

Tömeges adatok rendszerezett tárolása és a gyors hozzáférés biztosítása „ADATBANK”-ok útján

Lengyel Mihály alezredes

Az adatbank kifejezés tartalma meglehetősen homályos fogalom, amely nem más, mint nagyszámú felhasználók rendelkezésére álló központi információs raktár.

Az adatbankokra vonatkozó tervek távolinak feltételezett megvalósítása és a még kiforratlan szóhasználat miatt ezen a területen még bizonyos fokú tisztázatlanság figyelhető meg. A frissebb keletű kísérleteket szem előtt tartva próbálom tisztázni az adatbankok fogalmát, azok fajtaíát és felhasználásuk módját.

I.

Az adatbankok fogalma

1. Származás

Az adatbank fogalma az Egyesült Államokban született meg, a magánszektor és a közigazgatás „manager”-ei igényeinek hatására. A managerek, hogy ellenőrizhessék az ügyvitelt, előkészíthessék a döntéseket és előrejelzéseket készíthessenek — valahányszor erre lehetőség nyílt —, az esetenkénti felmérések helyett fel akarták használni, méghozzá rendszeresen azokat az információkat, amelyeket az ügyviteli eljárások automatizálása tett hozzáférhetővé.

Ekkor ütköztek bele az első akadályba, megállapítva, hogy a látszólag azonos, de speciális adatfeldolgozási célokból összegyűjtött és feldolgozott információk az esetek többségében összeegyeztethetetlenek, sőt összefüggéstelenek.

Valóban, a más-más célokra alkalmazott meghatározások és kódolási rendszerek annyira különbözőek, hogy lehetetlenné teszik az adatok összevetését, bár ez nélkülözhetetlen ahhoz, hogy szintetizált eredmények jöjjenek létre. (Például valamely intézménynél a személyügyi nyilvántartásokban felhasznált meghatározások nem egyeznek meg a bérszámfejtéshez szükséges nyilvántartásokban szereplő meghatározásokkal.) Bármiféle vizsgálat elvégzése — a költségek alakulásáról, a munkaerőhelyzetről vagy

a személyi állomány stabilitásáról — az adatok összehasonlítását teszi szükségessé, de a nyilvántartások eltérő szempontjai miatt ezek — általában — lehetetlenné válnak, vagy ha azok mégis elvégezhetők, akkor az teljesen véletlenszerű.

A helyzet még kiélezettebb a közigazgatás területén, ahol az információk összegyűjtésénél és feldolgozásánál a párhuzamosság az uralkodó. Bár különféle formában, de ugyanazokat az információkat gyűjtik össze — a különböző közigazgatási szervektől származó — megszámlálhatatlan kérdőíveken, az apró meghatározásbeli különbségek azonban lehetlenné teszik az elszigeteltség leküzdését és hozzájárulnak ahhoz, hogy a dolgoknak ez az irracionális és rendkívül költséges állapota fennmaradjon. Ennek az állapotnak az eredete pedig pszichológiai okokkal magyarázható tartózkodásra és a történelem hagyatékának továbbélésére vezethető vissza. Ennek megfelelően az információk minőségének és hozzáférhetőségének megjavítása nem képzelhető el másként, csak a közigazgatási eredetű információk kezelésének fokozatos racionalizálásával.

A közigazgatásban működő intézmények management-tevékenységéhez szükséges információ igény állandó növekedése azt eredményezte, hogy lassanként olyan rendszerek koncepciója alakult ki, amelyek összefüggő, egységesített és megközelíthető információkat tartalmaznak. Az ilyen rendszerek megvalósítása elengedhetlenné teszi az információáramlás megszervezésének és az információgyűjtési folyamat állomásainak mélyreható vizsgálatát, szem előtt tartva a gazdaságossági és összefüggőségi követelményeket.

Ezzel a megkötéssel közös, általános alapelveken — az adatbank közös alapelvén — nyugvó rendszer valósítható meg.

