereplő kampányokhoz tartozó paramétercsoportokat.

8. Azonosító listák használata II.

```
ggdbx dd clt=ps hl=gos lst=id.lst frm=xml
```

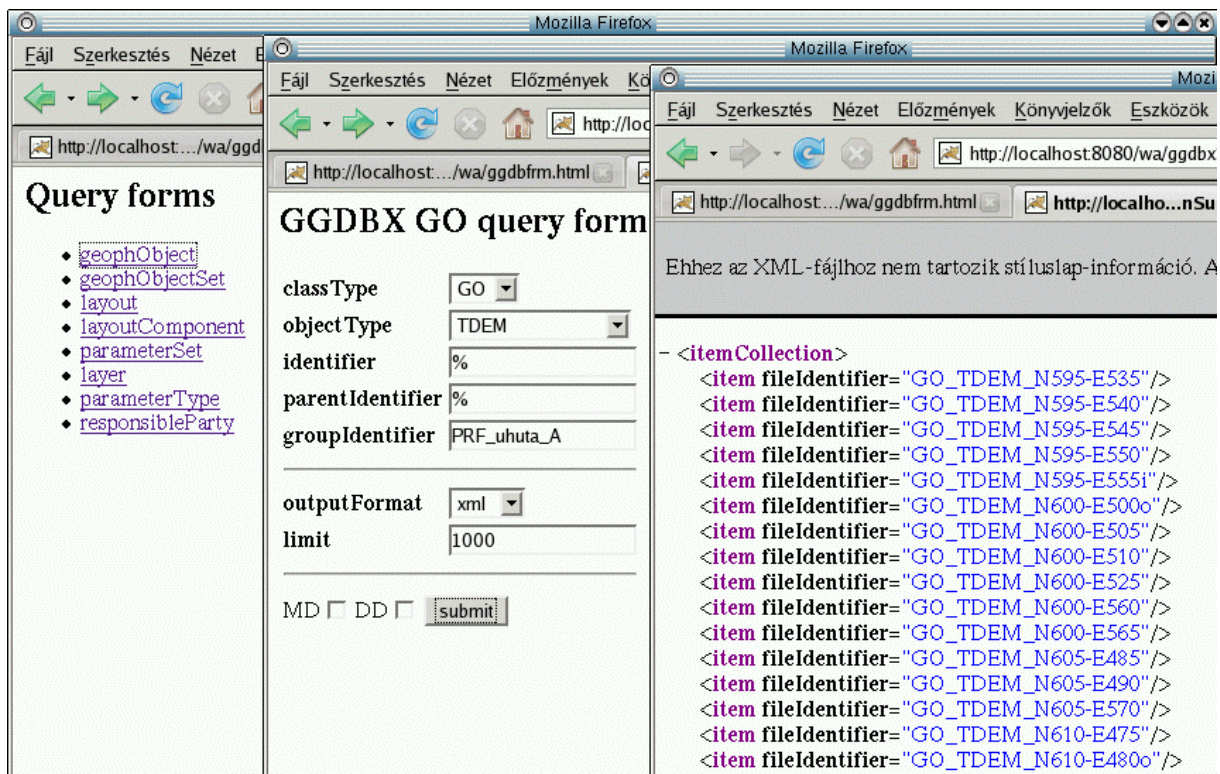
```
<itemCollection>  
<item id="CMP_ARBN-1">  
<parameterSet uuid="PS_0_elgi.hu_CMP_ARBN-1">  
<parameter><parameterCode>FLGT_DIR</parameterCode><GG_ParameterDouble><value>0.0</value></GG_P  
arameterDouble></parameter>  
<parameter><parameterCode>LN_DIST</parameterCode><GG_ParameterDouble><value>250.0</value></GG_  
ParameterDouble></parameter>  
<parameter><parameterCode>TIELN_DIST</parameterCode><GG_ParameterDouble><value>10.0</value></G  
G_ParameterDouble></parameter>  
<parameter><parameterCode>FLGT_SPD</parameterCode><GG_ParameterDouble><value>180.0</value></GG  
_ParameterDouble></parameter>  
...
```

A dd parancs hatására megkapjuk a listában szereplő kampányok paramétereit. A kimenet egy kollekció (itemCollection) melynek minden eleme tartalmaz egy paramétercsoportot. Az <item> elem id attribútuma a kampányazonosító.

A kimenetet átküldhetjük egy XSL konverteren, és tetszőleges formázást alkalmazhatunk. Az XSL fájlban elhelyezett XPATH szűrőkkel további adatszűrésre van mód.

ggdbxServlet

A ggdbx alkalmazás magja egy GGDBX java osztály, amely nem csak parancssoros környezetben használható. A ggdbxServlet program egy egyszerű weblapon keresztül teszi ugyanezeket a funkciókat elérhetővé. (28. ábra) A képen bemutatott lekérdezésben a PRF_uhuta_A azonosítójú szelvényhez (másodlagos csoport) tartozó tranziens mérések listáját kapjuk meg.



28. ábra Adatbázis lekérdezés web lapokon keresztül

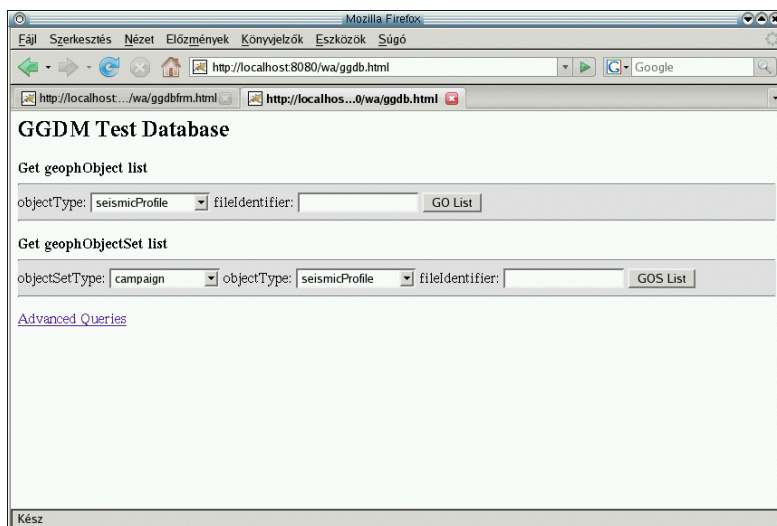
Első lépés az osztály típus kiválasztása. Ezután következik a paraméterek megadása, végül a keresés indítás.

Ha az egyébként rendkívül puritán weblapot két példányban nyitjuk meg, a két kereső ablak szimultán használatával nagyon hatékony és kényelmes adatkezelést végezhetünk. Az egyik lapot a listák, a másikat a listázott objektumok részleteinek megjelenítésére használhatjuk.

A GGDB böngésző

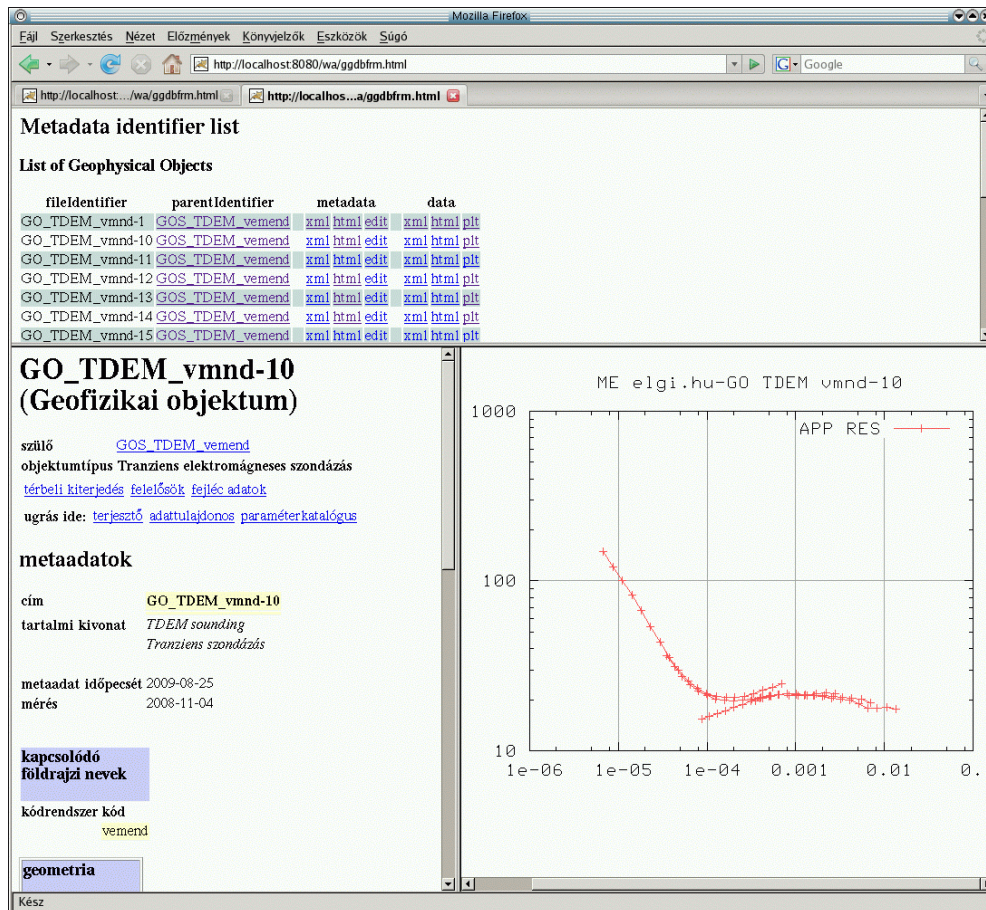
A GGDB adatbázis browser egyszerű és könnyen használható eszköz, egy olyan web alkalmazás, amely kényelmes navigációt, a hierarchikus rendszeren való le-fel mozgás

lehetőségét biztosítja a felhasználó számára. (29. ábra) A panel első sora geofizikai objektumok listaszerű megjelenítésére szolgál. A legördülő menüben megadható az objektum típusa, és wildcard karakterekkel az azonosító. A második sor az objektumcsoportok listázására szolgál. Itt az objektumtípus mellett az objektumcsoport típusa is megadandó. A belépő oldalról egy HTML frame-be jutunk, ahol felül a kért lista jelenik meg. (30. ábra) A lista elemein (fileIdentifier) kattintva a gyermek objektumokat tartalmazó listákba jutunk. A szülő elemekhez (parentIdentifier) tartozó linkek az azok gyermekeit tartalmazó listákat hozzák fel. A lista előállítására előtt a program ellenőrzi a szülő típusát. Elsődleges csoportoknál a gyermekek, másodlagos objektumcsoportoknál a csoporttagok listája jelenik meg. (Ez a kettő más típusú lekérdezést jelent.) Lehetőség van a metaadatok és adatok különböző formátumú (XML, HTML, grafikus) megjelenítésére. A metaadatok lent baloldalon, az adatok lent jobboldalon jelennek meg. Ezáltal a felhasználó jól áttekinthető képet kap az adatbázisban tárolt mérések összes adatáról.



29. ábra A GGDGM böngésző alkalmazás belépő felülete.

Hangsúlyozandó, hogy az itt bemutatott alkalmazások elsődleges célja elősegíteni az adatmodellben, és a hibrid megvalósításban rejlő lehetőségek, nehézségek felismerését, valamint megkönnyíteni a rendszer alapmoduljainak tesztelését. A kifinomult felhasználói igények kielégítésére szolgáló professzionális környezet kialakítása a fejlesztés jelenlegi fázisában nem cél, ez a jövő feladata. Mindazonáltal az itt bemutatott eszközök a hozzáértő felhasználók részére biztosítják a feltételeket a teszt adatbázis próbaüzem alatti használatához.



30. ábra A GGDB böngésző alkalmazás munka közben.

