

**MIKOVINY SÁMUEL FÖLDTUDOMÁNYI DOKTORI ISKOLA**

**Doktori értekezés tézisei**

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

**Írta:  
SŐRÉS LÁSZLÓ**

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

**Miskolci Egyetem  
Geofizikai Tanszék  
Miskolc  
2011.**

## **Absztrakt**

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 (Geoscience Markup Language) 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, módszerfüggetlen 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 (Extendable Markup Language) 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.

## **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 (Extendable Markup Language) 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.

## ***I. Az értekezés célkitűzései***

A geofizikai adatrendszerek heterogenitása nagyon megnehezíti az egységes elvek szerint működő információs rendszerek kialakítását. Az alapvető problémát leginkább az jelenti, hogy hiányzik a mérési módszerek és a hierarchikus adatrendszerek egységes leírására alkalmas általános adatmodell. Az elterjedt ipari adatszabványok túlságosan specifikusak, főleg az olajiparban használt módszerekre korlátozódnak (szeizmika, mélyfúrás-geofizika), és hiányzik belőlük az általános megközelítés igénye. Jelentős változás kezdődött el a GeoSciML földtudományi leírnyelv fejlesztésével, amely az alapokról indulva, szabatosan építi fel a méréseket, megfigyeléseket (OM, Observations & Measurements) (OGC 2007a), szenzorokat (SensorML) (OGC 2007b) reprezentáló modelleket. Jelenleg a geofizika még nem esik a nyelvet fejlesztő team érdeklődési körébe. Doktori értekezésem alapvető célja egy általános geofizikai adatmodell (GGDM, General Geophysical Data Model) megalkotása, és egy egységes elven felépülő geofizikai információs rendszer alapjainak lerakása. Egy ilyen rendszerrel szemben támasztott követelmények a következők:

- Adjon keretet a rendkívül polimorf geofizikai adatrendszerek minél átfogóbb leírásához
- Tegye lehetővé az adatrendszerek hierarchiájának leképezését
- A metaadatok kiterjedt használata révén javítsa az adatok fellelhetőségét, támogassa az internetes kereső szolgáltatásokat
- Az általános elvekre épülő adatmodell és az XML technológia révén könnyítse meg az adatcserét és az elterjedt ipari formátumok közötti a konverziókat

- Biztosítsa a leíró, mérési és eredmény adatok hatékony tárolását és a rugalmas lekérdezéseket
- Segítse a magas szinten integrált elosztott adatbázisok létrehozását
- Támogassa a különböző geofizikai módszerek együttes inverziójából származó több paraméteres modellek leírását

Ezek az elvárások megkövetelik egy általános geofizikai adatmodell megalkotását, amelyben helye van a leíró információknak, mérési, feldolgozási adatoknak, és az értelmező munka során megszülető földtani modelleknek. Egy nemzetközi és egy hazai metaadat projekt eredményein keresztül bemutatom a geofizikai metaadatok használatát. Céлом volt megvizsgálni az általános adatmodellre épülő adatbázis megvalósításának lehetőségeit, és bemutatni annak különböző módjait. Ezen elvekből kiindulva, az Eötvös Loránd Geofizikai Intézet adatrendszereihez és igényeihez igazodva egy konkrét adatbázis alkalmazás megvalósításával szándékozom bemutatni a felvázolt rendszerben rejlő lehetőségeket.