2. Általános alapelvek

Az adatbank alapfeladata az, hogy a különböző forrásokból származó információkat összesítse, és azokat időben bocsássa a felhasználók rendelkezésére. Ennek megfelelően az adatbank három fő részből tevődik össze:

- a) az adatok szervezett összesítése: az adatbázis;
- b) az adatbázist információkkal ellátó források és az adatbázisok közötti információcserét biztosító hálózat;
- c) az adatbázist lekérdező és az eredményeket megkapó felhasználók és az adatbázis közötti információcserét biztosító hálózat.

3. Adatbázis

Az információs rendszer keretei közötti adatbankban elhelyezett információk halmazát úgy kell rendezni, hogy az a felhasználók kérdéseit megválaszolhassa. Ez a rendezés azt jelenti, felhasználási nyilvántartásokat (azaz file-okat) kell létrehozni és azokat állandóan naprakészen tartani. A válaszadást szolgáló file-ok a forrásokból származó adatokból kerülnek kialakításra és azok az alábbi felépítést teszik szükségessé:

- ki kell alakítani keresztezett hivatkozású file-ok halmazát és ezt el kell látni egy, az összes információt tartalmazó állandó katalógussal;
- ki kell alakítani egy fejlett file-kezelési rendszert.

Egy olyan rendszer létrehozásánál, amelyben az információk az adatbázisban rendezettek, az alábbi szempontokat kell szem előtt tartani:

— a redundáns adatok kiküszöbölése (ugyanis a file-ok keresztezett hivatkozási rendszere lehetővé teszi, hogy elkerüljük a file-ok tartalmának ismétlődését egyik file-tól a másikig);

— a tárolási költségek csökkentése (az előző célkitűzés következménye);

— a naprakészség megkönnyítése, mivel a különböző felhasználási file-okat egyidejűleg tesszük naprakésszé ugyanazon információ (információk-események) alapján, ezzel a file-okban szereplő információk módosítása vagy megszüntetése nagymértékben leegyszerűsödik;

— egy-egy konkrét felhasználási területet érintő adatokat nem kell minden program esetében újra meghatározni, mivel ezt az adatbázis részét képező információkatalógus feleslegessé teszi.

Hogy a kép szemléletes és egyszerűbb legyen, az mondható, hogy a művelet hagyományos képét (az adatok felvonulása egy programban) felcseréljük egy új köppel, ami nem más mint egy program felvonulása az adatbázisban.

II.

Az adatbankok osztályozása

Az információforrások vagy a felhasználók jellegét és az adatbázis jellemzőit figyelembe véve, többféle osztályozás alakítható ki.

1. Az információforrások

Az információk forrásai lehetnek:

— elsődleges információszolgáltatási egységek (természetes személyek, intézmények stb.) Ebben az esetben az adatbank közvetlenül az alapdokumentumokat dolgozza fel;

— olyan szervezetek, amelyek már összegyűjtötték és megváltoztatták az elsődleges információkat, file-okban rendezve őket. A file-okat az adatbank forrás-file-okként használja fel. Ezek a file-ok lehetnek részletezők, tehát a bennük szereplő információk elsődlegesek vagy pedig a forrás-szervezet által kidolgozott statisztikákat tartalmazó összevont file-ok.

2. Az adatbank felhasználói

A felhasználók három csoportra oszthatók:

— döntések előkészítésével és előrejelzések kidolgozásával megbízott gazdasági és statisztikai vizsgálatokat végző felelő személyek;

— maguk a döntést hozók;

— olyan különféle intézmények, amelyek fel akarják használni az adatbankban tárolt információkat vezetési-irányítási feladataikban és eljárásaikban.

Az esetek nagy többségében a felhasználók a rendszer információinak naprakésszé tételét alapvető jellegűnek tekintik.

A naprakészségről kétféleképpen lehet gondoskodni:

- a felhasználók költségére speciális felmérések elvégzése útján vagy
- szisztematikusan felhasználva azokat az információkat, amelyeket mások saját céljaikra gyűjtöttek, és azok segítségével rendszeresen naprakészé tenni a tárolt információkat.