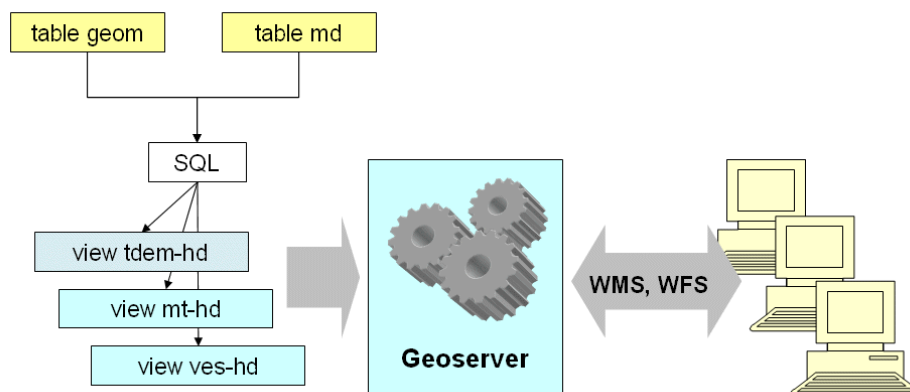
4.2.4. Térinformatika

A GGDM alapú adatbázis egyik fontos célja a térinformatikai rendszerekhez való illeszkedés biztosítása. Ezt a rendszer több szinten is támogatja. Az ISO 19115 metaadat szabvány térbeli kiterjedésre vonatkozó része kötelezően tartalmazza az objektumokat határoló négyszög koordinátáit, valamint opcionálisan a forrás adatok geometriáját. A GGDM adatmodellben geofizikai objektumokra nézve ez kötelező. A metaadatokban a geometria leírására a GML földrajzi leíró nyelvvel történik. Az OGC szabványokat követő térinformatikai rendszerekben, valamint az ezek által preferált WEB szolgáltatásokban a GML használata általános. A WMS (Web Map Service) WFS (Web Feature Service) szolgáltatások ezekre épülnek, sőt a geotudományokban egyre terjedő GeoSciML is. A geofizikai méréseket reprezentáló objektumok azok típusától függően általában pont, lineString, vagy polygon geometriával írhatók le. Ez a reprezentáció a GEOMIND Profilban alkalmazott geofizikai objektum – feature megfeleltetés

szintjén igaz. (Pl. egy VESZ szondázás, vagy egy szeizmikus szelvény egyaránt feature, az előző pont, az utóbbi lineString geometriával)

Az általános adatmodell ME_Measurement osztálya ennél sokkal részletesebb geometriát tartalmaz, (GG_LocalCRS, GG_Box) de ez a feldolgozás szempontjából fontos, a megkutatottsági térképek szintjén érdektelen.

A metaadat rekordban leírt geometriának harmonizálni kell a mérési geometriával, erről az XML feldolgozó programoknak gondoskodniuk kell. A térbeli kiterjedést leíró GML-t a mérési pozíciót definiáló helyi koordináta rendszer (GG_LocalCRS) és a terítési komponensek (GG_LayoutComponent) geometriája alapján kell meghatározni. Mivel különböző rendszerek ugyanannak a geometriának különböző reprezentációját használják, ezeket vagy menetközben állítjuk elő, vagy redundánsan tároljuk. A térképi megjelenítéseknél a gyorsaság számít, ezért a GGDM a második megoldással él.



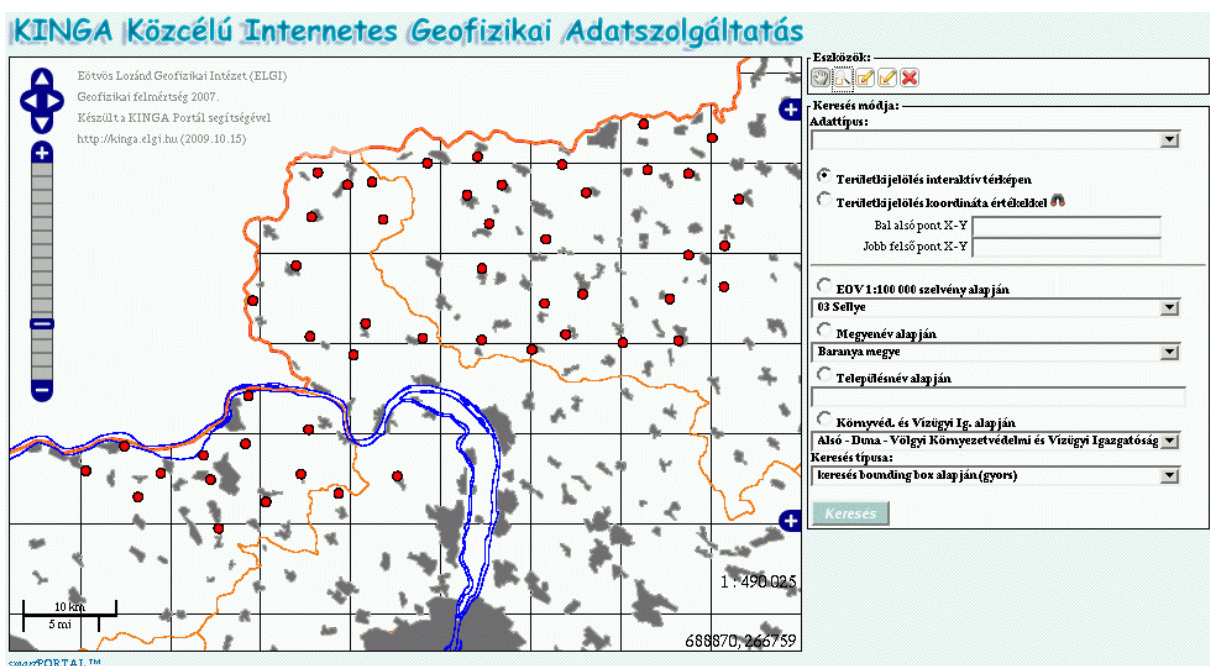
31. ábra Az adatbázis és a Geoserver kapcsolata

A geometriát és fejléc paramétereiket tartalmazó postgresSQL view-k a Geoserver WMS és WFS szolgáltatása révén a külvilág számára láthatók

Van egy harmadik szint is, amely az optimalizáció és a más rendszer komponensekhez való illeszkedés érdekében további „redundanciát” hoz be. Az egyes objektumok, objektum csoportok és jelentések geometriáját speciális táblákba kigyűjtve külön tároljuk. A postgresSQL támogatja a geometria típusú mezők használatát. A postGIS függvénykönyvtár magas szintű GIS függvények használatát is lehetővé teszi az SQL lekérdezésekben. Ezeket a funkciókat használjuk ki a GGDM alapú adatbázisokon végzett gyors GIS keresések során. Ugyanezt a lehetőséget aknázzuk ki az adatbázis – Geoserver kapcsolatban is. Az

objektumok, objektum csoportok és jelentések geometriáját natív PostgreSQL mezőkben tároló táblákból és a metaadat rekordok fejléc adataiból tárolt lekérdezésekkel ideiglenes táblákat (header view) hozunk létre. Ezek a valódi táblákkal egyenértékű, önálló táblákként érhetők el. Ezeket a Geoserver mint térképi adatforrást látja, és a WMS, WFS szolgáltatásokon keresztül a külvilág számára láthatóvá teszi. (31. ábra)

A WMS és WFS szolgáltatások lehetővé teszik, hogy GIS alkalmazások az Interneten keresztül szabványos lekérdezésekkel kommunikálhassanak az adatbázissal. A KINGA rendszer felhasználói felülete mögött épp ilyen mechanizmus működik. (32. ábra)



32. ábra A KINGA portál GIS modulja a WMS által szolgáltatott térképi tartalommal

A képen látható kartográfiai rétegeket a Geoserver ESRI shape fájljokból, a geofizikai méréseket jelző pontokat a PostgreSQL adatbázisból veszi. A kép szabványos WMS kérésre adott válaszként, kép formátumban jut el a KINGA klienshez.

Tér adatok elérése WMS szolgáltatáson keresztül

A WMS (Web Mapping Service) szabványos protokoll térképi információk Interneten keresztül történő továbbítására. Első verzióját az OGC 2000-ben publikálta, amit folyamatos fejlesztés követett. (Open Geospatial Consortium Inc. 2006a) A WMS-re épülő térkép megjelenítés lényegesen különbözik a megszokottól. A hagyományos térinformatikai alkalmazások esetében a felhasználói kérés és annak kiszolgálása egy fekete dobozon belül

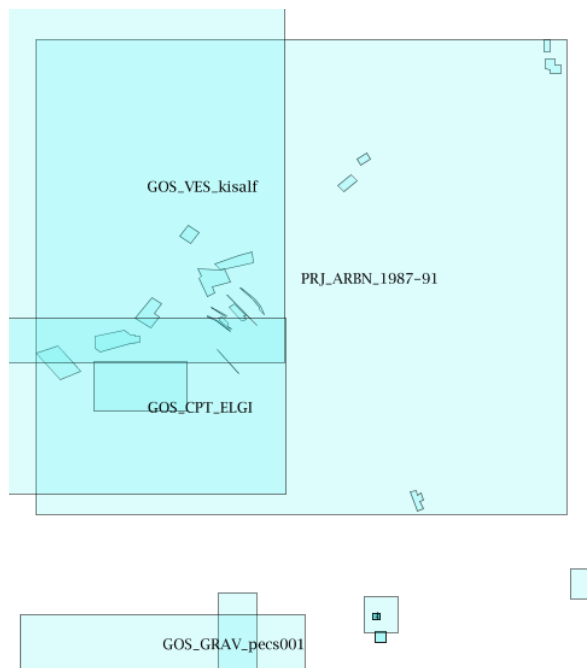
történik. A felhasználó utasításai, amelyeket a scrollbar mozgásával, az egér kattintásaival ad ki, közvetlenül kerülnek feldolgozásra a program által meghatározott formában, láthatatlanul. A WMS esetében a felhasználói kérések HTTP csatornán keresztül közvetített szabványos üzenetek formájában jutnak el a szerverhez. A szerver válasza vagy egy XML-be ágyazott üzenet, vagy egy kép a kért térképkivágással. Egy WMS kérés nem igényel semmilyen különleges szoftvert, csak egy szokásos böngészőt. A térkép kezelő WEB alkalmazásnak nem kell a grafikával, vetületi rendszerekkel bajlódnia, csak a megfelelő kéréseket megfogalmazni és továbbítani a szerver felé, majd a választ egy HTML lapon megfelelő módon lehelyezni.

A WMS szolgáltatás révén a GGDM adatbázisból egyszerű HTTP kommunikációval térképi adatokat nyerhetünk ki. A WMS-t vektoros térképi adatokhoz használhatjuk. A kereső stringben meg kell adni a térképi réteg nevét, a koordináta négyszöget, és a grafikus válasz formátumát. Ennek illusztrálására bemutatok egy egyszerű lekérdezést, amely a kliens oldalon közvetlenül felhasználható képi információt szolgáltat. A kérést HTTP get formájában kell a Geoservernek címezni:

1.példa:

```
HTTP://localhost/geoserver/wms?bbox=507398.75,77068.0,679686.25,295572.0&styles=&Format=image/gif&request=GetMap&version=1.1.1&layers=ggdb:gos-poly&width=484&height=550&srs=EPSG:23700
```

Az URL a GGDM adatbázisban lévő objektumcsoportok körvonalainak lekérdezésére szolgál. A megjelölt formátumnak megfelelően (image/gif) eredményül egy képet kapunk, ami a megadott kivágásban mutatja a kért térképi réteget (ggdb:gos-poly). Az „srs” (spatial reference system) paraméter értéke EPSG:23700, ami az EPSG adatbázisában az EOV rendszerint azonosítja. A választ, vagyis a kliensnek küldött gif képet a **33. ábra** mutatja. A formátum megadásakor több választási lehetőség is van. A Geoserver többek között a PNG, SVG, PDF, és KML formátumot is támogatja. A KML segítségével a GGDM téradatok a GoogleEarth program szolgáltatásainak körébe is bevonhatók.



33. ábra. A WMS kérésre a Geoserver által a küldött gif formátumú kép

Tér adatok elérése WFS szolgáltatáson keresztül

A WFS (Web Feature Service) a térképi objektumok mögött lévő adatbázis információ lekérdezésére szolgál. Szintén az OGC által támogatott szabvány, amely párhuzamosan fejlődött a WMS-sel. (Geospatial Consortium Inc. 2005) A feature általános térképi objektumot jelent. A szó magyar fordítása alakzat, de ez nem fedi a fogalom lényegét, miszerint a feature nem csak geometriával, hanem adatbázis tartalommal is jellemezhető. A WFS számos alapszabványra épül. A feature fogalmának pontos meghatározása a GML séma definíciós csomag része.

A WFS szolgáltatás lehetővé teszi, hogy a GGDM adatbázis geom táblájában lévő attribútumok és térbeli adatok felhasználásával lekérdezéseket végezzünk. A WMS-sel ellentétben nem térképi grafika, hanem a kapcsolt adatbázis tartalom kezelésére való. A WFS nyelvi elvárásainak megfelelően komplex kereső feltételeket adhatunk meg. A kérést HTTP post formájában kell a Geoserver címére eljuttatni. A kérés és a válasz is természetesen XML szöveg. Néhány egyszerű példa:

2. példa:

DescribeFeatureType: Adattípus (feature) attribútumainak lekérdezése:

Az alábbi példa a tranziens fejléc adattábla szerkezetét kérdezi le. A kérdésben mindössze az adattípus névtérrel (namespace) kiegészített azonosítóját kell megadni.

