

ÁTÁLLÁS AZ ÚJ INFORMÁCIÓTECHNOLÓGIÁRA A MAGYAR STATISZTIKAI HIVATALBAN

Mrs. Gizella Baracza
Hungarian Central Statistical Office

1.) A RENDSZERKONCEPCIÓ AKTUALITÁSA, A RENDSZER BEHATÁROLÁSA

A KSH új hardver/szoftver eszközeinek birtokbavétele után aktuálissá vált az adatok, a feldolgozási rendszer áttétele a régi környezetből az újra. Az átállással (migráció) kapcsolatban az a döntés született, hogy ez nem lehet a rendszer egyszerű, változatlan formában való áttétele az új környezetbe (erre technikailag sincs lehetőség, minthogy a régi eszközök nem állnak rendelkezésre az új számítógépes környezetben, sem a Központban sem a területen), hanem az átállásnak együtt kell járnia a teljes statisztikai információrendszer modernizálásával, az új információtechnológia bevezetésével. Ez azt jelenti, hogy az átállás a statisztikai információrendszer újratevezését igényli.

Az átállás során az első és legfontosabb feladat a **házon belüli felhasználók statisztikai munkájának** támogatása az új eszközök segítségével. Az erre a célra tervezendő rendszer kiterjed a statisztika számára szükséges összes funkcióra, a teljes adatfeldolgozási folyamatra.

2.) A KIINDULÓ HELYZET

A régi számítástechnikai környezetet az 1.sz. melléklet írja le.

A **kiinduló helyzet problémái** a következőkben foglalhatók össze:

A régi számítástechnikai környezet meghatározza a feldolgozási stílust, ami úgy jellemezhető, hogy

- a területi és központi feldolgozás nem alkotott egységes, integrált rendszert,
- a feldolgozások döntő többsége kötegelt (batch) módon zajlott, amiből következik, hogy a számítástechnika túlságosan elkülönült, nem épült

be a statisztikusok munkavégzésébe. (Ez alól kivétel a területi adatbevitel.)

- on-line lekérdező rendszer (SOLAR) készült a statisztikusok igényeinek kielégítésére, de minthogy saját fejlesztésű szoftverről van szó, nem lehetett elegendő fejlesztői kapacitást biztosítani, sem a továbbfejlesztésre, sem a karbantartásra.
- az adatbázis leírására létezett ún. metarendszer, de gondot okozott a batch karbantartás és az, hogy on-line lekérdezési lehetőséget nem tartalmazott - mert a MARK-IV. szoftver ezt nem tette lehetővé - csak kinyomtatott katalógusok formájában állt a felhasználók rendelkezésére.

A helyzetet tovább nehezíti az a tény, hogy magában a statisztikai rendszerben is nagy változások mennek végbe

- egyrészt az európai szabványok bevezetése (módszerek, nomenklatúrák),
- másrészt a társadalmi-gazdasági folyamatok változásának feszített ütemű statisztikai követése miatt.

A kiinduló helyzet erős oldalai, amelyekre az új rendszer kifejlesztésénél támaszkodni lehet, a következők:

- A gazdaság- és település statisztikában egységes adatbázisrendszer működik, amely
 - általános elvek szerint építi adatmodelljét,
 - kidolgozott az adatbázis elemek azonosítórendszerét,
 - az adatbázis tartalmát metaadatbázis írja le.
- Szabályozott munkamegosztás követi a feldolgozási folyamat alapvető funkcióit, ezek
 - adatelőkészítés,
 - adatbáziskezelés,
 - adatbázis-lekérdezés.
- A területi adatelőkészítésben olyan technológia alakult ki, hogy a statisztikusok on-line rendszerrel viszik be és javítják az adatokat.

- A statisztikusok körében PC-s szoftverek kezdtek elterjedni (LOTUS, SAS, SPSS).
- A SOLAR fejlesztése során sok tapasztalat gyűlt össze statisztikai adatle-kérdező rendszer tervezéséhez, megvalósításához.
- Mód volt arra - ha szűk körben is - hogy néhány rendszertervező megis-merkedjen a relációs adatbáziskezelés elveivel, gyakorlatával.

3.) ÚJ LEHETŐSÉGEK, CÉLKITŰZÉSEK

Az új számítástechnikai környezetet a 2.sz. melléklet írja le.