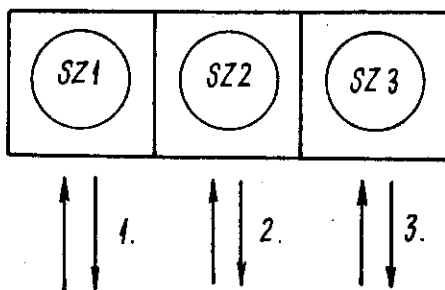
A második eljárás az, amellyel a leggyakrabban találkozhatunk és amely az adatbank tulajdonképpeni fogalmának legjobban megfelel. Ugyanakkor elképzelhető az eljárás keretein belül az, hogy kiegészítő jellegű kisebb felmérést is elvégezzünk, különösen a nagy fontosságú témakörökkel kapcsolatban.

3. Felhasználó és az adatszolgáltató közötti kapcsolatok

Ebből a szempontból négyféle típust különböztethetünk meg:

a) A feldolgozási adatbank.

Az egyes felhasználók önmaguk töltik fel és keresik vissza az információkat saját file-aikban anélkül, hogy megosztanák az információkat más felhasználókkal. Amennyiben a file-ok nem lehetnek közösek, tulajdonképpen a szó szoros értelmében nem is beszélhetünk adatbankról. Jó példák erre a különféle, egymástól független felhasználók rendelkezésére álló, time sharing üzemmódban dolgozó számítóközpontok.



b) A statisztikai adatbank

Szűkebb értelemben ez már adatbanknak mondható, amelyben az adatbázist egy sor adatszolgáltató szolgáltatja néhány felhasználó részére.

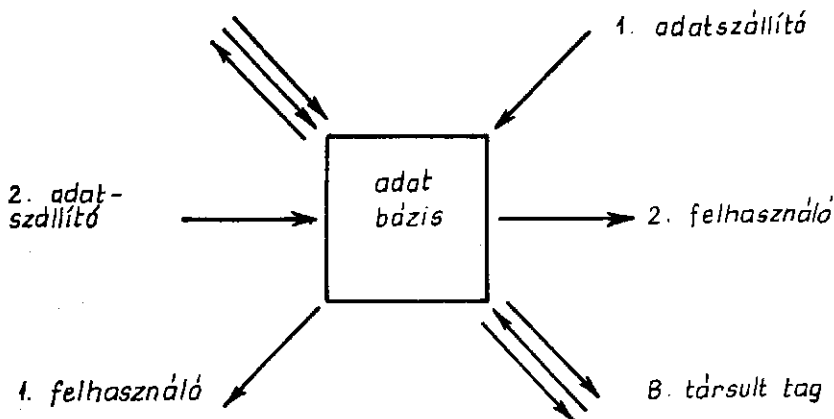
Elképzelésem szerint ilyen jellegűnek kell lennie a katonai tárgyú tájékoztatási adatbankoknak is.

c) Az információs szövetkezet

Ennél a típusnál az adatbázist egymással társult felhasználók-adatszállítók egyidejűleg hozzák naprakész állapotba és kérdezik le (pl. több légitársasági vállalat közös helyfoglalási rendszere).

d) Az információcsere-rendszer

Ez az a rendszer, amely tulajdonképpen a társult tagokat, adatszállítókat és felhasználókat az egymás mellé helyezett vagy az adatbázis kezein belül összekapcsolt wile-ok összes lehetséges kombinációjával köti össze.



4. Az adatbázis jellemzői

Az előbbieken alapján megállapítható, hogy az adatbázis képezi az összekötő kapcsokat az információk forrásai és a felhasználók között. Ennek megfelelően szorosan kapcsolódik az információnyerési és visszakeresési (retrieval) eljárásokhoz. Az adatbázisok — általában — összetett és egyszerű adatokat tartalmazhatnak, ennek megfelelően a visszakeresésnél nincs szükség tárgyszavak felhasználására vagy thesaurus kialakítására, tehát az automatizált dokumentációs rendszerek technikai fegyvertárára.