A HTTP kérés:

```
<DescribeFeatureType version="1.1.0" service="WFS" xmlns="HTTP://www.opengis.net/wfs"
  xmlns:ggdb="http://www.elgi.hu/ggdb" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <TypeName>ggdb:tdem-hd</TypeName>
</DescribeFeatureType>
```

A válasz:

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:ggdb="http://www.elgi.hu/ggdb"
  xmlns:gml="http://www.opengis.net/gml" elementFormDefault="qualified"
  targetNamespace="http://www.elgi.hu/ggdb">
  <xsd:import namespace="http://www.opengis.net/gml"
  schemaLocation="http://localhost:80/geoserver/schemas/gml/3.1.1/base/gml.xsd"/>
  <xsd:complexType name="tdem-hdType">
    <xsd:complexContent>
      <xsd:extension base="gml:AbstractFeatureType">
        <xsd:sequence>
          <xsd:element maxOccurs="1" minOccurs="0" name="goid" nillable="true"
            type="xsd:int"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="fileidentifier" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="parentidentifier"
            nillable="true" type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="arr_type" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="azm" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="loop_sz_min" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="loop_sz_max" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="tm_offs_min" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="tm_offs_max" nillable="true"
            type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="the_geom" nillable="true"
            type="gml:PointPropertyType"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
```

```

</xsd:complexType>
<xsd:element name="tdem-hd" substitutionGroup="gml:_Feature" type="ggdb:tdem-hdType"/>
</xsd:schema>

```

A válaszban a tranziens fejléc adattábla (mint featureType) sémadefinícióját kapjuk meg egy complexType XSD elem formájában..

3. példa:

GetFeature: téradatrekord (feature) lekérdezése attribútum érték alapján

A kérés:

```

<wfs:GetFeature service="WFS" version="1.1.0"
xmlns:ggdb="http://www.elgi.hu/ggdb" xmlns:wfs="http://www.opengis.net/wfs"
xmlns:ogc="http://www.opengis.net/ogc" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/wfs
http://schemas.opengis.net/wfs/1.1.0/wfs.xsd">
  <wfs:Query typeName="ggdb:tdem-hd">
    <ogc:Filter>
      <PropertyIsEqualTo>
        <PropertyName>fileidentifier</PropertyName>
        <Literal>GO_TDEM_N545-E530</Literal>
      </PropertyIsEqualTo>
    </ogc:Filter>
  </wfs:Query>
</wfs:GetFeature>

```

A GetFeature kérés egy ogc:filter elemben megadva az adattípus teljes nevét és egy szűrőt tartalmaz, ami az attribútum nevéből és értékéből áll. Válaszul azt a tdem-hd rekordot kapjuk, amelynél a fileidentifier="GO_TDEM_N545-E530" feltétel teljesül.

A válasz:

```

<wfs:FeatureCollection numberOfFeatures="1" timeStamp="2009-10-27T12:00:43.732Z"
xsi:schemaLocation="http://www.elgi.hu/ggdb
http://localhost:80/geoserver/wfs?service=WFS&version=1.1.0&request=DescribeFeatureTyp
e&typeName=ggdb:tdem-hd http://www.opengis.net/wfs
http://localhost:80/geoserver/schemas/wfs/1.1.0/wfs.xsd"
xmlns:ogc="http://www.opengis.net/ogc" xmlns:tiger="http://www.census.gov"
xmlns:ggdb="http://www.elgi.hu/ggdb" xmlns:wfs="http://www.opengis.net/wfs"
xmlns:topp="http://www.openplans.org/topp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:sf="http://www.openplans.org/spearfish" xmlns:ows="http://www.opengis.net/ows"
xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www.w3.org/1999/xlink">

```

```

<gml:featureMembers>
  <ggdb:tdem-hd gml:id="tdem-hd.nfm-bab5cda_12495b79d04_-75ef">
    <ggdb:goid>128</ggdb:goid>
    <ggdb:fileidentifier>GO_TDEM_N545-E530</ggdb:fileidentifier>
    <ggdb:parentidentifier>GOS_TDEM_uhuta2002</ggdb:parentidentifier>
    <ggdb:arr_type>CIL</ggdb:arr_type>
    <ggdb:azm>0</ggdb:azm>
    <ggdb:loop_sz_min>50.00</ggdb:loop_sz_min>
    <ggdb:loop_sz_max>50.00</ggdb:loop_sz_max>
    <ggdb:tm_offs_min>6.900000e-06</ggdb:tm_offs_min>
    <ggdb:tm_offs_max>7.070000e-04</ggdb:tm_offs_max>
    <ggdb:the_geom>
      <gml:Point srsName="urn:x-ogc:def:crs:EPSG:23700">
        <gml:pos>615450.0 95300.0</gml:pos>
      </gml:Point>
    </ggdb:the_geom>
  </ggdb:tdem-hd>
</gml:featureMembers>
</wfs:FeatureCollection>

```

A válasz XML egyetlen tagú kollekciónból áll, ami a rekord és a szülő objektum azonosítóját, valamint az összes fejlécparamétert (elrendezés típusa, azimuth, keret méret, mintavételi időtartomány) tartalmazza. Az utolsó elem a geometria, amely egy pont objektum gml leírásából áll.

4. példa:

GetFeature máshogy: téradatrekord (feature) lekérdezése koordináta ablak alapján

A kérés:

```

<wfs:GetFeature service="WFS" version="1.1.0" xmlns:ggdb="http://www.elgi.hu/ggdb"
xmlns:wfs="http://www.opengis.net/wfs" xmlns:ogc="http://www.opengis.net/ogc"
xmlns:gml="http://www.opengis.net/gml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/wfs http://schemas.opengis.net/wfs/1.1.0/wfs.xsd">
  <wfs:Query typeName="ggdb:tdem-hd">
    <wfs:PropertyName>ggdb:fileidentifier</wfs:PropertyName>
    <ogc:Filter>
      <ogc:BBOX>
        <ogc:PropertyName>the_geom</ogc:PropertyName>
        <gml:Envelope srsName="http://www.opengis.net/gml/srs/epsg.xml#23700">
          <gml:lowerCorner>615920 95320</gml:lowerCorner>
          <gml:upperCorner>615970 95470</gml:upperCorner>
        </gml:Envelope>
      </ogc:BBOX>
    </ogc:Filter>
  </wfs:Query>

```

```
</wfs:GetFeature>
```

Azon pontok azonosítóját (fileidentifier) kérjük, amelyek a megadott koordinátaablakon belül helyezkednek el.

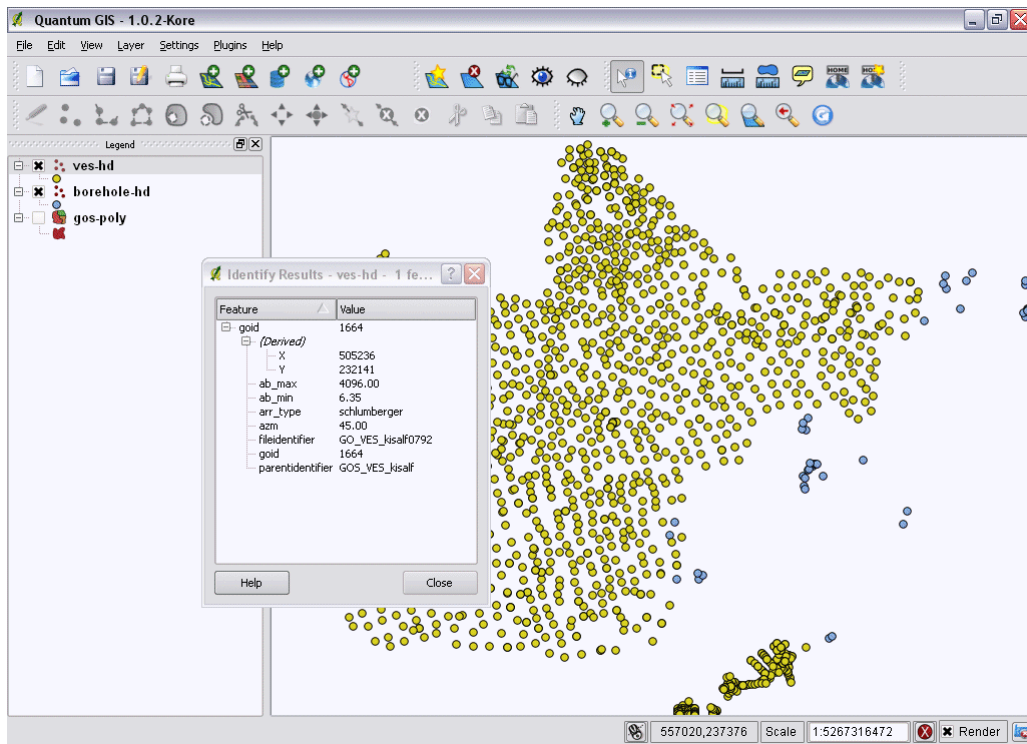
A válasz:

```
<wfs:FeatureCollection numberOfFeatures="3" timeStamp="2009-10-27T12:08:58.052Z"
xsi:schemaLocation="http://www.elgi.hu/ggdb
http://localhost:80/geoserver/wfs?service=WFS&version=1.1.0&request=DescribeFeatureTyp
e&typeName=ggdb:tdem-hd http://www.opengis.net/wfs
http://localhost:80/geoserver/schemas/wfs/1.1.0/wfs.xsd"
xmlns:ogc="http://www.opengis.net/ogc" xmlns:tiger="http://www.census.gov"
xmlns:ggdb="http://www.elgi.hu/ggdb" xmlns:wfs="http://www.opengis.net/wfs"
xmlns:topp="http://www.openplans.org/topp" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:sf="http://www.openplans.org/spearfish" xmlns:ows="http://www.opengis.net/ows"
xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www.w3.org/1999/xlink">
  <gml:featureMembers>
    <ggdb:tdem-hd gml:id="tdem-hd.nfm-bab5cda_12495b79d04_-7072">
      <ggdb:fileidentifier>GO_TDEM_N595-E535</ggdb:fileidentifier>
    </ggdb:tdem-hd>
    <ggdb:tdem-hd gml:id="tdem-hd.nfm-bab5cda_12495b79d04_-7071">
      <ggdb:fileidentifier>GO_TDEM_N595-E540</ggdb:fileidentifier>
    </ggdb:tdem-hd>
    <ggdb:tdem-hd gml:id="tdem-hd.nfm-bab5cda_12495b79d04_-7070">
      <ggdb:fileidentifier>GO_TDEM_N595-E545</ggdb:fileidentifier>
    </ggdb:tdem-hd>
  </gml:featureMembers>
</wfs:FeatureCollection>
```

A kollekción 3 pontból áll. Mivel a kérésben megadtuk a „fileidentifier” attribútumot, a listában csak ez szerepel.

A WMS és WFS szolgáltatás részletes ismertetése dolgozatomban nem célja. A fenti példák csak a technológia alapelveit mutatják be, és érzékeltetik a benne rejlő lehetőségeket. A szabványos, XML-be ágyazott WMS és WFS szolgáltatások ennél sokkal többre is képesek, és a hálózati térinformatika terén nagyon hatékonyan alkalmazhatók. Rohamosan terjednek az olyan térinformatikai alkalmazások, melyek használják az XML-re épülő web szolgáltatásokat, és a távoli internetes szerverek térképi rétegeit, kapcsolt adatbázis tartalmakat képesek a hagyományos adatrendszerekkel együtt kezelni. Példaként bemutatom,

hogy a kiváló, nyílt forráskódú QuantumGIS (QGIS) program miként használja a GGDM adatbázis VESZ méréseket és fúráspontokat tartalmazó rétegeit. (34. ábra)



34. ábra GGDM adatbázisban tárolt VESZ mérések fejléc adatainak lekérdezése a QGIS programmal.

A program eszköztára tartalmaz két gombot a WMS és WFS rétegek beszurásához. A távoli adatforrások használatához mindössze a szerver internetes címét kell megadnunk, és kiválasztani az általa felajánlott adatforrásokat. A WFS lehetővé teszi a távoli adatforrás térképi elemeinek kiválasztását, és a hozzájuk kapcsolódó attribútum táblák megszokott használatát. A bemutatott példában a VESZ fejléc adatok lekérdezése közvetlenül a GGDM adatbázisból történik. Ez a folyamat nagyjából a következő:

- A felhasználó által megjelölt térképi pont koordinátája és az aktív réteg neve alapján a kliens program összeállít egy GetFeature kérést. A kérésben megfogalmazott szűrési feltétel biztosítja, hogy a kiválasztandó adatbázis rekordhoz tartozó geometriai elem a képernyőn megjelölt pont adott sugarú környezetébe essen.
- A kérést a kliens elküldi a távoli gépnek, amelyet a Geoserver WFS modulja megkap, majd értelmez. Az adott réteghöz kapcsolt adatbázis és adattábla ismeretében a

Geoserver összeállít egy geometriai SQL kérést, amit a PostgreSQL adatbázisnak küld tovább.

- A PostgreSQL válaszát a Geoserver becsomagolja egy WFS-konform XML szövegbe, és továbbítja a klienshez.
- A kliens a kapott szöveget értelmezi és megjeleníti.

Hangsúlyozom, hogy a Geoserver bármely geometriát tartalmazó PostgreSQL view-t (tárolt lekérdezés) képes térképi réteggként kezelni. A GGDM adatbázis flexibilitása gyakorlatilag korlátlan lekérdezési lehetőségeket kínál. Tetszőleges, geometriát tartalmazó táblázat, vagy view térképi réteggként kezelhető, és a hagyományos ESRI shape fájlhoz képest egy rendkívül dinamikus térképi tartalom állítható elő. A KINGA rendszer hasonló módon végzi a metaadat keresések találati halmazainak térképi megjelenítését.

A teljesség kedvéért megemlítem még a WCS (Web Coverage Service) szolgáltatást, ami hasonló a WMS-hez, de nem vektoros térképi rétegek, hanem raszter fedvények (képek, rácsadatok) hasonló módon történő elérését és térképi adatrendszerekbe történő integrálását teszi lehetővé. A Geoserver a WCS szolgáltatást is támogatja.

A nagy térinformatikai szoftvergyártók, mint az ESRI, vagy az InterGraph is felismerték az internetes térképi szolgáltatásokban rejlő lehetőségeket, és elkészítették saját szerver és kliens moduljaikat, melyek kapcsolódnak a meglévő rendszerekhez (ArcIMS WMS / GeoMedia WebMap). A legismertebb és legsikeresebb nyílt forráskódú, platform független rendszer a Geoserver. A kitűnő, szintén ingyenes és platform független QuantumGIS rendszer komoly vetélytársa a kereskedelmi programoknak. Az OGC szintén támogatja az ingyenes kliens alkalmazások létrehozását.

5. Az általános geofizikai adatmodell alkalmazásának lehetőségei

Az általános geofizikai adatmodell jelentősége, és felhasználási köre egyelőre nem belátható. A GGDM-en alapuló adatbázisokkal kapcsolatos kísérletek ígéretesek. Az adatmodell elterjedésében fontos szerepet játszanak azok a még el nem készült alkalmazások, amelyek az adatbázisokat a hétköznapi gyakorlathoz kapcsolják. Az egységes elven történő adatleírás nemcsak az adatok cseréjét könnyíti meg, de a geofizikai feldolgozó programok adatkezelő moduljait is átláthatóbbá teszi. Az „ahányféle adat, annyi adatbázis” gyakorlatát felválthatja egy áttekinthetőbb hatékonyabb adatkezelés. Ebbe az irányba mutat néhány terv, amely a létező adatbázisok egységesítésére, és a GGDM alapú adatbázisokra épülő WEB szolgáltatások fejlesztésére irányul.

5.1. A GAIA rendszer átalakítása

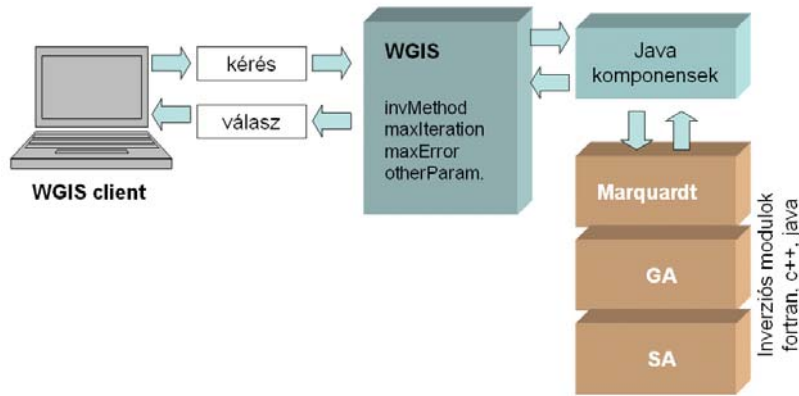
Az ELGI által fenntartott Országos Geoelektromos Adatbázis jelenleg a GAIA programrendszer keretében működik. A sajátfejlesztésű adatbáziskezelőre épülő rendszer a hatékony adatkezelés mellett számos feldolgozó, és megjelenítő funkcióval rendelkezik, melynek kiépítése éveket vett igénybe. A GAIA kifejlesztése óta eltelt idő az informatikában sok újdonságot hozott, ezért felmerült a rendszer szabványosításának és modernizálásának igénye. Az adatkezelést végző alrendszer kiváltása, a GGDM alapú adatbáziskezelőre való átállás elkezdődött. A GAIA rendszer elemei közötti adatcsere ún. DIF (Data Interchange File) formátumú állományokkal történik. Ezek strukturált szövegfájlok, amelyek szerkezete leképezhető a GGDM adatmodellre. A DIF–GGDM konverzió ilyenformán viszonylag könnyedén megoldható. A GAIA által nyújtott adatkezelő szolgáltatásoknak megfelelő GGDM adatbázis funkciók kialakítása is folyamatban van. A végeredmény egy olyan rendszer lesz, amely a felhasználó számára változatlan környezetet biztosít, de amely mögött egy egységesített adatrendszer és nyílt forrású adatbáziskezelő rendszer működik. Az egységes adatmodellnek köszönhetően nincs akadálya annak sem, hogy a GAIA rendszerben kizárólag VESZ és tranziens mérések kezelésére használt funkciókat más geofizikai módszerekre is kiterjesszük.

5.2. WEB adatszolgáltatások

Az utóbbi időben felmerült a GGDM alapú adatbázisok internetes elérésére és azok integrációjára alkalmas adatszolgáltatás kifejlesztésének terve. A Web Geophysical Data Service (WGDS) működésének alapelve, hogy az adatbázis elérése egy WEB szervizen keresztül, szabványos kérések és parancsok formájában történik. A szolgáltatás alap funkciói a WFS szabványnak felelnének meg, illetve ahol szükséges kibővítenék azokat. A szerviz lehetővé tenné a GGDM formátumú geofizikai adatok távoli gépeken történő tárolását, lekérdezését. A WGDS szerverek hálózata egy teljesen egységes osztott adatbázis rendszert valósítana meg, amelyben az egyes szerverek egyenértékű csomópontokként működnek. A WEB szolgáltatás alapja a hibrid adatbázisra épülő alkalmazásokat bemutató fejezetben hivatkozott GGDBX java osztály lenne.

5.3. WEB inverziós szolgáltatások

Az egységes adatszerkezet használata megkönnyíti az általános célú geofizikai alkalmazások fejlesztését. A WEB Geophysical Inversion Service (WGIS) egy olyan tervezett szolgáltatás, amely lehetővé teszi geofizikai adatok távoli gépeken történő inverziós feldolgozását. A kliens GGDM formátumban elküldi a mérési adatokat és a start modellt, majd a szervertől megkapja az inverzió eredményét. (35. ábra) A feldolgozás módját, az alkalmazandó módszert, a különböző technikai paramétereket a felhasználó szintén be tudná állítani. Az inverziós szerver feldolgozó moduljai egységes interfésszel rendelkező programok, amelyek a WEB szerviz moduljaival kommunikálnak. A moduláris felépítés miatt a matematikai funkciók és az adatkezeléssel kapcsolatos feladatok elválnak egymástól, ami a fejlesztő munka szempontjából kedvező. A szolgáltatásnak sok egyéb előnye is lehet. Az erőforrásigényes feladatok nagy teljesítményű inverziós szerverekre koncentrálhatók, bizonyos inverziós szolgáltatások akár üzleti alapon is végezhetők. A WEB szerviz mögött számítógépes klaszterek, vagy gridek is működhetnek. Az inverziós szerver a mérési adatokat biztonsági okokból nem tárolja, így a távoli felhasználónak nem kell aggódnia, hogy adatai illetéktelen kezekbe kerülnek.



35. ábra Geofizikai adatok inverziója a WGIS web szerízen keresztül

5.4. Geofizika és az INSPIRE irányelvek

Hazánkban is elkezdődött az Internetes téradat infrastruktúra kiépítésének törvényi szintű megalapozása. Ennek megfelelően az Európai Unió által lefektetett INSPIRE irányelvekben meghatározott, közérdeklődésre számot tartó témákban készült téradatok internetes hozzáférését minél nagyobb körben biztosítani kell. Első lépcsőben a téradatokra vonatkozó metaadatoknak, később maguknak a téradatoknak a nyilvánosságát a közadatokat kezelő intézményeknek biztosítani kell. A Metaadatok vonatkozásában az ISO 19115 szabvány, az adatoknál pedig a GeoSciML használata követendő. A magas szintű interoperabilitás érdekében az adatszolgáltatásnak a modern WEB technológiákra, konkrétan az OGC szabványos WMS, WFS szolgáltatásokra kell épülnie.

Az INSPIRE dokumentumokban definiált tematikus témacsoportok és térképi rétegek a geofizika tárgyszót expliciten nem jelölik meg. Az alap téradatok témacsoportjai között viszont a földtan szerepel, és mint a föld belső, fizikai szerkezetét feltáró tudomány, a geofizika egy része is nyilvánvalóan ide tartozik. Az ELGI az alap geofizikai paraméterterképek, obszervatóriumi adatok, alaphálózati mérések metaadatainak nyilvánossá tételével kapcsolatban elkötelezte magát az INSPIRE irányelveknek megfelelő adatszolgáltatás mellett. A GEOMIND és KINGA rendszerek kiépítésével és üzemeltetésével ezzel kapcsolatos feladatainak egy részét el is látta. A GGDM adatmodellre épülő WEB szolgáltatások kiépítése újabb lépés lehet ebben az irányban.

5.5. GGDM és a GeoSciML

A GeoSciML egy dinamikusan fejlődő XML alapú nyelv, amely fokozatosan terjeszkedik a földtudományok különböző részterületein. (Sen, M., Duffyb, T., 2005). Több tematikus modellező nyelvre épül. A tér adatok, valamint a tér- és időbeli megfigyelések adatainak leírását geometriai alapfogalmakon keresztül valósítja meg. Az OGC által is támogatott nyelv az INSPIRE egyik alappillére lett. A GeoSciML fő alkotóelemei a következők:

- GML földrajzi leíró nyelv, a tér- és időbeli objektumok leírásának szabatos eszköze. (Open Geospatial Consortium Inc. 2005b)
- ISO 19115 metaadat szabvány
- Observations & Measurements (OM) séma definíciós csomag, mérési adatok és megfigyelések GML-re épülő leírását teszi lehetővé. (Open Geospatial Consortium Inc. 2007a)
- Sensor Modeling Language (SensorML), szenzorok tulajdonságainak leírására használható nyelv. (Open Geospatial Consortium Inc. 2007b)

A GeoSciML látványos fejlődése a GEOMIND projekttel nagyjából párhuzamosan indult el. Túl az alapok létrehozásán megtörtént a geológiai objektumok és fúrások struktúráinak kialakítása. Az INSPIRE témacsoportokhoz kapcsolódó adatszerkezetek kidolgozásával foglalkozó team-ek a természetes és mesterséges környezet számos elemének szabatos leírására törekednek. Nem vitás, hogy a GeoSciML előbb-utóbb a geofizikában is megjelenik. A GGDM a geofizikai adatok leírásában a GeoSciML-hez hasonló alapelvekre támaszkodik. További hasonlóság az ISO 19115 és a GML használata. Nagy különbség, hogy a GGDM a geofizikai adatrendszerek ábrázolására a kutatás teljes vertikumát átfogó hierarchikus rendszert használ, amelynek alján a terítési komponensek, tetején a geofizikai objektumcsoportok állnak. A GeoSciML-ben ilyen hierarchia egyelőre nem létezik. A két nyelv közötti kapcsolódás a szenzorok és érzékelők szintjén megteremthető. Ha a GGDM terítési komponensek leírásában az OM, és SensorML eszköztárát használjuk, a GGDM egyszeriben a GeoSciML geofizikai kiterjesztésévé válhat.

Összefoglalás

A geofizika világában számos adatszabvány létezik, de a mérési és feldolgozási adatrendszerek átfogó leírására alkalmas adatmodell a mai napig nem született. A legismertebb szabványok főleg az olajiparban alkalmazott legfontosabb módszerekre lettek kidolgozva. Átfogó adatmodellek helyett történeti hagyományokra, és módszer-specifikus elvekre épülnek. A GeoSciML az első olyan rendszer, amely a földtudományokban bevezette a mérések, észlelések szabatos leírását megvalósító, geometriai alapelvekre épülő leírási módot. Jelen állapotában a GeoSciML geofizikai adatok és adatrendszerek leírására még nem használható. Dolgozatomban bemutatok két különböző szintű adatmodellt, amely erre a célra alkalmasnak bizonyult. Egyik a GEOMIND profil, másik az Általános Geofizikai Adatmodell (General Geophysical Data Model, GGDM). Az első egy geofizikai metaadat szabvány, amely az elmúlt években elindított két internetes geofizikai portál (GEOMIND: <http://www.geomind.eu>, KINGA: <http://kinga.elgi.hu>) adatforgalmának alapját képezi. A második egy olyan modell, amely szakítva a „geofizikai módszer” hagyományos fogalmával a mért, feldolgozott adatok, és inverziós modellek általános, módszer független leírására alkalmas. E modellek hatékonyan csökkentik a geofizikai adatokra jellemző diverzitást egységes elvekre épülő geofizikai adatbázisok és információs rendszerek létrehozását teszik lehetővé. A geofizikai objektum, objektumcsoport és riport fogalmának bevezetésével bonyolult adatrendszerek (projektek, mérési és feldolgozási kampányok) és dokumentációs rendszerek (jelentéstárak, archívumok) képezhetők le szabványos metaadat rekordok halmazára. A GGDM olyan XML nyelv, mellyel a legbonyolultabb geofizikai mérési rendszerek is egyszerűen leírhatók a mérések során használt szenzorok és források jellemzőinek pontos leírásával.

Az Általános Geofizikai Adatmodellre épülő adatbázis megvalósítására két, gyakorlatban is kipróbált implementációt mutatok be. Az „XML dokumentumtár” egy kereső motor által generált speciális indexállomány révén rugalmasan kezelhető, hagyományos fájl archívum, amely viszonylag könnyen megvalósítható. A másik – a „Hibrid adatbázis” – egy olyan relációs rendszer, amely a gyakran keresett vezér attribútumokat hagyományos mezőkben, az önálló egységként kezelhető, mély struktúrákat pedig XML típusú mezőkben tárolja. Ezzel egyszerű adatbázis szerkezet mellett nagyfokú flexibilitást biztosít. A hibrid rendszer fejlesztési igénye jóval nagyobb. A térinformatikai funkciók támogatása mindkét rendszer esetében könnyen megoldható a nyílt forráskódú Geoserver rendszer használatával. Az XML

technológia alkalmazásával lehetőség nyílik új geofizikai web szolgáltatások kialakítására is, melyek eddig elképzelhetetlen interoperabilitást és az egymástól elkülönült rendszerek magasabb szintű integrációját teszik lehetővé.

Köszönetnyilvánítás

Köszön témavezetőmnek, Dr. Turai Endre egyetemi docensnek, hogy szakmai támogatásával segítette dolgozatom elkészítését. Köszönöm Dr. Fancsik Tamásnak az ELGI igazgatójának, a Miskolci Egyetem Geofizikai Tanszékének, és a GEOMIND Konzorciumnak, hogy munkámhoz az anyagi és intézményi háttérrel biztosították. Köszönetet mondok közvetlen kollégáimnak, akik a dolgozatban leírt munkák elvégzésében az elmúlt időszakban segítettek, különösen Vértesy Lászlónak, a Térképezési Főosztály vezetőjének, aki a téma fontosságát hangsúlyozva mindig mellettem állt és biztatott. Köszönöm Mikael Pedersennek (GEUS, Dánia), Klaus Kühnnek, Jörg Kudernek (GGA, Németország) és Valdas Rapseviciusnak (ITG, Litvánia) amiért szakértelmükkel, lelkesedésükkel és kritikai megjegyzéseikkel a GEOMIND projektben a tervezés során segítségemre voltak.

Irodalmi hivatkozások

Apache Lucene, 2009: Apache Lucene text search engine. <http://lucene.apache.org>

Apache Solr, 2009: Apache Solr search server. <http://lucene.apache.org/solr>

GEOMIND Consortium 2007a. Data Dictionary for GEOMIND ISO19115 extension classes. <http://geomind.elgi.hu/doc/GEOMIND-extensions-1.3.xml>,
<http://geomind.elgi.hu/doc/GEOMIND-extensions-1.3.pdf>

GEOMIND Consortium 2007b. GEOMIND Profile Schema Definition Package. http://www.geomind.eu/portal/public_files/schema/geomind/geomind/2007

Gyulai Á., Ormos T. 1998: A new procedure for the interpretation of VES data: 1.5-D simultaneous inversion method. Applied Geophysics, 1999

Halassy B 2002: Adatmodellezés. Nemzeti tankönyvkiadó, Budapest

ISO/TC 211, 2004. ISO/TC 211/WG 4/PT 19136 Geographic information - Geography Markup Language (GML). International standard specification

Open Geospatial Consortium Inc. 2006: OpenGIS Web Map Service (WMS) Implementation Specification <http://www.opengeospatial.org/standards/wms>

Open Geospatial Consortium Inc. 2005a: Web Feature Service Implementation Specification <http://www.opengeospatial.org/standards/wfs>

Open Geospatial Consortium Inc. 2005b: OpenGIS Geography Markup Language (GML) Encoding Standard <http://schemas.liquid-technologies.com/OpenGis/gml/3.1.1/>

Open Geospatial Consortium Inc. 2007a: Observations and Measurements <http://schemas.liquid-technologies.com/OpenGis/om/1.0.0/>

Open Geospatial Consortium Inc. 2007b: OpenGIS Sensor Model Language (SensorML) <http://schemas.liquid-technologies.com/OpenGis/sensorML/1.0.1/>

Prácsér E. 1986: Computing of transient response of layered halfspace, problems in apparent resistivity inversion. GEOFIZIKAI KÖZLEMÉNYEK - GEOPHYSICAL TRANSACTIONS 32(3) 221-234.

Prácsér E. 1992: Fast Computing of Transient Electromagnetic Field on the Surface of a Layered Half-space. GEOFIZIKAI KÖZLEMÉNYEK - GEOPHYSICAL TRANSACTIONS 37(2-3) 159-176.

Sen, M., Duffyb, T., 2005. GeoSciML: Development of a generic GeoScience Markup Language. Computers & Geosciences 31(9) 1095-1103

Sörös L., Pedersen M., Rapsevicius V., Kühne K., Kuder J. 2007a. Specification of standards for digital geophysical content, http://geomind.elgi.hu/doc/D6_2.pdf

Sörös L., Pedersen M., Rapsevicius V., Kühne K., Kuder J. 2007b. Specification of standards for digital geophysical content, Annex A, Data dictionary
http://geomind.elgi.hu/doc/D6_2_A.pdf

Sörös L., Pedersen M., Rapsevicius V., Kühne K., Kuder J. 2007c. Specification of standards for digital geophysical content, Annex B, Implementation & examples
http://geomind.elgi.hu/doc/D6_2_B.pdf

Sörös L. 2009b: GGDM schema definition package.
<http://geomind.elgi.hu/GGDM/ggdm-1.4.2.xsd>

Timár et al. 1996: Építsünk könnyen lassan adatmodellt. Egyetemi Nyomda, Veszprém.

Timár et al. 1997: Így még könnyebb az adatmodellezés. Egyetemi Nyomda, Veszprém.

W3C 2007a. XML Path Language (XPath) 2.0. <http://www.w3c.org/TR/xpath20>

W3C 2007b. XSL Transformations (XSLT) Version 2,0 <http://www.w3c.org/TR/xslt20>

W3C 2008. Extensible Markup Language (XML) 1.0 (Fifth Edition).
<http://www.w3c.org/TR/REC-xml/>

Felhasznált szakirodalom

Axmark D., Widenius M. M., 2004: MySQL Reference Manual. <http://www.mysql.com>

Bloch J., 2001: Effective Java: Programming Language Guide. Addison Wesley, ISBN: 0-201-31005-8

Bradley N. 2000: Az XML-kézikönyv. Veszprém : OOK Press

Dobróka M., 2001: Bevezetés a geofizikai inverzióba. Egyetemi Kiadó, Miskolc.

Graham S. et al. 2002: Building Web services with Java , XML, SOAP, WSDL, UDDI. Kiskapu kiadó Budapest

Reese G., Yanger R. J., King T. 2003: A MySQL kezelése és használata. O'Reilly, Kossuth kiadó.

Sörös L., 2000: Az Országos Geoelektromos és Elektromágneses Adatbázis című projekt eredményei 2000 december 31-ig. ELGI jelentés, Budapest.

Sörös L., 2003: Zárójelentés az Országos Geoelektromos és Elektromágneses Adatbázis téma 2001 és 2003 között született eredményeiről. ELGI jelentés, Budapest.

Sőrés L., 1999: An other GIS: Geoelectric Information System. EEGS'1999 Extended Abstr, Budapest.

Sőrés L., Prácsér E, Gulyás Á, Kiss J., 2004: The Hungarian National Geoelectric Database. EAGE'2004 Extended Abstr, Paris.

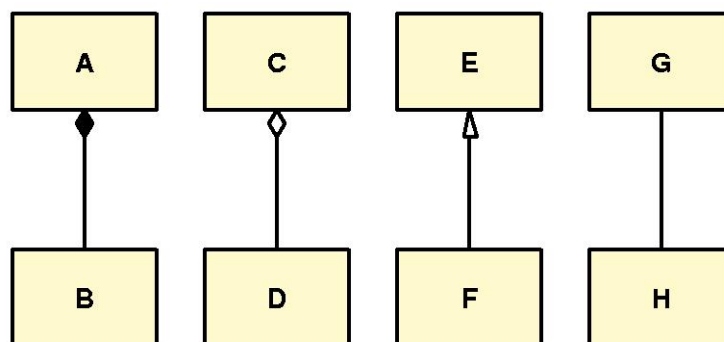
Stolnicki Gy. 1995: SQL kézikönyv. Computerbooks, Budapest

Takács E., 1981: Geoelektromos kutatómódszerek 1-2 rész. Tankönykiadó, Budapest.

1. Függelék, UML: az egységesített modellező nyelv

A dolgozatban található adatszerkezeti diagramok az UML (Unified Modelling language) jelölésrendszer szabályait követik. Az UML egy objektum orientált grafikus nyelv, amelynek a számítástechnikától az üzleti tervezésig sok alkalmazási területe van. A szabványos jelölésrendszert alkalmazó nyelv számos diagramtípust használ a valóságot modellező szerkezetek és folyamatok leírására. Dolgozatban az osztálydiagramokat használom, ezért röviden áttekintem ezek használatát. A teljesség kedvéért felsorolom a legelterjedtebb diagram típusokat:

- Use case diagrams (esetdiagram)
- Class diagrams (osztálydiagram)
- Object diagrams (objektumdiagram)
- Sequence diagram (szekvencia-diagram)
- Collaboration diagram (együtműködési diagram)
- Statechart diagram (állapotdiagram)
- Activity diagram (aktivációs diagram)
- Component diagram (komponensdiagram)
- Deployment diagram (konfigurációs diagram)



36. ábra UML osztálydiagramok jelölésrendszere:

A-B: kompozíció, C-D: aggregáció, E-F: származtatás, G-H: asszociáció

Az osztálydiagramok egy adott rendszer statikus szerkezetét és az elemek közötti kapcsolatokat (asszociációkat) ábrázolják. A dolgozatban előforduló kapcsolatokat: kompozíció

(composition), aggregáció (aggregation), asszociáció (association) és származtatás (subclassing/generalisation). Ezek jelölését a **36. ábra** mutatja. Az aggregáció és kompozíció rész-egész kapcsolat. Definiálják a befoglaló osztályt és a befoglalt részegységeket. A kompozíció jele a kitöltött rombusz. Az aggregáció jelölésénél a rombusz üres. A különbség a kettő között a kapcsolat szorosságában van. A kompozíció egy erős aggregációként értelmezhető. Akkor használjuk, ha a befoglaló elem a részegységektől függetlenül nem létezhet, pl. adatrekordok esetén a befoglaló elem törlésével a részegységeket is törölni kell. (pl.: mérés – adatrekord) Aggregáció esetén a részegység a befoglaló elemtől függetlenül is létezhet (pl.: mérés – műszer) Az asszociáció azt jelzi, hogy két objektum között valamilyen kapcsolat van.

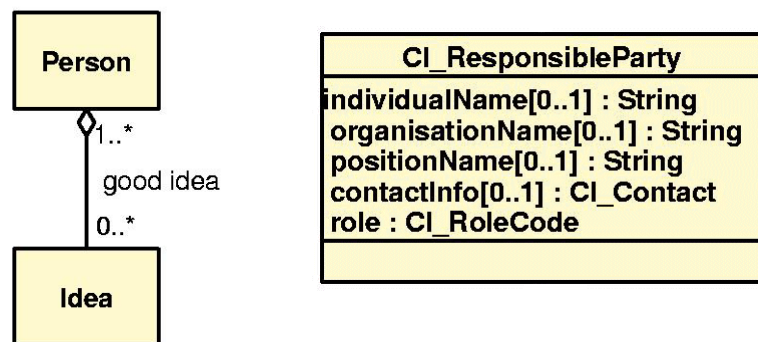
A származtatás (subclassing/generalization) – a másik irányból nézve általánosítás – azt jelenti, hogy a származtatott osztály örökli a szülő osztály tulajdonságait, és újakkal kiterjeszti azt. (pl.: személy – alkalmazott) A diagramokon a kapcsolatok számossága (cardinality) is jelölhető:

[0..1] nulla, vagy egy (opcionális)

[0..*] bármennyi

[1..*] legalább egy

[n] pontosan n



37. ábra Osztály-attribútumok megjelenítése és kapcsolat két osztály között.

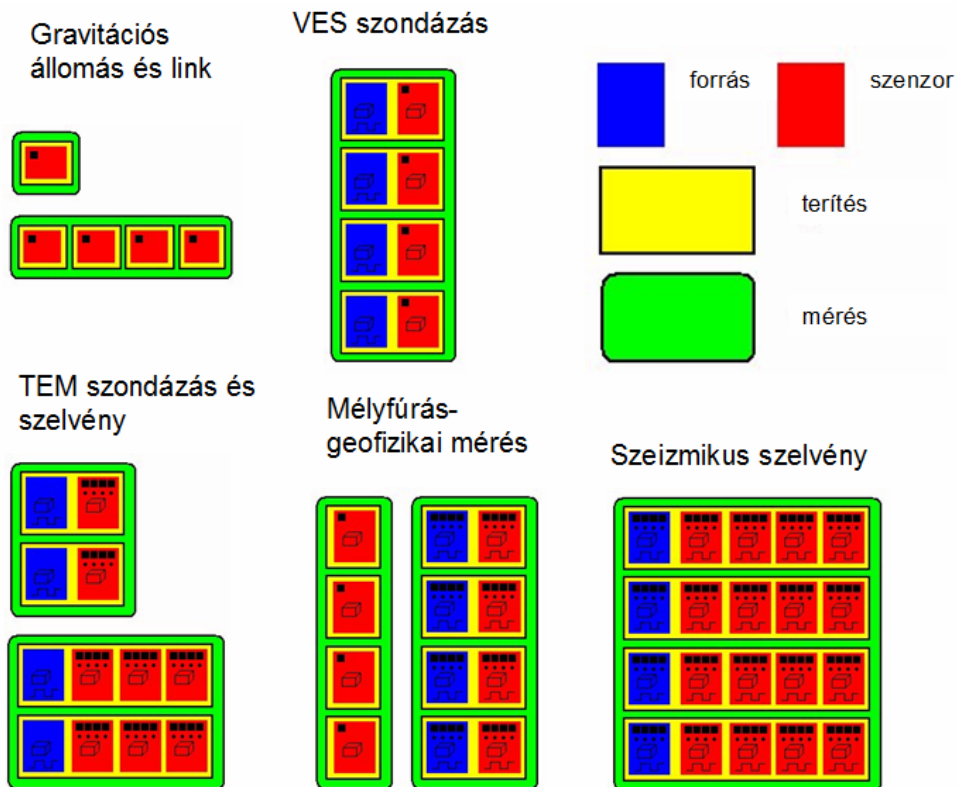
Az osztálydiagramon az attribútumok neve, típusa, a kapcsolat neve és a számossága is feltüntethető.

A **37. ábra** bal oldala példaként egy aggregáció részleteit mutatja. Egy kutatói beszélgetés során rögzített ötletek adatbázisában az osztálydiagram szerint minden személynek lehet bármennyi jó ötlete, és az is előfordulhat, hogy két személynek ugyanaz az ötlete támadt.

Az osztálydiagram lehetőséget ad a belső szerkezet részletes megjelenítésére. A jobb oldali ábrán az osztályt reprezentáló szövegdoxban felül az osztály neve, középen a hozzá tartozó attribútumok kerülnek számosságuk és típusuk megjelölésével. A dox alsó részén lehetőség van az osztályhoz kapcsolódó metódusok (függvények, szubrutinok) feltüntetésére is.

2. Függelék, XML példák

A GGDM adatmodell absztrakt elemeivel a legkülönbözőbb mérési módszerek adatai leírhatók. Az XML példák bemutatása előtt vessünk egy pillantást a 38. ábrára, amely ezt illusztrálja. Minden mérés (zöld négyszög) tartalmaz egy vagy több terítést (sárga négyszög), amely terítési komponensekből áll. A kék négyszög forrást, a piros szenzort jelent. A terítési komponensek belsejébe rajzolt piktogramok a mérési paramétereket, és adattömböket (DomainSet, RangeSet) jelölik. A stilizált kocka a terítési komponens geometriáját leíró „box” elemeket szimbolizálja.



38. ábra Geofizikai mérések összeállítása a GGDM adatmodell alapelemeiből

Gravitációs mérési adatok

Az első példa egy gravitációs állomás adatait mutatja be. Az üres *metaadat* elem csak az eléréshez szükséges azonosító kódot tartalmazza. A mérési adatok tárolására a *ME_Measurement* elem szolgál. Az állomás koordinátáit a *localCRS* elemben két különböző koordináta rendszerben is (EOV és WGS84) megadtuk. Magukat a mérési adatokat a

parameterSet tartalmazza. A szövegben a két paraméternek (MGH50, TTC) csak az értékét találjuk. A rájuk vonatkozó típusdefiníciók a metaadat rekordban hivatkozott paraméterkatalógusban vannak.