## **II. Tudományos előzmények**

### **A GAIA**

A geoinformatika terén végzett munkám hajtóereje az 1984 óta geofizikusként gyűjtött adatkezelői és adatfeldolgozói tapasztalatom volt. A dolgozatomhoz kapcsolódó kutatás közvetlen előzménye az Eötvös Loránd Geofizikai Intézetben kifejlesztett GAIA információs rendszer kiépítése volt. Ez egy VESZ és tranziens adatok tárolását, feldolgozását és megjelenítését segítő rendszer, amely jelenleg az Országos Geoelektromos Adatbázis (OGA) szoftveres bázisa. A GAIA fejlesztés 1997-ben kezdődött, 1998-tól kezdve fogadta az adatokat. Mára 42000 VESZ és 5500 tranziens szondázást tartalmazó informatikai rendszerre terebélyesedett. A GAIA rendszer három szervesen egymásra épülő szintből áll. Ezek a saját fejlesztésű adatbázis motor, az adatbázis kezelő köztes alkalmazások, és a grafikus felhasználói felületek. A moduláris felépítésű rendszer nagyfokú rugalmasságot és hatékony feldolgozói munkát tesz lehetővé. A GAIA kiterjedt használata során körvonalazódtak annak korlátai, és egyben a továbblépés iránya. A térinformatikai funkciók bővítése, a hálózati elérhetőség javítása, az egyéb geofizikai módszerek bevonása és az adatstruktúrák szabványosítása olyan elvárások voltak, amelyek túlmutattak a GAIA keretein. A WEB technológiák fejlődése, a térinformatikai szabványok terjedése, az XML alapú leírónyelvek sora egy sokkal általánosabban megfogalmazott rendszer irányába mutatott. 2006-ban indult el a GEOMIND projekt, amely alapvető fordulatot hozott a dolgozatomban felvázolt adatmodell és geofizikai információs rendszer megalkotása szempontjából. Egy kulcsfontosságú munkacsomag vezetőjeként lehetőségem volt részt venni egy nemzetközi együttműködésben létrehozott geofizikai metaadat portál kialakításában.

### **Metaadat szabványok fejlődése**

A metaadatok fogalma a könyvtárakkal egyidős. A digitális könyvtárak és az internetes keresőrendszerek megjelenésével a metaadat fogalma az informatikai köztudatba is berobbant. A térinformatikai rendszerek révén, főleg a térképi adatrendszerek jellemzésére a földtudományok körében is terjed a használatuk. Számos nemzetközi szabvány jött létre, melyek használata főleg az őket támogató szoftverek népszerűségétől függ. A valószínűleg leggyakrabban használt szabvány az amerikai Szövetségi Földrajzi Adat Társaság nevével fémjelzett FGDC (Federal Geographic Data Committee, 1998). Szabatoságát, és általánosságát

tekintve a szabványok sorából kiemelkedik az ISO 19115 (ISO/TC 211 2003). XML implementációja, az ISO 19139 az OGC (Open Geospatial Consortium) által támogatott rendszereknek alapvető komponensévé vált. Dolgozatomban ezért választottam a geofizikai adatrendszerek magas szintű leírásának eszközeként.

### **A térinformatikai szabványok, WEB szolgáltatások fejlődése**

A világhálón elérhető térképi információk nyílt megosztásának igénye kikényszerítette a térinformatikai szabványok terjedését. Az OGC által támogatott szabványok használata minden eddiginél magasabb szintű integrációt tesz lehetővé. A térinformatikai WEB szolgáltatások és az ezekhez kapcsolódó alapszabványok fejlődése az elmúlt évtizedben töretlen volt. Ezek ma az Európai Unió INSPIRE direktíváiban lefektetett célok megvalósításának alapját jelentik. A gomba módra szaporodó WEB szolgáltatások köréből a téma szempontjából legfontosabb a WMS és a WFS.

A WMS (Web Mapping Service) szabványos protokoll térképi információk Interneten keresztül történő továbbítására. Első verzióját az OGC 2000-ben publikálta. (Open Geospatial Consortium Inc. 2006) A WMS-re épülő térkép megjelenítés lényegesen különbözik a megszokottól. A felhasználó kapcsolata a GIS adatbázissal egy http csatornán folytatott kliens – szerver kommunikációra épül.

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. (Open Geospatial Consortium Inc. 2005a) A WFS séma lehetővé teszi, hogy a térképi adatbázisokból az objektumok attribútumainak, illetve geometriájuknak, kölcsönös térbeli helyzetüknek alapján bonyolult szűréseket végezzünk.

A nagy térinformatikai szoftver gyártók, 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.

### **GeoSciML**

A GeoSciML leírónyelv látványos fejlődése a disszertációban többször említett GEOMIND projekttel nagyjából párhuzamosan indult el. A főleg GML-re (Geography Markup Language) épülő nyelv fokozatosan terjeszkedik a földtudományok különböző területein (Sen, M., Duffyb, T., 2005). 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. Az általam felvázolt adatmodell a GeoSciML-hez hasonló elvekre épül, és átfogó keretet adhat a geofizikai objektumok és objektumcsoportok GML alapú leírásához.