Ami magát az adatbázist illeti:

— a felhasználási file-ok szerkezete megegyezhet vagy eltérhet a forrás-file-ok szerkezetétől;

— megegyezés esetén az információgyűjtés megszervezése a visszakeresés szempontjai szerint történik. Olyan rendszerről van itt szó, amelyben az adatbank foglalja el a központi helyet, ez pedig ritkán fordul elő. Az ellenkező eset a leggyakoribb;

— ha a hírközlési eszközök ezt lehetővé teszik, nem szükséges az összes felhasználási file-t egy helyen koncentrálni, elegendő, ha a file-ok az információkatalógusban (index) össze vannak kapcsolva;

— az adatbázis magját képező index egyetlen helyen kell, hogy elhe-

lyezkedjen és irányításáról egyetlen adatfeldolgozó berendezésnek kell gondoskodnia, amely összegyűjti a kérdéseket és azokkal irányítja az adatbázis karbantartását.

Az adatbank tartalmának vizsgálata után megpróbáltam csoportokba sorolni az adatbankokat. Ezek után célszerű megvizsgálni azokat a problémákat, amelyeket az ilyen rendszerek konkrét felállítása vet fel.

III.

Az adatbankok létesítése

1. A személyi állomány

Ha összehasonlítunk egy bonyolult adatfeldolgozó rendszert és egy adatbank megvalósításának körülményeit, a legnagyobb különbséget az információk elemzésének tulajdonított pontosságban találjuk meg. Ezt a különbséget megtalálhatjuk az elemzés különböző szakaszaiban, ami kifejezésre jut a személyi állománnyal szemben támasztott követelményekben is.

A személyi állomány:

— az adatok szakértője, az információk szerkezetének és a file-szervezésnek a szakemberei;

— az adatok elemzői, akik összegyűjtik és elemzik az információk meghatározását és kódolási rendszerének összeállítását;

— a különleges képzésben részesült programozók, akiknek szakterülete az adatbázis adatainak karbantartása és naprakész állapotának biztosítása.

Az adatok szakértői valószínűleg a legfontosabb emberek a műveletek végrehajtásánál. Az információk összegyűjtésének vízszintes síkon történő megszervezése és ennek alapos elemzése azt eredményezi, hogy a döntéseket igen nagy fokú hierarchiában kell meghozni. Egyébként az igények meghatározása nagyszámú és igen közeli kapcsolatot tesz szükségessé a potenciális felhasználókkal. A megfelelő technikai hozzáértésen kívül az előzetes tanulmányokkal megbízott kutatócsoportnak olyan jó felkészültségű vezetőre van szüksége, aki alkalmas egy nagyszabású szervezet megfelelő irányítására, a döntések meghozatalára és a különféle tárgyalópartnerekkel a kapcsolatok kialakítására.

Az adatbankoknak és a hozzá kapcsolódó rendszereknek a megtervezése olyan hozzáértést igényel, amely biztosítja:

— a már létező rendszer elemzését;

— a kielégítendő igények meghatározását.

A fenti tényezők alapján kell meghatározni:

— az igénybe veendő információforrásokat;

— a kimenő információkat;

— a rendszer belső technikai jellemzőit és

— az ennek megfelelő feltételfüzetet.

A rendszer problémáinak megközelítése akkor hatékony, ha a vizsgálathoz hipotézisek kerülnek felállításra ahhoz, hogy milyen információkat nyújthat a rendszer, és ezután ezeknek a hipotéziseknek az alapján árnyaltabban kerülnek meghatározásra a szükségletek. Ezzel egyidejűleg mind technikai, mind szerkezeti-felépítési szempontból kibontakoztatható egy, a realitásokkal számoló vázlat, amely ezután végül elvezet az átfogó koncepció létrehozásához.