```
<geophObject>
  <dataProvider>http://www.elgi.hu</dataProvider>
  <identifierName>GO GRAV 58014</identifierName>
  <parentIdentifier>GOS GRAV jano976</parentIdentifier>
  <objectType>gravimetry</objectType>
  <metadata uuid="MD_elgi.hu_GO_GRAV_58014"/>
  <ME Measurement uuid="ME elgi.hu_GO_GRAV_58014">
    <identifierName>GO_GRAV_58014</identifierName>
    <localCRS>
      <origin>
        <srsName>EPSG:23700</srsName>
        <GG_CartesianCRS>
          <x>532918</x>
          <y>39563</y>
        </GG_CartesianCRS>
      </origin>
      <origin>
        <srsName>EPSG:4326</srsName>
        <GG_SphericCRS>
          <latitude>17.5101704</latitude>
          <longitude>46.975816</longitude>
        </GG_SphericCRS>
      </origin>
      <elevation>
        <verticalDatum>BalticSea</verticalDatum>
        <elevation>234.5</elevation>
      </elevation>
    </localCRS>
    <parameterSet uuid="PS_1_elgi.hu_GO_GRAV_58014">
      <parameter>
        <parameterCode>MGH50</parameterCode>
        <GG_ParameterDouble>
          <value>751652.0</value>
        </GG_ParameterDouble>
      </parameter>
      <parameter>
        <parameterCode>TTC</parameterCode>
        <GG_ParameterDouble>
          <value>428.0</value>
        </GG_ParameterDouble>
      </parameter>
    </parameterSet>
  </ME Measurement>
</geophObject>
```

Típusdefiníciók a PCAT_GRAV_ELGI paraméterkatalógusból:

```
<parameterType>
  <parameterCode>MGH50</parameterCode>
  <parameterName>MGH50</parameterName>
  <unitOfMeasure>microGal</unitOfMeasure>
  <dataType>double</dataType>
  <description>ELGI style measured gravity field. Reference year
1950</description>
  <parameterSet>
    <parameter>
      <parameterCode>DTM GR</parameterCode>
      <GG_ParameterString>
        <value>Potsdam</value>
      </GG_ParameterString>
    </parameter>
  </parameterSet>
```

```

        </parameter>
      </parameterSet>
    </parameterType>
  </parameterType>
  <parameterType>
    <parameterCode>TTC</parameterCode>
    <parameterName>TotalTopoCorrection</parameterName>
    <unitOfMeasure>microGal</unitOfMeasure>
    <dataType>double</dataType>
    <description>Manual topographic correction</description>
  </parameterType>

```

VESZ szondázási adatok

A második példa egy VESZ szondázás leírását mutatja be. A teljes metaadat rekordot közöljük, de a mérési adatsorból hely hiányában csak a 2 m-es és a 20 m-es MN-el mért sorozat első 4-4 elemét mutatjuk meg. A tömörebb kódolás érdekében 1D virtuális terítési komponenseket használunk (ld. 3.1.4 fejezet). A tápelektroda távolságok és a látszólagos fajlagos ellenállás értékek két egymást követő *layout* elemben foglalnak helyet. Egy ortodox leírásban minden táp- és mérőelektroda pár (AB-MN) külön terítésbe kerülne. VESZ mérések esetén az azonos MN-el mért sorozatok összevonása megfelel az általános gyakorlatnak.

```

<geophObject>
  <dataProvider>http://www.elgi.hu</dataProvider>
  <identifierName>GO VES kisalf0001</identifierName>
  <parentIdentifier>GOS VES kisalf</parentIdentifier>
  <objectType>VES</objectType>
  <metadata uuid="MD_elgi.hu-GO_VES_kisalf0001">
    <hierarchyLevel>geophObject</hierarchyLevel>
    <language>EN</language>
    <characterSet>utf8</characterSet>
    <metadataContact uuid="custodian=sores@elgi.hu">
      <person>László Sörös</person>
      <organization>ELGI</organization>
      <role>custodian</role>
    </metadataContact>
    <dateStamp>2009-08-26</dateStamp>
    <referenceSystemInfo>EPSG:23700</referenceSystemInfo>
    <identificationInfo>
      <citation>
        <title>
          <value>GO_VES_kisalf0001</value>
        </title>
        <date>
          <dateType>measurement</dateType>
          <theDate>1986-09-09</theDate>
        </date>
        <responsibleParty uuid="operator=csapo_i.@elgi.hu">
          <person>Csapó I.</person>
          <role>operator</role>
        </responsibleParty>
        <presentationForm>documentDigital</presentationForm>
      </citation>
      <abstract>
        <value>VES-IP sounding</value>
        <translation>
          <language>HU</language>
          <value>VESZ-GP szondázás</value>
        </translation>
      </abstract>
      <status>completed</status>
      <extent>

```