A számítástechnikai környezet leírásából láthatók az **új lehetőségek**:

- Egységes a hardver/szoftver környezet a Központban és a területi igaz-gatóságokon. Ebből következően egységes, integrált feldolgozó rend-szer alakítható ki.
- A szoftverrendszer egyik legfontosabb eleme az ORACLE adatbázis-kezelő rendszer, amely lehetővé teszi az adatok egységes, biztonságos kezelését, interaktív elérését, a korszerű adatvédelmi eljárások alkalmazását.
- Az ORACLE 4. generációs fejlesztő eszközei lehetővé teszik felhasználó-barát alkalmazások fejlesztését, amelyek segítségével a felhasználó szá-mítástechnikai ismeret nélkül, menü alapján választhatja ki a számára szükséges funkciókat, és a képernyőn megfogalmazott feladat eredmé-nyét azonnal ugyanott megkaphatja.
- Minimális számítástechnikai ismeret szükséges az ORACLE végfelhasználói eszközeinek használatához. Ennek segítségével a statisztikus ösz-szetett feladatokat fogalmazhat meg, és oldhat meg saját maga.
- A statisztikusok számára már ismert LOTUS, SAS, SPSS, valamint a WORD újabb Windows-os változata közvetlenül elérhetik az ORACLE adatbázis adatait. Az adatbázis adatai e szoftverekkel továbbfeldolgoz-hatók.

- Hálózati szoftver teszi lehetővé, hogy a területi igazgatóságok és a Központ között átadhatók legyenek adatok, illetve bármelyik hálózatra kapcsolt munkahelyről elérjük egy távoli adatbázis adatait, ha azt az adatvédelmi szabályok megengedik.

Az új lehetőségeket figyelembe véve a **célkitűzés** egy olyan rendszer kialakítása, amelyben

- a központi és területi alkalmazások egységes, integrált rendszert alkotnak,
- a rendszer egységes elvek szerint kialakított adatbázisra épül,
- a felhasználók (statisztikusok) saját munkahelyeikről a hálózaton keresztül elérhetik az adatbázist
 - olyan felhasználóbarát alkalmazásokon keresztül, amelyeket a rendszer tervezői az adott téma igényei szerint fejlesztettek ki,
 - általános végfelhasználói eszközökkel (ORACLE végfelhasználói eszközök, LOTUS, SAS, SPSS, WORD), amelyekkel bonyolultabb igények is kielégíthetők,
- az adatgazda statisztikus szerepe és lehetőségei jelentősen megváltoznak. Minthogy az adatbázis nem csak a feldolgozás végeredményeként keletkező adatokat tartalmazza, hanem a feldolgozási folyamat közbeni állapotait is, az adatgazda statisztikus folyamatosan "kézben tarthatja" saját adatait, s időben tájékozódhat az adatok minőségét illetően, javíthat, módosíthat az erre a célra kifejlesztett alkalmazások segítségével,
- azon igények kielégítése, amelyek több számítástechnikai ismeretet igényelnek, továbbra is a hagyományos módon az informatikusok feladata marad.

A tervezett új rendszerben megváltozik a statisztikusok és informatikusok szerepe. A statisztikusok egyre több feladatot önállóan képesek megoldani, az informatikusok szerepe pedig az, hogy megtervezzék, üzemeltessék az adatbázisrendszert, és olyan eszközöket fejlesszenek, amivel a statisztikusok önállóan dolgozhatnak. Ez az átalakulás azonban várhatóan igen hosszú folyamat eredményeként állhat elő.

4.) AZ ÚJ RENDSZER ÁTFOGÓ MODELLJE

4.1 A rendszer elemei

A rendszer átfogó modelljét a 3.sz. melléklet mutatja.

A rendszer **adatbázisa** négy fő részből áll, ezeket számmal azonosítja az ábra:

1. Metaadatok
2. Regiszterek
3. Adatgyűjtés-szervezési adatok
4. Statisztikai adatok

A rendszer **fő funkcióit** az ábra betűvel jelöli:

- A. A statisztikai információrendszer tervezése
- B. Metaadat-kezelés
- C. Regiszterek karbantartása
- D. Adatgyűjtés-szervezés
- E. Adatelőkészítés
- F. Feldolgozás
- G. Adatbázis-lekérdezés
- H. Tájékoztatás, elemzés

Az ábra nagyvonalakban mutatja az adatbázis fő részei és a fő funkciók közötti kapcsolatokat.