A részletes tanulmányok szempontjából a tanulmányok lefolytatásának során külön ki kell emelni a tervezet előrehaladásának, pontos ellenőrzésének jelentőségét, továbbá az előrejelzések mind szélesebb körű felhasználásának és az előrejelzések „visszacsatolással” történő naprakésszé tételének szükségességét.

Technikai szempontból a feldolgozást a következők jellemzik:

— rendkívül terjedelmes file-ok feldolgozására van szükség (a terjedelem több milliárd karakteres nagyságrendet is elérhet);

— jól meghatározott kritériumokat kielégítő információkat kell kiválasztani;

— gyakran kívánatos, hogy ezeket a file-okat „real-time” vagy párbeszédés üzemmódban lehessen lekérdezni;

— olyan egyszerű információvisszakeresési nyelvet kell a felhasználók rendelkezésére bocsátani, amely lehetővé teszi, hogy nehézségek nélkül hozzájuthassanak az adatbankban tárolt információkhoz. Az eredmény eléréséhez az szükséges, hogy a rendszer technikai kialakításának szintjén eldöntsük azt, hogy saját magunk állítsunk-e össze megfelelő programot, vagy pedig egy már kész software-t szerezzünk-e be („build or buy analysis”). Az esetek többségében gazdaságosabb egy már kész software beszerzése. Ennek okai az alábbiak:

— az adatbankot „nyílt” rendszerként kell felépíteni; más-más forrásokból származnak az információk és más-más felhasználók lépnek fel információ iránti igényekkel. Egy speciális programrendszer nehezen egyeztethető össze ezzel a „nyíltsággal”;

— az adatbázisba szervezett információk feldolgozására számos software áll rendelkezésre. Általában tartalmaznak egy ügyviteli és egy lekérdezési nyelvet is;

— az a befektetés, amelyet ilyen bonyolult rendszerek egyedi elemzése és programozása tesz szükségessé, rendkívül jelentős. A csőd szélére vezethetne az, ha minden egyes különleges alkalmazásnál új software-t akarnánk „gyártani”.

Ennek az elemzésnek az eredménye tehát az, hogy meg kell keresni az igényeinknek megfelelő software-t, tehát azt, amely képes biztosítani az adatbank feladatainak elvégzését.

Ez a keresés bonyolultabb, mint amilyennek első látásra tűnik, mivel a gyártó cégek és a szolgáltató vállalatok sok általános file-kezelési software-t dobtak piacra. (Pl. Data Management, Data Base Management Systems vagy a General Purpose File Management).

Az általános file-kezelési software kiválasztásának szempontjai első-

sorban arra épülnek tehát, hogy milyen mértékben képes az adott software megoldani az adatbank működtetésével kapcsolatos feladatokat. Ha egy adott software ebből a szempontból kielégítő, akkor meg kell vizsgálni:

- a teljesítményeket (például meg kell állapítani, hogy a kiválasztott software értelmező avagy fordító-e);
- a vételi vagy bérleti költségeket;
- a tényleges beszerezhetőséget.

2. A titkosság problémája

Ez a program talán egyike a legvitatottabbaknak, pedig a megoldás technikai előfeltételei már rendelkezésre állnak. Hogy tisztább legyen a kérdés, két szintet kell megkülönböztetni:

- a konzultálás titkát (ezt hozzák fel leggyakrabban, mert egyszerre érinti az információk összegyűjtését és a lekérdezőket);
- az adatbázis védelme, naprakésszé tétele és módosítása.

Főként az első, az információk bizalmas jellegével és néha a statisztikai titkokkal is kapcsolatos probléma áll előtérben. A leggyakrabban alkalmazott megoldások megfelelő kulcsok és átmeneti szavak használatán alapulnak. A software számolja az egymás utáni próbálkozásokat és elutasítja a kérelmet, ha a próbálkozások mennyisége meghaladja az előre rögzített nagyságot, így elkerülhetők a tapogatózó kísérletek. Ismeretesek más megoldások is, mint például a konzol segítségével megvalósított védelem.

IV.