### **III. Az elvégzett vizsgálatok, az alkalmazott módszerek**

A dolgozatban tárgyalt adatmodell és a felvázolt információs rendszer az informatika és a geofizika számos eredményeinek együttes felhasználásával jött létre. Az informatika területéről ide tartozik az adatbázisok elmélete, és gyakorlata, különösen az adatmodellezés, adatbázis kezelés. A megfelelő adatmodellek kidolgozása megkívánja a modellezett környezet, jelen esetben a geofizikai mérések, feldolgozási, inverziós módszerek és adatrendszerek alapos ismeretét. A tervező munka során felhasználtam az UML (Unified Modelling Language) nyelv eszközeit. A GGDM (General Geophysical Data Model) specifikációját XSD sémadefiniációs csomag formájában készítettem el.

A már realizált adatbázisokban az adatok tárolásának, elérésének, a keresések végrehajtásának eszköze az SQL (Sequence Query Language), a relációs adatbázisok nyelve. Mivel a geofizikai adatok térbeli adatrendszerek, ahol a mért és származtatott paraméterek térképi megjelenítése fontos, ezért az adatmodelleknek és az adatbázisoknak támogatniuk kell a térinformatikai rendszerekhez való kapcsolódást. Nélkülözhetetlen az OGC által támogatott alapszabványok tanulmányozása. Ezek közül legfontosabbak a GML (Geography Markup Language) (Open Geospatial Consortium Inc. 2005b), a tér- és időbeli objektumok egzakt leírására szolgáló nyelv, és az ISO 19115, amely földrajzi adatrendszerek metaadat szintű leírásának a szabványa. Mindkettő XSD sémadefiniációkra épülő rendszer, amely számos további szabványt (ISO 19107, ISO 19108, ISO 19117, ISO 19123) integrál magába. Felfelé ezek újabb, még magasabb szintű rendszerek, a térinformatikai webszolgáltatások részét képezik (WMS, WFS). Az építmény alján az Internetes technológiák alapköve, az XML foglal helyet, annak számos eszközével és lehetőségével (XPath, XQuery, XLink, XSLT, stb) Mindezen eszközöket munkám során nemcsak megismerni, de a megfelelő helyen alkalmazni is kellett.

A programozási feladatokhoz az objektumorientált JAVA (Java EE) programozási nyelvet használtam, amelyet támogatja az XML állományok kezelését, az adatbázisok használatát, webalkalmazások fejlesztését, így kiváló eszköz a komplex alkalmazások fejlesztéséhez. Munkám során amikor csak lehetett a Unix operációs rendszer alapelemeire támaszkodó „eszköztelen” munkastílust követtem, és törekedtem a minőségi, nyílt forráskódú, platform független megoldások használatára. Ugyanakkor kiemelek két szoftvert, amely fejlesztő munkámban nagy segítségemre volt: az Altova XMLSpy Professional Edition 2008-as verzióját, amelyet az XML állományok kezeléséhez, és az XSD sémák tervezéséhez, finomításához használtam, valamint a SUN NetBeans 6.1-es programfejlesztő rendszerét, amelyet a Java alkalmazások fejlesztésében alkalmaztam.

Az adatbáziskezelő rendszerekkel kapcsolatos kísérleteimet kezdetben MySQL-lel végeztem, majd áttértem a PostgreSQL használatára, amely az SQL szabvány implementációjában, az XML és geometriai típusú mezők kezelésében jóval fejlettebbnek tekinthető. A WMS-el és WFS-el kapcsolatos munkákban a GeoServer rendszerre hagytam, amely kiválóan együttműködik a PostgreSQL-lel. A térinformatikai tesztek QuantumGIS rendszerrel történtek. Az XML dokumentumtár indexállományainak előállításához az Apache Lucene keresőmotort és az Apache SolR web alkalmazást használtam.