4.2 Az adatbázis fő részei

Az adatbázis fő részeit a következők szerint írhatjuk le röviden:

- 1.) **Metaadatok:** Ez a rész leírja a statisztikai rendszer elemeit. Ide tartoznak például az adatgyűjtések, kérdőívek, statisztikai fogalmak, statisztikai módszerek, mutatók, számítási eljárások, nomenklatúrák, ezek elemei és összefüggései, az adatbázis-elemek leírása, valamint a visszakeresést segítő tárgykörök, tárgyszavak rendszere. Ezekhez a rendszerelemekhez szöveges megnevezéseket, jellemző ismérveket, tartalmi leírásokat rendel, s tárolja az elemek közötti kapcsolatokat.

- 2.) **Regiszterek:** Ez a rész nyilvántartja a statisztikától függetlenül létező és változó sokaságok egységeit, amelyekre a statisztikai megfigyelés vonatkozik. Pl. gazdálkodó egységek, települések, országok.
Tárolja az aktuális állapoton kívül az egyes egységek változásait, a statisztikai megfigyelés időpontjára vonatkozó állapotot, esetleg a visszatekintő történeti állapotot.
- 3.) **Adatgyűjtés-szervezési adatok:** Ez a rész tartalmazza a regiszter egységek és az adatgyűjtések kapcsolatát. Kijelöli az egyes adatgyűjtések adatszolgáltatóit, jelzi az adatszolgáltatás teljesítését, vagy a nem-teljesítés okát, a pótlás iránti igényt, az expedálás információit. Adatokat tartalmaz az adatszolgáltatókkal való kapcsolattartáshoz.
- 4.) **Statisztikai adatok:** Ez a rész tárolja a statisztikai célra begyűjtött, a más intézménytől átvett, illetve az ezekből képzett adatokat. Ide tartoznak a feldolgozási folyamat során keletkező különböző állapotok, pl. minta, teljeskörűsített minta, a különböző aggregátumok, valamint az összehasonlíthatóság érdekében képzett homogenizált adatállapotok is.

4.3 A rendszer fő funkciói

A fő funkciók részletezését a 4.sz. melléklet foglalja össze a **rendszer hierarchikus funkciódiagramjában**. A fő funkciók feladatai röviden a következők:

A. A statisztikai rendszer tervezése

Az igények elemzése után a statisztikus megtervezi az adott adatgyűjtés tartalmát (mutatókör, vonatkozási kör, képzett adtok, módszerek, eljárások), majd megtervezi a kérdőívet és meghatározza a feldolgozással szembeni igényeit.

Ennek alapján az informatikus megtervezi az adatgyűjtés adatmodelljét, meghatározza, hogy milyen regiszterek szükségesek az adott témához, hogy történik az adatgyűjtés-szervezés és meghatározza a feldolgozási folyamatot. Dönt arról, hogy a téma adatai adatbázisba kerüljenek-e, milyen legyen az adatelőkészítés és a feldolgozás módja.

B. Metaadat-kezelés

A metainformáció rendszer jelenti a statisztikai információrendszer dokumentációját. Ezen metaadatok bevétele, folyamatos karbantartása a metaadat-kezelés. A metaadatok bizonyos részének aktualizálását statisztikusok végzik (pl: adatgyűjtések, fogalmak, módszerek, általános nomenklatúrák), más részét informatikusok aktualizálják (pl: az adatbázis tartalmának leírása).

A metaadatok használatára szinte minden funkciónak szüksége van. Ez az esetek egy részében tájékozódást jelent, vagyis a metaadatok az egyszerűen használható dokumentáció szerepét töltik be. Más esetekben a metaadatok központilag tárolt információk, melyeket a teljes feldolgozási folyamat egységesen használ (pl: egy nomenklatúrával leírt kódrendszert az adatelőkészítés a bevitel ellenőrzésére, az adatbázis-lekérdezés a kódhoz tartozó szöveg kinyomtatására használ.)