```

        <geographicIdentifier>
          <code>Ostffyasszonyfa</code>
        </geographicIdentifier>
        <geometry>
          <gml:Point gml:id="PNT-GO VES kisalf0001">
            <gml:pos>0.00 0.00</gml:pos>
          </gml:Point>
        </geometry>
        <boundingBox>
          <west>0.00</west>
          <east>0.00</east>
          <south>0.00</south>
          <north>0.00</north>
        </boundingBox>
      </extent>
      <language>HU</language>
      <characterSet>8859part2</characterSet>
    </identificationInfo>
    <instrumentation>
      <name>DIAPIR E-4</name>
    </instrumentation>
    <header>
      <GG_HeaderVES>
        <ARR_TYPE>schlumberger</ARR_TYPE>
        <AB_MIN>6.35</AB_MIN>
        <AB_MAX>1625.50</AB_MAX>
        <AZM>0.00</AZM>
      </GG_HeaderVES>
    </header>
  </metadata>
  <ME Measurement uuid="ME elgi.hu-GO VES kisalf0001">
    <identifierName>GO_VES_kisalf0001</identifierName>
    <localCRS>
      <origin>
        <srsName>EPSG:23700</srsName>
        <GG_CartesianCRS>
          <x>0.00</x>
          <y>0.00</y>
        </GG_CartesianCRS>
      </origin>
      <elevation>
        <verticalDatum>AdriaticSea</verticalDatum>
        <elevation>0.00</elevation>
      </elevation>
      <azimuth>0.00</azimuth>
    </localCRS>
    <layout uuid="LO_0_elgi.hu_GO_VES_kisalf0001">
      <i>1</i>
      <layoutComponent uuid="elgi.hu-GO_VES_kisalf0001_0">
        <i>1</i>
        <comment>Apparent resistivity series, MN=2.00</comment>
        <deviceType>virtualDevice</deviceType>
        <layoutComponentType>sensor</layoutComponentType>
        <box>
          <size>
            <x>2.00</x>
          </size>
          <orientation>
            <x>1.0</x>
          </orientation>
        </box>
        <dimensions>1</dimensions>
        <domainSet>
          <i>1</i>
          <sequence uuid="SQ 0 elgi.hu-GO_VES_kisalf0001">
            <elements>12</elements>
            <parameterCode>AB DIST</parameterCode>
            <GG IrregularSequence>
              <sequenceData>
                <i>1</i>
                <value>6.35</value>
              </sequenceData>
              <sequenceData>
                <i>2</i>
                <value>8</value>
              </sequenceData>
              <sequenceData>
            </sequenceData>
          </GG IrregularSequence>
        </domainSet>
      </layoutComponent>
    </layout>
  </ME Measurement>

```

```

                <i>3</i>
                <value>10.08</value>
            </sequenceData>
            <sequenceData>
                <i>4</i>
                <value>12.7</value>
            </sequenceData>
        </GG_IrregularSequence>
    </sequence>
</domainSet>
<recording>
    <startTime>1986-09-09T00:00:00.0</startTime>
    <endTime>1986-09-09T00:00:00.0</endTime>
    <rangeSet uuid="RS 0 elgi.hu-GO_VES_kisalf0001">
        <parameterCode>APP RES</parameterCode>
        <dataElements>12</dataElements>
        <dataElement>
            <i>1</i>
            <mask>true</mask>
            <GG MeasDataDouble>
                <value>13.3</value>
            </GG MeasDataDouble>
        </dataElement>
        <dataElement>
            <i>2</i>
            <mask>true</mask>
            <GG MeasDataDouble>
                <value>14.6</value>
            </GG MeasDataDouble>
        </dataElement>
        <dataElement>
            <i>3</i>
            <mask>true</mask>
            <GG MeasDataDouble>
                <value>16.6</value>
            </GG MeasDataDouble>
        </dataElement>
        <dataElement>
            <i>4</i>
            <mask>true</mask>
            <GG MeasDataDouble>
                <value>19.8</value>
            </GG MeasDataDouble>
        </dataElement>
    </rangeSet>
</recording>
</layoutComponent>
<layoutComponent uuid="elgi.hu-GO_VES_kisalf0001_1">
    <i>2</i>
    <comment>Apparent resistivity series, MN=20.00</comment>
    <deviceType>virtualDevice</deviceType>
    <layoutComponentType>sensor</layoutComponentType>
    <box>
        <size>
            <x>20.00</x>
        </size>
        <orientation>
            <x>1.0</x>
        </orientation>
    </box>
    <dimensions>1</dimensions>
    <domainSet>
        <i>1</i>
        <sequence uuid="SQ_1_elgi.hu-GO_VES_kisalf0001">
            <elements>9</elements>
            <parameterCode>AB DIST</parameterCode>
            <GG IrregularSequence>
                <sequenceData>
                    <i>1</i>
                    <value>64</value>
                </sequenceData>
                <sequenceData>
                    <i>2</i>
                    <value>80.63</value>
                </sequenceData>
                <sequenceData>
                    <i>3</i>

```

```

                <value>101.59</value>
            </sequenceData>
            <sequenceData>
                <i>4</i>
                <value>128</value>
            </sequenceData>
        </GG_IrregularSequence>
    </sequence>
</domainSet>
<recording>
    <startTime>1986-09-09T00:00:00.0</startTime>
    <endTime>1986-09-09T00:00:00.0</endTime>
    <rangeSet uuid="RS_1_elgi.hu-GO_VES_kisalf0001">
        <parameterCode>APP RES</parameterCode>
        <dataElements>9</dataElements>
        <dataElement>
            <i>1</i>
            <mask>>false</mask>
            <GG MeasDataDouble>
                <value>69</value>
            </GG MeasDataDouble>
        </dataElement>
        <dataElement>
            <i>2</i>
            <mask>>false</mask>
            <GG MeasDataDouble>
                <value>71.1</value>
            </GG MeasDataDouble>
        </dataElement>
        <dataElement>
            <i>3</i>
            <mask>>false</mask>
            <GG MeasDataDouble>
                <value>70.1</value>
            </GG MeasDataDouble>
        </dataElement>
        <dataElement>
            <i>4</i>
            <mask>>true</mask>
            <GG MeasDataDouble>
                <value>63.2</value>
            </GG MeasDataDouble>
        </dataElement>
    </rangeSet>
</recording>
</layoutComponent>
</layout>
</ME Measurement>
</geophObject>

```

Rétegmodell, fúrási rétegsor

A harmadik példa a rétegmodell használatára mutat be egy különleges példát. A geofizikai inverzióval meghatározott 1D modellek tárolására kialakított szerkezet fúrási rétegsorok tárolására is tökéletesen alkalmas. Az alábbiakban – szintén metaadatok nélkül – egy csonkított rétegsor látható. A mélység és vastagság adatok mellett rétegparaméterként a MÁFI rétegekódjaival megadott formáció-azonosító szerepel.

```
<geophObject xmlns:gml="http://www.opengis.net/gml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="C:\Documents\geomind\wp9\ggdm\schema\ggdm.xsd">
  <dataProvider>http://www.elgi.hu</dataProvider>
  <identifierName>GO_BH_preter-205</identifierName>
  <objectType>borehole</objectType>
  <metadata uuid="MO_elgi.hu_GO_BH_preter-205"/>
  <MO Model>
    <identifierName>GO BH_preter-205</identifierName>
    <tag>pretercier</tag>
    <localCRS>
      <origin>
        <srsName>EPSG:23700</srsName>
        <GG CartesianCRS>
          <x>123345</x>
          <y>654543</y>
        </GG_CartesianCRS>
      </origin>
      <elevation>
        <verticalDatum>BalticSea</verticalDatum>
        <elevation>159.16</elevation>
      </elevation>
    </localCRS>
    <MO LayerModel>
      <layers>3</layers>
      <layer>
        <i>1</i>
        <top>0</top>
        <thck>1</thck>
        <property>
          <fixed>>false</fixed>
          <parameter>
            <parameterCode>LITHO_MAFI</parameterCode>
            <GG_ParameterString>
              <value>Qh2talaj</value>
            </GG_ParameterString>
          </parameter>
        </property>
      </layer>
      <layer>
        <i>2</i>
        <top>1</top>
        <thck>18</thck>
        <property>
          <fixed>>false</fixed>
          <parameter>
            <parameterCode>LITHO_MAFI</parameterCode>
            <GG_ParameterString>
              <value>f Qp</value>
            </GG_ParameterString>
          </parameter>
        </property>
      </layer>
      <layer>
        <i>3</i>
        <top>19</top>
        <thck>166</thck>
```

```
<property>
  <fixed>false</fixed>
  <parameter>
    <parameterCode>LITHO_MAFI</parameterCode>
    <GG ParameterString>
      <value>edPal</value>
    </GG_ParameterString>
  </parameter>
</property>
</layer>
</MO_LayerModel>
</MO_Model>
</geophObject>
```

3. Függelék, XML

Mi az XML?

XML az eXtensible Markup Language kifejezés rövidítéséből származik. Nem is annyira egy konkrét nyelv, hanem inkább egy filozófia megtestesítője. Strukturált rendszerek általános leírására szolgál. A hangsúly a strukturált és az általános jelzőkön van, ennek köszönheti óriási sikerét és egyre növekvő jelentőségét az informatikában. Különösen a hálózati adatsere területén elterjedt. A markup nyelvek területén (semmitmondó magyar fordításban jelölőnyelv) számos elődje volt (SGML, HTML). A legismertebb közülük a HTML (HyperText Markup Language). Bár nem követeli meg az XML-el szemben támasztott szigorú formai elvárásokat, lényegében az XML egy változatának tekinthető. (A XML elvárásoknak formailag is megfelelő HTML változat az XHTML.) Túllépve elődei gyengéin 1998-ban a W3C konzorcium először kiadta az XML szabvány specifikációját (W3C 2008). A nyelv szerkezete első megközelítésben igen egyszerű, néhány következetesen betartandó alapelvre épül. Legfontosabb építő kövei az elemek és attribútumok. Az elemek egymásba ágyazható szerkezetek, amelyek nyitó és záró címkékből (tag), opcionális attribútumokból, és tartalomból állnak. Példa egy XML elemre:

```
<book>Lord of the Rings</book>
```

<book> a nyitó tag, </book> a záró tag, a „Lord of the Ring” pedig a tartalom. Az elemek lehetnek üresek is, amit kétféleképpen jelölhetünk:

```
<book></book>
```

vagy

```
<book/>
```

Egymásbaágyazhatóak:

```
<book>  
  <title>The Lord of the Rings</title>  
  <author>J.R.R. Tolkien</author>
```

```
</book>
```

Itt a <title> és az <author> elem a <book> részeként jelenik meg. Az attribútumokat a nyitó tag-ben helyezhetünk el.

```
<title lang="en">The Lord of the Rings</title>
```

A „lang” attribútum értéke „en”. Egy elem bármennyi attribútumot és bármennyi befoglalt elemet tartalmazhat, de az elemek nem lapolódhatnak át. A formai szabályoknak megfelelő XML dokumentumot jól formázottnak (well-formed) nevezzük. Még egy formai szabály, amely betartása kötelező: Minden XML dokumentum deklarációval indul, ami számos információt tartalmazhat, mint pl. az XML verzió számát, és a karakter kódolás típusát.

```
<?xml version="1.0" encoding="ISO-8859-2"?>
```

További XML építőkövek:

Megjegyzés:

```
<!--comment -->
```

Feldolgozási parancs:

```
<? cat furasok.txt | grep koord | awk -F; '{print $1, $2, $3}' ?>
```

A <? és ?> jelek közötti szöveget a feldolgozó program nem az XML tartalom részének tekinti, hanem parancsnak, és végrehajtja azt.

Névterek (Namespace):

```
<book
  xmlns="http://talkien.hu"
  xmlns:my="http://mynamespace.hu"
  xmlns:your="http://yournamespace.hu">
  <title lang="en">The Lord of the Rings</title>
  <title lang="hu">A Gyűrűk Ura</title>
```

```
<author>J.R.R. Tolkien</author>
<price curr="EUR">14.98</price>
<my:opinion>excellent</my:opinion>
<your:opinion>too long</your:opinion>
</book>
```

Az xmlns:my egy névteret definiál. A megadott címnek kizárólag mint egyedi azonosítónak van jelentősége. A my névtérbe a my:opinion elem beletartozik. A névterek használatának értelme, hogy azonos nevű, de mást jelölő elemek nem ütköznek. Pl., a your:opinion nem keverhető össze a my:opinion elemmel. A dokumentum alapértelmezésbeli névterét az xmlns adja meg.

CDATA:

```
<![CDATA [
ASD8789SDVVCU89SDF789SDF489GFHIUOWF0W8UFVWE789SD
SDFJKLWER9804U78TRHU208I2RVNJOEFJI042U84RHJIUFNRN
348FRWHUWJI348U934TU8343HFRWHU29834H8U3FJOW9834R
ASD8789SDVVCU89SDF789SDF489GFHIUOWF0W8UFVWE789SD
]]>
```

A CDATA struktúra tetszőleges karaktersor XML-ben történő elhelyezésére szolgál. Akkor van rá szükség, ha a szöveget az XML értelmezés hatálya alól ki akarjuk vonni.

XML-re épülő technikák, eszközök

Az XML rohamos fejlődésének csak egyik oka, hogy gyakorlatilag bármilyen objektum szöveges kódolására alkalmas. Nemcsak az internetes adatcserében gyakori, de felhasználói programok ki és bementi formátumaként, programrendszerek konfigurációs fájljaiként, és önálló szakterületeket átfogó leírónyelvek hordozójaként is egyre terjed. A siker másik oka az XML-re épülő hatékony technológiákban keresendő.

DOM

A Document Object Model egy programnyelv független interfész, ami az XML dokumentumok kezeléséhez egy szabványos eszközkészletet biztosít a programozóknak. A DOM segítségével az XML dokumentumok elemei a számítógép memóriájában

hozzáférhető és manipulálható. A DOM reprezentáció az XML dokumentumot csomópontokból álló fa struktúrában képezi le. A csomópontoknak 12 típusa van, ezek részben megegyeznek a bevezetőben említett XML építő elemekkel. A csomópontokon keresztül a dokumentumok tartalma módosítható, elemek, attribútumok lekérdezhetők, megszámlálhatók, törölhetők, másolhatók stb. Bár a DOM az XML dokumentumok viszonylag alacsony szintű manipulációjára való, támogatja az XPATH kifejezések használatát, az érvényesítést (validation) és a XSL transzformációkat is (ld. bővebben később). A legtöbb modern programnyelv függvénykönyvtárakkal támogatja a DOM használatát.

SAX

A teljesség kedvéért meg kell említeni a SAX (Simple Application Interface for XML) szövegelemzőt (parser) is. A SAX a DOM népszerű alternatívája. Szemben a DOM-mal, ahol a teljes fa struktúra felépül a memóriában, ez soros elérést tesz lehetővé. A SAX szövegértelmező az egyes XML elemek beolvasásakor eseményeket generál, amelynek hatására az eseményekhez kapcsolt feldolgozási lépések végrehajtnak. Nagyméretű fájlok feldolgozására való igen gyors, kis erőforrás igényű, viszonylag egyszerű eszköz.

XPATH

Az XML egyik igen kellemes tulajdonsága, hogy a csomópontok (elemek, attribútumok) a legbonyolultabb szerkezetekben is hatékonyan kereshetők. Minden XML elem elérési útvonallal (path) rendelkezik. Az XPATH szintaxisa a unix fájl-elérési útvonalakhoz hasonló formát követ. Néhány példa egyszerű elérési útvonalakra:

```
/book/author          J.R.R. Tolkien
/book/my:opinion      excellent
```

Attribútumok kiválasztására @ karakter szolgál, amelyet az attribútum neve elé kell tenni:

```
/book/price/@curr     EUR
```

A fenti példák abszolút hivatkozások. Relatív hivatkozásoknál az aktuális csomóponthoz képest történik a kiválasztás. Pl., szülő kiválasztása:

```
/book/price/..          /book
```

Az aktuális csomóponttól lefelé kiválaszthatjuk az összes adott nevű csomópontot:

```
/book//title           The Lord of the Rings, A Gyűrűk Ura
```

Specifikus csomópontok kiválasztása pozíció, vagy érték alapján is lehetséges. Az ún. predikátumok (predicates) és relatív hivatkozások kombinációjával igen rugalmas keresések hajthatók végre. Keressük a második cím elemet:

```
/book//title[2]       A Gyűrűk Ura
```

Keressük azt a könyvet, amelynek címe "The Lord of the Rings":

```
//*[title="The Lord of the Rings"] /book
```

Keressük a könyv magyar nyelvű címét, vagyis azt a title elemet, amelynek lang attribútuma "hu":

```
/book//title[@lang="hu"]      A Gyűrűk Ura
```

Keressük a "The Lord of the Rings" című könyv erőben megadott árát:

```
//*[title="The Lord of the Rings"]/price[@curr="EUR"]      14.98
```

Az XPath nyelv sok XML feldolgozó technika alapja. Specifikációja a W3C konzorcium honlapján olvasható (W3C, 2007a)

XSLT

A rövidítés az „Extensible Stylesheet Language Transformation” kifejezésből származik. Az XSL egy XML alapú nyelv, amely XML dokumentumok transzformációjára szolgál (W3C, 2007b). Segítségével XML fájlok tetszőleges formátumba átkonvertálhatók. Gyakori alkalmazási területe az XML fájlok formázott megjelenítése HTML, PDF, vagy egyéb formában, de ennél sokkal többre is képes. Az XSLT egyfajta szűrőnek tekinthető, ami az

XML dokumentumokat képes bármilyen rendszer bemenetéhez illeszteni. Ezt használja ki pl., az AJAX Internet technológia.

XSD

Az XSD jelentése: XML séma definíció (XML Schema Definition). Ez a szintén XML alapú nyelv lehetővé teszi, hogy meghatározzuk egy XML állomány szerkezeti felépítését. Ezáltal ellenőrizhetővé válik, hogy egy fájl formailag, szerkezetileg megfelel-e az elvárásoknak. Ha igen, akkor az XML fájl érvényesített (validated). Az XML kezelő függvénykönyvtárak hatalmas eszközkészlettel rendelkeznek, többek közt az érvényesítés (validáció) támogatására. Ha az érvényesítés hiba nélkül lefut, biztosak lehetünk benne, hogy adataink formailag helyesek, azaz, a sémának megfelelnek. Tartalmi ellenőrzésre is van mód, erre a Schematron szolgál, ami egy speciális eszköz XPATH kifejezésekkel megadott adatokon végzett műveletek eredményének ellenőrzésére. Az XSD a korábban használt DTD-vel (Document Type Definition) leírással szemben egyre inkább túlsúlyba kerül, és az XML alapú új leíró nyelvek elsődleges eszközévé válik. A GEOMIND profil és a GGDM implementációja is XSD-re épül.

AJAX

A rövidítés az „Asynchronous JavaScript and XML” kifejezésből származik. A szerver – kliens együttműködés során, Az adatok a szervertől a kliens kérésére XML formátumban érkeznek. A böngésző felügyelete alatt futó JavaScript program XSL transzformációval átalakítja az adatokat a megfelelő formátumra, ami ezután közvetlenül hasznosítható a böngésző számára. Ezt a technikát használja a KINGA portál is a metaadatok kezelése közben.

Az XML alkalmazása és fejlődési irányai

Az XML legfontosabb jellemzője, hogy a szöveg az adatok mellett magától értetődő módon tartalmazza a szerkezetet is, ezért kiválóan alkalmas strukturált objektumok leírására. Ennek köszönhető, hogy meghódított minden olyan területet, ahol az objektum orientált megközelítés hatékonyan működik.

Szakmai leírónyelvek

Az elmúlt években az XML-ből a különböző szakterületeket leíró nyelvek sora fejlődött ki, és ez a tendencia robbanásszerűen terjed. Ennek oka, hogy egy adott szakterületre nézve specifikus adatmodellből származtatott szerkezeti séma olyan leírónyelvet definiál, ami a szakértők számára magától értetődő fogalmakkal és szakkifejezésekkel operál. Az egyes elemek kapcsolatai szintén a szakterület ismert összefüggéseit képezik le. Egy ilyen szerkezeti sémára épülő XML adatfájl a szakterület ismerője számára azonnal értelmezhető (leszámítva persze az XML alapjainak megismerésére fordítandó időt). Ugyanakkor meg kell jegyezni, hogy az XML általában nem emberi fogyasztásra készül, viszont az XML szövegek könnyedén konvertálhatók jól olvasható és tetszetős HTML formátumra, sőt, az SVG (Scalable Vector Graphics) nyelvnek köszönhetően grafikus formába is.

Szabványos alapsémák

Jellemző, hogy az alkalmazott szakterületeket leíró nyelvek általános leírónyelvek sémáira épülnek. Ilyen általános leírónyelv például a GML (Geography Markup Language) amely térbeli objektumok geometriájának szabatos leírására való. A MathML tetszőlegesen bonyolult matematikai kifejezések XML kódolását teszi lehetővé. A SensorML-t a mérő és feldolgozó rendszerek leírására, a SWE-t (Sensor Web Enablement) az adatgyűjtő rendszerek összekapcsolására és a megfigyelési adatok hálózaton keresztül történő továbbítására fejlesztették ki. Nagyon erős a tendencia, amely az alapsémák szabványosításának irányába mutat. Az XML szabványosítás folyamatát az OGC (Open Geospatial Consortium) és az ISO/TC-211 technikai bizottság (Technical Committee ISO/TC-211, Geographic information/Geomatics.) koordinálja. Ezeket az alapsémákat az ISO 191-es sorozatba tartozó szabványok foglalják össze. Az alapsémák integrációjával jött létre az O&M (Observation and Measurements) nyelv, ami tetszőleges észlelés vagy mérés leírására alkalmas szabvány. A jövőben geofizikai mérések leírásában is fontos szerepet kaphat. Az alapsémákat és az O&M-et is magába integráló magas szintű leírónyelvek egyik legmarkánsabb példája a GeoSciML (Geoscience Markup Language), a földtudományi leírónyelv. Ez a geológiai szerkezetek, földtani egységek, kőzetek leírására szolgál, beleértve azok geometriáját, korát, anyagát, fizikai tulajdonságait. A szakmai terminológia integrálása a szintén szabványos sémákra épülő szemantikai szótárak és többnyelvű teauruszok segítségével történik.

Szabványos WEB szervizek

Az adatrepresentáció mellett az XML másik kiemelt alkalmazási területe a hálózati kommunikáció. Az interoperabilitásban a szabványos adatformátumok mellett a szabványos adatszolgáltatásnak szintén fontos szerepe van. A WEB szolgáltatások, vagy szervizek (web services) szabványosítása szintén XSD sémák alapján történik. Ezek meghatározzák, hogy egy adott hálózati kiszolgáló milyen funkciókkal van felruházva, egy adott szolgáltatás hogyan érhető el, milyen paramétereket és hogyan kell megadni, hogy a kérés teljesíthető legyen. A WEB szervizek alkalmazási területe gyakorlatilag korlátlan. A földtudományokban legjellemzőbb alkalmazások a különböző WEB GIS szolgáltatások (WMS, WFS, WCS, CS-W), és az adattranszformációs WEB szervizek.

Az XML kiterjedt alkalmazásának látványos eredménye a hálózati integráció. A közös adatmodellnek és egységes adatformátumnak köszönhetően az egymástól elszigetelt rendszerek a korábbiakhoz képest sokkal hatékonyabban képesek együttműködni. A nyilvános metaadat katalógusokban fellelhető elérhetőségi adatok a távoli adatforrásokhoz könnyű hozzáférést biztosítanak. A rendszerek átláthatóságának nem feltétele, hogy a belső adatrepresentáció is azonos legyen. Elegendő a nyilvánosan elérhető adatszolgáltatások szintjén biztosítani a közös nyelvi környezetet. Gyakori megoldás, hogy a szakmai adatbázisok és az internetes adatszolgáltató alkalmazások között konverziós modulok (pl. transzformációs WEB szervizek) biztosítják a kapcsolatot. (Az persze elengedhetetlen, hogy a közös modell minden kulcs eleme valamilyen módon reprezentálva legyen a belső adatbázisban is)

INSPIRE

Az eddig elszigetelt rendszerek közötti határok elmosódásának eredményeként elindult, és rohamos ütemben halad a nemzetközi adatintegráció. Az Európai Unió által támogatott INSPIRE (Infrastructure for Spatial Information in Europe) program is ezt bizonyítja. Ennek célja egy határokon átívelő egységes európai térinformatikai infrastruktúra kialakítása, amely lefedi az élet számos területét a népegészségügyi adatoktól a földhasználatig, a meteorológiától a földtanig, vagy a közlekedési hálózatoktól a természeti kockázati zónáig. Az állami téradat szolgáltató intézmények informatikai rendszereinek harmonizálásáról van

szó, amely lehetővé teszi, hogy hasonló jellegű nyilvános adatokat az unió intézményei, vagy állampolgárai bármely európai országban egységes rendszeren keresztül egységes formában kaphassanak meg. Fontos szempont, hogy a párhuzamos adattárolás megszűnjön, és az információ elérése, kezelése, karbantartása hatékonyabbá és olcsóbbá váljon. A formai harmonizáció mellett az egyes témák területén jelenleg zajló adatspecifikáció célja a közös adatmodelleken keresztül a tartalmi egység megteremtése. Ez a munka a tematikus munkacsoportokban zajlik. (A geofizikai adatmodellel kapcsolatos munkám elismeréseként abban a megtiszteltetésben van részem, hogy részt vehetek a geológia munkacsoport tevékenységében, ahol feladatom a geofizikai objektumok INSPIRE modellbe történő integrálása)

A rendszer kiépítése után az INSPIRE belépési pontok a geo-portálok lesznek. Ezek számos funkciót látnak el, pl.: keresés (data discovery) megjelenítés (viewing services) letöltés (download), szerkesztés (data editing). E funkciók ellátása a legújabb informatikai technikák alkalmazását teszik szükségessé, és komoly kihívások elé állítják az adatszolgáltatókat.