## **V. Új tudományos eredmények (tézisek)**

### **1. Heterogén geofizikai adatrendszerek leírására szolgáló egységes szabvány kidolgozása az ISO 19115 metaadat szabvány alapján**

A földrajzi adatokra vonatkozó metaadat szabvány használata a különböző térinformatikai adatrendszerek esetében elterjedt. A metaadatok leginkább a nyilvánosságnak szánt analóg és digitális térképek tartalmával, elérhetőségével kapcsolatos információkat tartalmazzák. Emellett helyet kapnak a térképi adatrendszereket jellemző legfontosabb tudnivalók is, pl. az adatok elkészítésének módja, minősége, a feldolgozottság foka stb. A gondolat, hogy geofizikai adatrendszerekre is alkalmazzuk az ISO 19115 metaadat szabványt a GEOMIND projekt első koordinátorától, Tomasz Mardaltól a Lengyel Geológiai Szolgálat munkatársától származik. A geofizikai adatrendszerek sokfélesége miatt azonban már a kezdet kezdetén nehezen megválaszolható kérdések merültek fel. Hogyan feleltethetők meg a szabvány térinformatikai ihletésű fogalmai a geofizikában használt fogalmaknak? Hogyan lehet áthidalni a különböző mérési módszerek közötti alapvető különbségeket (pontszerű, szondázó, szelvénymenti, térképező mérések)? Hogyan kezeljük a geofizikai kutatás eredményeit megjelenítő, nem térképi adatokat (szelvények, 1D 2D 3D modellek)? Egyáltalán, mit tekintünk a geofizikai metaadatok alapegységének? A feltett kérdésekre az általam megalkotott fogalmak bevezetésével adható meg a válasz. Munkám során létrehoztam az ezek mögött lévő logikai struktúrákat és ezeket XML séma definíciókba fektettem le.

#### *1. tézis*

***A heterogén geofizikai adatrendszerek egységes leírására új fogalmi rendszert és adatmodellt dolgoztam ki. Ez magába foglalja a geofizikai objektum, geofizikai objektum csoport és a riport fogalmát, valamint ezek adatmodelljét.***

#### **a) A geofizikai objektum fogalma, és adatmodellje**

A szabvány alapértelmezés szerint minden adatszethez (dataset) rendel egy metaadat rekordot. Ez a térinformatikában általában térképi adatrendszert jelent. A térképek a hagyományos GIS hierarchia szerint alakzatokból (feature) állnak, amelyekhez attribútumok formájában mögöttes adatbázis tartalom tartozik. A geofizika térképi alakzatai a mérési és feldolgozási folyamatok különböző szintjeit képviselik. Egy pont alakzat egy geofizikai térképen lehet egy önálló gravitációs mérés, de lehet egy szeizmikus szelvény egyetlen pontja is, mely önmagában nem áll meg. A kettő nem azonos hierarchiaszintet képvisel. A kérdés általános kezelése érdekében bevezettem a geofizikai objektum fogalmát és meghatároztam a hozzá tartozó részletes logikai struktúrát (adatmodell). A geofizikai objektum olyan egyszerű térképi alakzatként megjeleníthető önálló egység, amelyhez metaadat rekord kapcsolódik. Ez a gyakorlatias meghatározás összecseng a hagyományos mérés fogalommal, amely különböző komplexitású dolgokat egyaránt mérésenként kezel. Az egyedi mérések fontosságuknak megfelelően kísérő információkkal láthatók el, a térképeken önálló alakzatként jelennek meg, belső tulajdonságaikat attribútumaik adják meg. A meghatározás magába foglalja a geofizikai modelleket is. Egy rétegmodell, egy 2D szelvény, vagy egy geofizikával lehatárolt 3D test szintén lehet geofizikai objektum. A mérési módszert a geofizikai objektum egy attribútuma, az objektumtípus adja meg. A metaadatok szintjén az eltérő típusú objektumok azonos struktúrával jellemezhetők, csak a fejléc adatokban különböznek. A fejléc (header) a metaadat rekord része, szerkezetét az XSD séma határozza meg.