Statisztikai adatbank

A fentiekben az adatbankot mint nagy adatmennyiségek gyors és rugalmas kiértékelési lehetőségét definiáltuk. Elvileg minden lexikon és minden kartoték ilyen adatbankot jelent. Azonban mindkettő a gyors és rugalmas kiértékelés körülményét csak nagyon korlátozottan teljesíti. Itt ezért adatbank alatt csak egy olyan adattárolást értünk, amely elektronikus adatfeldolgozó berendezéssel közvetlenül elérhető és lehetőleg kevés manuális munkát kíván.

1. Felhasználási cél

Csak a fentiek szerint pontosan definiált adatbankoknál vannak különböző formák, amelyek főként a felhasználási célból különböztethetők meg. Lényeges megkülönböztető ismérvük az alábbi:

- az adatkészletezés fajtája,
- az adatkészlet kezelésének módja,
- az információ kiértékelésének módja.

Az adatkészlet fajtájára vonatkozóan két formát különböztetünk meg:

- eredeti adatok tárolása,
- levezetett (származtatott) adatok tárolása.

Az eredeti adatok tárolásánál a bemenő információkat minden további feldolgozás nélkül teljesen megőrizzük. Az adattárolásnak ez a módja általában a kartotékoknál és regisztereknél fordul elő. Ritkábban találkozunk levezetett adatok tárolásával. Ez alatt azt kell érteni, hogy a bemenő eredeti adatokból csak az új felismeréseket vezetjük le és csak ezeket vesszük fel az adatbankba. Itt többnyire az eredeti anyag sűrítéséről van szó. Magától értetődően minden anyagsűrítés bizonyos információ-vesztést jelent.

Készletkezelés alatt értendő az adatok bevitele vagy kivitele az adatbankból. Mind a bevitelnél, mind a kivitelnél — legalábbis elméletileg — az alábbi lehetőségek fordulnak elő:

- nincs változás,
- szakaszos (periodikus vagy esetenkénti) változás,
- folyamatos változás.

Az alábbi táblázat mutatja a növekedésekből és csökkenésekből adódó lehetőségeket és példákat ad erre.

Növekedés Csökkenés	nincs	szakaszos	folyamatos
nincs	lezárt archívum (bírósi ítéletek 1970)	nyílt archívum (a havi statisztikai felvételek ered- ményei)	nyílt regiszter (születési regisz- ter)
szakaszos	megszűnő regisz- ter	nyílt archívum régi események lerakásával	nyílt regiszter régi események lerakásával (egy könyvtár adattá- ra)
folyamatos	figyelmeztető adattár egy meg- határozott beszá- moló körzet egy- szeri statisztikai felmérésére	figyelmeztető adattár egy meg- határozott beszá- moló körzet peri- odikus statisztikai felmérésére	folyamatos adat- tár (kutatói adattár, jelentő regiszter)

A fenti séma adattáraiba történő besorolás bizonyos mértékig szubjektív, mert a „folyamatos változás” fogalma az „esetleges”-től nehezen határolható el. Elvileg azonban a „folyamatos” alatt a mindenkori eset azonnali teljesítését kell érteni.

Az adatbank-formák további szisztematikus felosztása az információértékelés fajtájára vonatkozik. Ennél az alábbiakat különböztetjük meg:

- az adattárhoz történő hozzáférés időbeli meghatározása és
- az információkivétel formája.

Erről a következő séma ad áttekintést. Az „azonnali hozzáférés” alatt

az adattár közvetlen igénybevételét értjük számítógépen keresztül, míg az „esetenként” azt jelenti, hogy a közbelső fokozatok iktathatók be és a kérdések naponkénti feldolgozása esetleg elegendő. A „tisztá kereső folyamat” alatt az adatkivitelnél az egyes esetek megtalálását értjük, az adattárba többé vagy kevésbé komplikált szabályok szerint.