C. Regiszterek karbantartása

A regiszterek karbantartása valamely, a **statisztikától függetlenül létező és változó** egységek (pl. gazdálkodó szervezetek, települések, országok, címek) folyamatos aktualizálását jelenti. A nyilvántartás célja, hogy alapul szolgáljon adatgyűjtések tervezéséhez és azok megszervezéséhez.

D. Adatgyűjtés-szervezés

Az adatgyűjtés-szervezés minden esetben egy regiszterre épül. A regiszter egységeihez hozzárendeli a megfelelő adatgyűjtéseket és felügyeli a teljes folyamatot az adatkéréstől a beérkezéséig.

E. Adatelőkészítés

Az adatelőkészítés feladata a beérkezett adatok rögzítése, ellenőrzése, javítása.

F. Feldolgozás

A feldolgozás során készül el az adatbázis azon része, amit a végfelhasználók lekérdezhetnek. A feldolgozás közbenső állományai is hozzáférhetőek, de csak az adatgazda statisztikus számára.

A feldolgozáshoz tartozik a be nem érkező adatszolgáltatók pótlása, reprezentatív statisztikák esetén a becslés, a lekérdezést könnyítő aggregátumok előállítás, az összehasonlítható adatok képzése és az utólagos javítás.

G. Adatbázis-lekérdezés

Az adatbázis-lekérdezés feladata az adatbázisrendszer felhasználóinak kiszolgálása, annak biztosítása, hogy a felhasználó megkaphassa az őt érdeklő adatokat. Ez két módon történhet:

- a statisztikus önállóan használja az adatbázist,
- hagyományos úton a lekérdezést - megrendelés alapján - informatikus végzi.

A rendszer fokozatos kiépítésével a statisztikusok önálló adatbázishasználata egyre nagyobb arányú lesz, egyre kisebb lesz az igény a hagyományos, informatikusok által végzendő lekérdezések iránt.

H. Tájékoztatás, elemzés

A tájékoztatás a statisztikai munka végeredményét közvetíti a statisztika felhasználói számára. Ennek két fő formája van:

- előre megtervezett kiadványok formájában,
- az ad-hoc érdeklődők informálásával.

5.) A RENDSZER MEGVALÓSÍTÁSÁNAK SZOFTVERESZKÖZEI

Az adatbázis megvalósítása az **ORACLE relációs adatbázis-kezelő rendszerrel** történik. Az adatokat a rendszer kétdimenziós **adatbázistáblákban** tárolja. A táblák a relációs adatmodellnek megfelelően vannak kialakítva.

A funkciók megvalósításához a legfontosabb szoftvereszközök a következők:

- Általános adatbáziskezelő eszközök:
 - ORACLE7 RDBMS - relációs adatbázis-kezelő rendszer
 - SQL*Loader - adatbetöltést végző eszköz
 - EXP(ort)/IMP(ort) - segédprogramok az adatbázis mentésére és viszatöltésére

- SQL*DBA - az adatbázis adminisztrátorok munkáját segítő eszköz
- SQL*NET - az adatbázis hálózaton történő működtetéshez szükséges eszköz
- SQL*PLUS - általános adatkezelő eszköz
- 4. generációs alkalmazás fejlesztő eszközök:
 - SQL*FORMS - interaktív képernyős alkalmazások fejlesztésére szolgáló eszköz. Ennek segítségével készülnek az adatbázis karbantartását és lekérdezését szolgáló a felhasználóbarát alkalmazások.
 - SQL*MENU - menüvel vezérelt alkalmazások fejlesztésére szolgáló eszköz. Segítségével a különböző lekérdező, listázó, aktualizáló tevékenységek egy menüben összefoghatók.
 - SQL*REPORTWriter - táblázási feladatok eszköze

(Az SQL*FORMS és SQL*MENU a szoftver új verziójában ORACLE FORMS néven az SQL*ReportWriter ORACLE REPORTS néven szerepel.)