## **b) A geofizikai objektumcsoport fogalma és adatmodellje**

A geofizikai objektumcsoport definícióm, és az azt leképező adatmodell értelmében összetartozó objektumok halmaza. Az összetartozás minősége szerint beszélünk elsődleges és másodlagos csoportokról. Elsődleges csoportok azok, amelyek a hozzájuk tartozó objektumokkal kizárólagos szülői viszonyban vannak. A mérési kampányok tipikus elsődleges csoportok. Egy adott mérés egy és csakis egy mérési kampányhoz tartozhat, ahhoz, amelyben létrejött. A másodlagos csoportok olyan halmazok, amelyek nem kizáró kapcsolatban vannak a tag objektumokkal. Egy területi adatrendszerből szerkesztett szelvény másodlagos objektumcsoport, mivel minden pontja tartozhat más szelvényekhez is. Az objektumcsoportok magasabb egységekbe szervezhetők. Pl. a mérési projektek tipikus csoport aggregációk, amelyek mérési kampányokból állnak. Az objektumcsoportok csomópontok az objektum hierarchiában és sokszor tisztán metaadatokból állnak. Kivételt képeznek a geofizikai fedvények, amelyek a szabvány terminológiájának megfelelően szintén adatszetteknek (geophysical object set) tekintendők. Ezek mögött a metaadatokon túl analóg, vagy digitális térképek, adatrácsok, vektor alakzatok állnak.

## **c) A riport származtatott metaadat osztály**

A geofizikai objektumok és objektumcsoportok bevezetésével komplex adatrendszerek leképezhetők, dokumentálhatók. Egy harmadik fogalom bevezetésével a földtani dokumentáció (jelentések, cikkek, kéziratok, terepi jegyzőkönyvek) szintén a metaadat rendszer részévé tehető, és azzal analóg módon kezelhető. Ez a fogalom a riport. A riport, mint származtatott metaadat osztály a hagyományos metaadatoktól két dologban tér el: földrajzilag meghatározott, és hozzákapcsolható a mérési adatokhoz. A riport osztályra épülő metaadatbázis egyik legizgalmasabb alkalmazása a nyílt, internetes publikációs portál, amely közvetlenül kapcsolódik a mérési adatokhoz és modellekhez.

### *2. tézis*

***Kidolgoztam a földrajzi adatokra vonatkozó ISO 19115 metaadat szabvány geofizikai kiterjesztését.***

## **d) A földrajzi adatokra vonatkozó metaadat szabvány geofizikai kiterjesztésének megalkotása**

A fent leírt fogalmak felhasználásával a GEOMIND projekt keretében elkészült a geofizikai metaadat szabvány adatmodellje, és az erre épülő XML nyelv, az XSD sémával definiált GEOMIND Profil. Az adatmodell kialakítása a vezetésem alatt álló nemzetközi munkacsoportban történt (Sőrés et al., 2007a). Az XSD séma kialakítása szoros együttműködés keretében a Dánia és Grönland Geológiai Szolgálatának (GEUS) feladata volt. A GEOMIND Profil lényege, hogy az ISO 19115 szabvány alapeleméből az a, b és c pontban leírtaknak megfelelően három származtatott osztályt hozunk létre. Ezek (GE\_GeophObject, GE\_GeophObjectSet, GE\_Report) öröklik a szabvány magjának összes elemét, és kiterjesztik azt a geofizika szempontjából fontos további elemekkel. Lehetőséget adnak a mérések körülményeinek, a műszerezettségnek leírására, és kibővítik az alapszabványban rendelkezésre álló kódlistákat. A geofizikai metainformációk a GE\_GeophysicalInfo nevű elemekben kaptak helyet, amely három fő részből áll. Ezek a mérések körülményeit részletező



GG\_MeasuringConditions, a műszerezettséget leíró GG\_Instrumentation, és a legfontosabb mérés technikai paramétereket tartalmazó GG\_Header.

## 2. Az Általános Geofizikai Adatmodell és leíró nyelv megalkotása.

A GEOMIND projekt másik fontos célkitűzése volt az ipari szabványok köréből kieső legkeresettebb geofizikai módszerek mérési adatainak cseréjét támogató szabvány kialakítása. Ez a feladat szintén az általam felügyelt munkacsomag része volt. A módszerként eltérő szabványok kialakítása helyett egy jóval általánosabb megoldás született. A projekt során elkészítettem az általános geofizikai adatmodell első verzióját. Végleges kidolgozására és tesztelésére a GEOMIND projekt zárása után, az ELGI adatbázis fejlesztési programjának keretében került sor.