Hozzáférés	Esetenként	Azonnal
A kivétel formája		
Tiszta kereső folyamat	figyelemzettő adattár	kutatási (nyomozási) adattár
Komplex és váltakozó kiértékelés	speciális statisztikai kiértékelés	szorosabb értelemben vett management információs rendszer

2. Technikai formák

Ma különböző lehetőségek vannak, hogy nagy adatmennyiségeket úgy tároljunk, hogy az elektronikus feldolgozás részére gyorsan elérhetőek legyenek. Elsősorban az alábbi tárolási eszközök jöhetnek szóba:

mágnesszalagok,

mágnesfilmes tárolók,

mágneslemezes tárolók,

mágnesdobos tárolók.

Mágnesszalagok jelentik a legolcsóbb tárolási lehetőséget az említettek közül. Azonban megvan az a hátrányuk, hogy nem biztosítanak közvetlen hozzáférést egy meghatározott információhoz. Ha a kívánt adatok nem állnak a tekerecs kezdetén, ezeket először — programvezérléssel — le kell csévelni.

A mágnesszalagokkal elvileg csak szekvenciálisan lefutó folyamatokat lehet racionálisan megoldani. A többi tárolóeszköznél az úgynevezett elérési idő minden tetszés szerinti információhoz sokkal rövidebb (a másodpercek tartománya). Ezek gyakorlatilag mindenütt szükségesek, ahol az azonnali feldolgozás az előző fejezet értelmében szükséges. Az említett három közvetlen hozzáférésű tárolóeszköz között lényeges különbségek vannak, a hozzáférési idő és a tárolókapacitás vonatkozásában. A szalagtárolók (mágnesszalag) rendelkeznek a leglassúbb hozzáférési idővel, azonban viszonylag olcsók és nagy kapacitásúak (mintegy egymilliárd jel közvetlen kapcsolatban az elektronikus adatfeldolgozó berendezéssel). Fordítva van ez a viszony a dobtárolóknál. A lemezek minden vonatkozásban a középen helyezkednek el.

Adatok bevitelére és kibocsátására az adatbankban több lehetőség van. A legfontosabbak az alábbiak: lyukkártya, lyukszalag, mágnesszalag, írógép, gyorsnyomtató, képernyős berendezés. Az azonnali adatfeldolgozásra gyakorlatilag csak az írógép, nyomtató és képernyő alkalmas. Természetesen az azonnali adatfeldolgozás valóban csak kevés esetben szük-

séges. Magánál a kutatásnál ma még nem dolgoznak ebben az „azonnali” értelemben.

Az adatbank használatánál meghatározott alkalmazások esetében különleges szerepet játszik az adatátvitel. Segítségével több, területileg távolfekvő adatszolgáltató léphet kapcsolatba egy elektronikus adatfeldolgozó berendezéssel és ezáltal egy adatbankkal, amikor is mind az új adatok bevétele, mind az információnyerés az adatbankon keresztül történhet. Elméletileg manapság több mint 100 hely kapcsolódhat egy megfelelően kiépített adatbankhoz. A gyakorlatban az ilyen rendszereknél mégis felmerül néhány nehézség, amelyet legnagyobb részt az adatátvitel okoz. Egy ilyen rendszer költségei ugyancsak jelentősek.

3. *Egy statisztikai adatbank felépítésének általános irányelvei*

Az adatbank létesítésére vonatkozó elgondolásoknál az alábbi szempontoknak van különös jelentőségük:

Egy elektronikus adatfeldolgozó berendezés konfigurációjának — amely adatbankhoz kapcsolódik — hatékonyságát döntően a felhasználási cél határozza meg. Minden adatbank létesítésénél ezért különös gonddal végzett tervezésre van szükség az összes érdekelt, különösen a felhasználók együttműködésével.

Jelenleg sok helyen — gyakran azonos okból — tervezik adatbank felállítását. Feltétlenül szükséges, hogy ezek az elgondolások koordináltan kerüljenek végrehajtásra. Kétségtelen, hogy belátható időn belül felmerül a különböző adatbankok között az adatscere problémája. Ha azonban közben különböző módszereket alkalmaznak, ez az adatscere többnyire nehézkes lesz, ha egyáltalán nem válik lehetetlenné.