- Az adatbázis tervezését szolgáló eszközök:
 - CASE*DESIGNER - grafikus rendszertervező eszköz
 - CASE*DICTIONARY- adatszótár a rendszertervezés információi számára
 - CASE*GENERATOR- alkalmazások generálásához.
- Az adatok feldolgozását segítő további eszközök:
 - Pro*C - ORACLE adatok C programozási nyelvvel történő feldolgozásához
 - Pro*COBOL - ORACLE adatok COBOL programozási nyelvvel történő feldolgozásához
 - SAS - bonyolult statisztikai feldolgozásokhoz (pl: becslések, hibaszámítás, matematikai statisztikai eljárások)
 - SQL*TextRetrieval - szöveges adatok kezelését biztosító eszköz
- Az adatelőkészítés speciális eszközei:
 - BLAISE - a Holland Statisztikai Hivatal által kifejlesztett adatelőkészítő rendszer, amely lehetőséget nyújt

a kérdőív tervezésével egyidőben történő végleges ellenőrző program előállítására, így használatával lényegesen lerövidülhet az adatbeviteli program előállításához szükséges idő. (Bizonyos kérdőívtípusok leírására, kezelésére nem alkalmas pl: kötetlen számú sort tartalmazó tábla).

- PDE
 - saját fejlesztésű adatelőkészítő program a TPA gépen lévő SERIES IV. programok kiváltására, valamint a BLAISE-ben nem kezelhető kérdőívek feldolgozására (erre a kérdőívek átalakítása után a későbbiekben valószínűleg nem lesz szükség.)

- Az adatbázis lekérdezés speciális eszközei:
 - DataQuery
 - interaktív lekérdező eszköz
 - Data Browser
 - grafikus interaktív lekérdező eszköz
 - LOTUS
 - PC-s táblázatkezelő, amelyből közvetlenül elérhetők az ORACLE adatbázisban tárolt adatok.
 - SAS
 - statisztikai elemző programcsomag, amely közvetlenül használhatja az ORACLE adatbázist.
 - TPL
 - általános táblázási feladatokra, nagytömegű adatok esetén is.
 - SPSS
 - statisztikai elemző programcsomag, amely közvetve használhatja az ORACLE adatbázist
 - Word for Windows - szöveges anyagok készítéséhez, kiadványok összeállításához, amelybe közvetlenül behívhatók az ORACLE adatbázis adatai.

Számítástechnikai környezet az átállás előtt

Központ:

Hardver

IBM 4381-14

memória: 32Mb

lemez kapacitás: 18Gb

mágnesszalag egység: 16db

terminál: 60db

sornyomtató: 3db

MDS 6402/6403 adatrögzítő berendezés

PC: 50db

Szoftver:

Operációs rendszer: MVS

Adattárolás: MARK IV, RAPID

Táblázás: MARK IV, TPL, TAB68

Programozási nyelv: COBOL, PL/1, FORTRAN, RPG

Adatbázis-lekérdezés: SOLAR (saját fejlesztés)

Elemzés: SAS, SPSS

Szövegszerkesztés PC-n: MS-WORD (DOS alatt)

Táblázatkezelés PC-n: LOTUS (DOS alatt)

Megyei igazgatóságok:

Hardver:

TPA

memória:768Kb

lemez kapacitás: 240-300Mb

terminál: 8-16db

Szoftver:

Operációs rendszer: RSX11M

Adatbevitel: SERIES IV

Programozási nyelv: FORTRAN, C, PASCAL, DIBOL

Adattárolás , lekérdezés:DATATRIEVE.

Az új számítástechnikai környezet legfontosabb elemei

Hardver:

Központ:

Szerverek: HP9000-890, -867,-847,-817(3db)
memória: 32Mb-512Mb
lemezkapacitás:1.3Gb-48Gb
optikai lemeztár:57Gb
DAT kazettás egységek:8*8Gb
hálózati kapcsolat az IBM 4381-14-hez (lásd
1.sz.melléklet)

Egyéb berendezések:

8db UNIX munkaállomás HP9000/715
270db PC (HP Vectra és korábbi beszerzések)
26db HPLaserJet nyomtató
3db sornyomtató

Megyei igazgatóságok:

Szerverek: HP9000-867, -817(18db)
memória:32Mb-80Mb(megyenagyságtól függően)
lemezkapacitás:2.3Gb-4Gb (- ||-)

Egyéb berendezések:

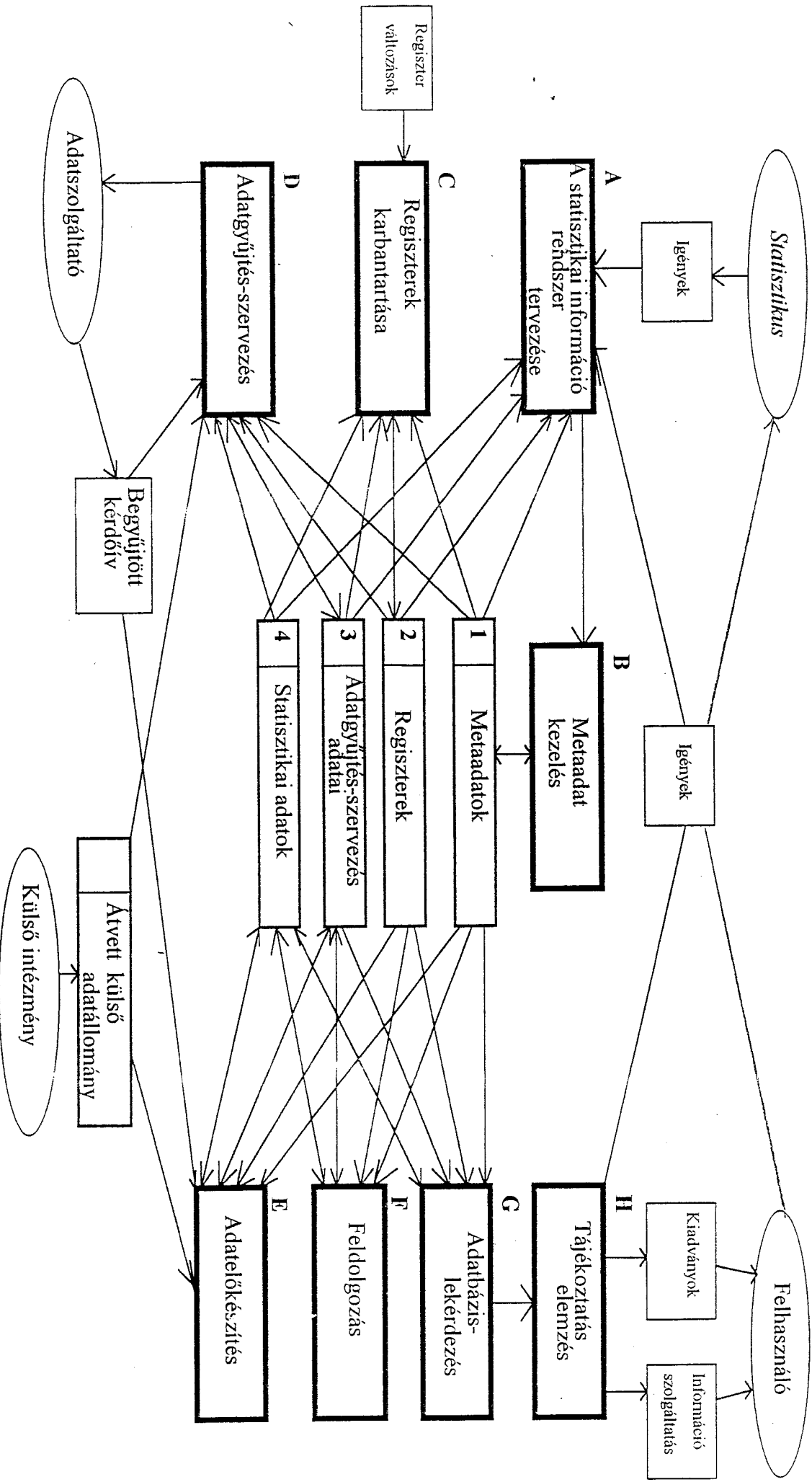
19db UNIX munkaállomás HP9000/715
400db PC
19db HPLaserJet nyomtató
19db sornyomtató

Szoftver:

Központ és megyei igazgatóságok:

Operációs rendszer: HP-UX9.0
Adatbáziskezelés: ORACLE
Alkalmazásfejlesztés: PDE, BLAISE, ORACLE, SAS
Elemzés: SAS, SPSS, TPL
Programozási nyelv: COBOL, C,FORTRAN, PASCAL
Szövegszerkesztés PC-n: Word for Windows
Táblázatkezelés PC-n: Lotus 1-2-3 for Windows
Hálózatkezelés: LAN Manager, HP Arpa
Elektronikus levelezés: OpenMail

Átfogó rendszermodel I
 adatfolyam diagram



Átfogó rendszermodell I
Hierarchikus funkciódiagram