### 3. tézis

*A geofizikai mérések és inverziós modellek egységes leírásához kidolgoztam az Általános Geofizikai Adatmodellt. Bevezettem a forrás, szenzor, terítési komponens, terítés, rétegmodell, rácsmodell és általános modell fogalmát. E fogalmakra építve kidolgoztam a geofizikai mérés, a geofizikai modell és az inverzió adatmodelljét.*

#### a) A geofizikai mérés adatmodellje

Az Általános Geofizikai Adatmodell (General Geophysical Data Model, GGDM) szakít a „mérési módszer” hagyományos koncepciójával, amely különböző struktúrákban definiálja a különböző méréseket. Eltekint a régi számítástechnikai korlátok miatt használt kényszerű egyszerűsítésektől, és a mérési elrendezések részletes dokumentációjára törekszik. A méréseket mozaikszerű kirakós játéknak tekinti, ahol az építkezés alapelemei a forrás és a szenzor. A mérési elrendezések egyszerűen forrás-szenzor konfigurációk, és ezek együttese adja a magát a mérést. Egy mérés tehát terítések (layout) összessége. A terítések terítési komponensekből (layout component) állnak. Típusuk lehet forrás (source) vagy szenzor (sensor). A komponensek geometriájának és a hozzájuk kapcsolódó technikai paramétereknek, mérési adatoknak megadásával bármilyen észlelés (nem csak geofizikai mérés) tökéletesen leírható. Az adatok tárolása a GML nyelvben használt érték tartomány (domain set) és érték készlet (range set) objektumok mintájára történik. Ez a gondolatmenet rokon a GeoSciML felfogásával, amely szintén GML alapokra építi a fel a szenzorokat, méréseket és megfigyeléseket leíró adatszerkezeteket. Ez biztosítja a majdani könnyű átjárást a GGDM és a GeoSciML között.

#### b) Geofizikai modell és inverzió adatmodellje

Geofizikai modellen a numerikus modellezéssel, vagy inverzióval kapott digitális modelleket értem. Ezek a mérésektől elkülönülő, önálló objektumok, bár gyakran a mérési adatokkal együtt kezelik őket. A mérés és modell közti egy-egy értelmű megfeleltetés nem kielégítő, hiszen egy mérésből több modell is származtatható (különböző feldolgozási verziók), és egy modell számos különböző mérés együttes inverziójának (joint inversion) eredménye is lehet. A mérések és modellek N:N egyed-reláció kapcsolatát realizáló köztes objektum maga az inverzió. Az inverzió önálló adatbázis objektumként való szerepeltetése lehetővé teszi a feldolgozási folyamatot jellemző technikai paraméterek, vektorok, mátrixot tárolását, utólagos vizsgálatát. A geofizikai modellnek három típusát különböztetjük meg. Rétegmodell (layer model), rácsmodell (grid model), és általános modell (general model). A rétegmodell olyan

1D modellnek felel meg, amely tetszőleges számú réteggel és rétegenként tetszőleges számú fizikai tulajdonsággal jellemezhető. A réteg attribútum szöveges érték is lehet, ezért fűrészi adatbázis kialakítására is használható. A rácsmodell a 2D, 3D, 4D inverziók leírására szolgál, tetszőleges számú cellával, és cellánként tetszőleges számú fizikai tulajdonsággal. Az általános geofizikai modell GML-ben leírható általános geometriájú komponensek összessége. A modell komponensek homogén testeknek tekintendők.

#### **4. tézis**

***Az Általános Geofizikai Adatmodell alapján elkészítettem az XML séma definíciókra épülő geofizikai leírónyelvet.***

#### **c) A GGDM geofizikai leírónyelv**

Az általános geofizikai adatmodell implementációját a GGDM XSD séma határozza meg. A séma olyan XML szöveg, amely előírja a használható elemek nevét, azok szerkezetét, sorrendjét. Az XML technológia használatának egyik nagy előnye az érvényesítés (validáció) lehetősége. Ennek során a validációt végző program az XML szöveget összeveti a sémával, és szóról szóra ellenőrzi az egymásnak való megfelelést. Az XML érvényesítés biztosítja, hogy a feldolgozásra, adateserére szánt szövegek megfelelnek a leíró nyelv által szabott formai követelményeknek. Az XSD séma az adatmodellt leíró adatszótárból (data dictionary) egy szövegfeldolgozó rutin segítségével automatikusan áll elő. A GGDM használatát bemutató példa állományok a <http://geomind.elgi.hu> címen találhatóak.

### **3. Az Általános Geofizikai Adatmodellre épülő adatbázis megvalósítása, és alkalmazása**

A fent vázolt adatmodellre építve az általános geofizikai adatbázis fizikai megvalósításának számos módja van. Dolgozatomban két lehetséges változatot mutatok be. Az egyik egy egyszerű és gyorsan megvalósítható, de kevésbé általános, a másik egy munkaigényesebb, de hatékony, és rugalmas megoldás.

#### **5. tézis**

***Az Általános Geofizikai Adatmodellre épülő adatbázis gyakorlati megvalósítására két kísérleti rendszert készítettem: az XML dokumentumtárat, amely egy hagyományos fájl archívum és az Apache Lucene keresőmotor összekapcsolására épül, valamint egy hibrid adatbázist, amely a hagyományos relációs adatbázist az XML alapú adatkezeléssel ötvözi.***

#### **a) XML dokumentumtár**

Az XML dokumentumtár érvényesített XML fájlok rendezett halmaza, amelyben egy speciális index állomány segítségével hatékony keresések hajthatók végre. A fájlok tárolása történhet egyszerű adatbázis táblában, vagy akár hagyományos könyvtárszerkezetben. Az adatok tárolásakor a rendszer megadott adatmezőkből egy index állományt készít. Az indexek a mezőértékeket a fájlok elérési útvonalához kapcsolják. A keresés eredménye lehet valamely mezőérték, vagy az elérési útvonal, amelynek ismeretében a felhasználó a keresett teljes rekordot kapja meg a rendszertől. Alkalmos XSL transzformációk segítségével az adatok tetszőleges formában megjeleníthetők. A keresések hatékonysága az indexállomány szerkezetétől függ. Az XML dokumentumtár megvalósítása az Apache Lucene keresőmotor

és az Apache SolR webalkalmazás összekapcsolásával viszonylag egyszerűen megoldható. Ez az implementáció elsősorban archiválási és adatszolgáltatási célra szánt rendszerekben javasolt. A KINGA portál, amely az informatikai kivitelezést végző InterComp Kft. által fejlesztett dokumentum szerver alkalmazására épül, hasonló elven működik.

## **b) Hibrid adatbázis**

A teljes GGDM adatmodell leképezése tisztán relációs adatbázisba több száz táblás rendszert eredményezne. A hibrid megoldás egyesíti a relációs adatbázis, és az XML adattárolás előnyeit. A gyakran használt kereső paraméterek hagyományos attribútumokként, a zárt egységnek tekinthető mély adatstruktúrák XML mezőkben vannak tárolva. Ez a megoldás viszonylag egyszerű szerkezet mellett nagyfokú hatékonyságot biztosít. Az önállóan használható elemek önmagukban is elérhetők, ami egyrészt megkönnyíti a nagy adattömeget tartalmazó mérések kezelését, másrészt rugalmasabb kereséseket tesz lehetővé. A modern adatbázis-kezelő rendszerek támogatják az XML szövegeken végzett, SQL-be ágyazott XPATH alapú kereséseket. Ezek a műveletek persze viszonylag lassúak, ezért mindig ajánlatos hagyományos mezők használatával csökkenteni a találati halmazok méretét. Az adatok tárolása előtt az XML fájlok ún. faktorizáción esnek át. Ennek során a külön táblában tárolandó elemek a szülőből törődnek, és helyükre egy egyedi azonosítóval ellátott hivatkozások kerülnek. Az adatok lekérdezésekor a faktorizációval ellentétes folyamat, az összeolvasztás zajlik. A hivatkozással megadott XML fragmentumok a megfelelő táblából visszakerülnek a szülő elembe. Mind a faktorizáció, mind pedig az összeolvasztás a többszörös beágyazások miatt rekurzív eljárás. Az adatbázis optimális szerkezetének kialakítása a teszt adatbázisok használata közben gyűjtött tapasztalatoknak megfelelően több lépcsőben történt. A hibrid adatbázis implementációja olyan rendszereknél javasolt, ahol a napi feldolgozás során az adatok intenzív használata és a rész struktúrákat érintő, kifinomult lekérdezések iránti igény a jellemző.

## **VI. Az eredmények hasznosítása**

A dolgozatomban leírt geofizikai metaadat modell fontos komponense az Európai Unió *eContentplus* programjának támogatásával megvalósult nemzetközi GEOMIND projekt által létrehozott internetes, nemzetközi információs rendszernek. A metaadatok cseréje a GEOMIND profilnak megfelelő XML állományokon keresztül zajlik. A GGDM adatmodell és leírónyelv a GVOP által támogatott KINGA projektben készült geofizikai metaadat portál működésében kapott jelentős szerepet. A metaadatok a nyilvános adatrendszerekhez való hozzáférés javításában alapvető fontosságúak. Az adatok interneten keresztül történő keresését a metaadatok felhasználása jelentősen megkönnyíti. A nyilvános publikációs rendszerekben elterjedten használt metaadat katalógusok lehetővé teszik, hogy az adatforrásokat a szerzők, az elkészítés ideje, vagy az adatokhoz kapcsolódó kulcsszavak, egyéb információk alapján könnyebben megtaláljuk, a milliányi adatrekord közül kiválaszthassuk azt, amire szükségünk van. A geofizikai adatrendszerek publikálásában a metaadatok szerepe még viszonylag kicsi. E tekintetben a nemzetközi összefogással készült többnyelvű GEOMIND metaadat portál és a hazai KINGA portál fontos előrelépést jelent. Bár az adatszolgáltatók és a portálok kezelő felületei még koránt sem használják ki a metaadat szabványban rejlő lehetőségeket, kétségtelenül érzékeltetik azokat. Külön hangsúlyozandó a nyilvános földtudományi dokumentumok (jelentések, esettanulmányok, oktatási anyagok) egységes térinformatikai rendszerbe foglalásának lehetősége, amely mindkét rendszer esetében adott. A KINGA rendszer pl.

kiindulási alapja lehetne egy intézmények közötti nemzeti tudásbázisnak, melyhez tetszőleges számú adatszolgáltató csatlakozhat. Ehhez csak a földtani dokumentációt ismertető szabványos metaadatokat kell elkészíteni és feltölteni a jelenleg már működő rendszerbe.

Az általános geofizikai adatmodellre épülő XML alapú adatsere és adattárolás lehetőségeit ma még nehéz megítélni. Az ad hoc fájlformátumok káoszában a GGDM hidat jelent az adatgyűjtő és adatfeldolgozó alkalmazások között. Lehetőséget teremt az egymástól elszigetelt adatrendszerek egységes kezelésére, az internetes technológiák szélesebb körű kihasználására. A dolgozatomban bemutatott hibrid adatbázis, amely a jövőben fokozatosan kivált néhány ELGI-ben kezelt heterogén felépítésű módszertani adatbázist, e technológia első alkalmazása. Az egységes elvekre épülő geofizikai web szolgáltatások kiépítésével lehetővé válik távoli adatszolgáltatók adatainak nagy központi adatbázis szervereken történő tárolása, illetve kis, elkülönülő adatbázisok egységes, osztott rendszerbe integrálása.

## **VII. Az értekezés témaköréből készült saját publikációk jegyzéke**

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](http://www.geomind.eu/portal/public_files/schema/geomind/geomind/2007)

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

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.

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

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](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](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](http://geomind.elgi.hu/doc/D6_2_B.pdf)

## **VIII. A téziszüzetben hivatkozott publikációk jegyzéke**

Federal Geographic Data Committee, 1998: Content Standard for Digital Geospatial Metadata

ISO/TC 211 2003: ISO/FDIS 19115 Geographic Information – Metadata

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/>

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

Sóres 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](http://geomind.elgi.hu/doc/D6_2.pdf)