Ha az adatbank-rendszereket erősen a felhasználási szempontok szerint kell kialakítani, akkor az adatbankot lehetőleg kizárólag szakmai és nem az intézmények szempontjai szerint kell létrehozni. Itt általában lehetséges, hogy a jelenlegi munkamegosztást — a különböző ágak között — valamelyest módosítsuk. Azonban semmi esetre sem lenne célszerű ilyen okok miatt nagyon különböző felhasználási célokra, illetve feladatokra központi adatbankot létrehozni.

Egy adatbank kapacitása — a gazdaságosság szempontjait figyelembe véve — ma még nem elégíti ki minden kívánságot. Ezért nagyon kritikusán meg kell vizsgálni, hogy milyen adatokat vegyünk fel az adatbankba. Ezen ok alapján is egyetlen általános adatbank-központ egy nagyobb regionális vagy szakmai terület részére kevésbé ésszerű.

Mielőtt az azonnali feldolgozásról távolsági adatátvitellel terveket készítenénk, alaposan meg kell vizsgálni, hogy ez valóban szükséges-e. A számítógépes adatbank költségei közvetlen feldolgozás nélkül lényegesen kisebbek, mint a közvetlen feldolgozás beruházásai.

4. *A katonai statisztika adatbankja*

Már a bevezetésben utaltam egy adatbank jelentőségére a hivatalos statisztikában. Ennek az az eredménye, hogy mind az adatszolgáltatás, mind a kiértékelő szervezet összes munkája szerkezetileg átalakítható.

Előnyét az jelenti, hogy a statisztikai kiértékelés lehetőségei a felvételnél többletráfordítás nélkül és a katonai vezetésben a munkaerők számának emelése nélkül növekszenek. Sok helyen a megtakarítás a képzetlen munkaerők számánál érhető el. Ennek előfeltétele természetesen a gondosan átgondolt adatrendszer és a hatékony programozási segítség a kiértékelés részére.

Egy ilyen adatbank tervezésénél az érdekeltek között kiterjedt megbeszélést kell folytatni, amely a részletekre is kiterjed. Egyébként az előrehaladás helyett visszaesés következne be, mert a technikai területen és részben az anyag szakmai kiértékelésében is minden együttműködés kudarcba fulladna.

5. *Katonai adatbank tartalma*

Egy adatbank felépítésénél legelőször is a tartalomra vonatkozó kérdést kell tisztázni. Az utóbbi időben az „adatbank” címszó sok helyen jelentős aktivitást váltott ki olyan következtetéssel, hogy sok szakosított adatbankot kell létrehozni. Kétségtelen, hogy egy adatbankot a felhasználás céljának megfelelően kell kialakítani. Azonban téves felfogás lenne ebből az a végső következtetést levonni, hogy nagyon sok egyedi adatbankot kellene létesíteni. Ez ugyanis az információk kapcsolatát lényegesen megnehezítené. Ennek vagy információ-vesztés lenne a következménye (ugyanis meghatározott adatok kapcsolódását a széttagoltság lehetlenné teszi), vagy ugyanazon tények tárolása különböző helyeken válna szükségessé. A koncepció kialakításánál tehát az ésszerű középutat kell választani a nem racionális, nehezen kezelhető mammut adatbank és az adatok szétaprózása között. Ahogyan több oldalról már felismerték, a hivatalos statisztika — amelynek a gazdasági és társadalmi események nagy részét kell kiértékelnie — nyújtja a legjobb alapot a katonai és gazdasági tevékenységek nyilvántartására szolgáló információs központ létesítésére. Valószínű azonban, hogy néhány információt, amelyet jelenleg még más közigazgatási szervnél gyűjtenek, a saját statisztikai adatbankjába kell átvennie.

A statisztika információinak tárolása semmi esetre sem vonatkozhat egyetlen alapegységre. Sokkal inkább különböző egységek adatait kell rögzíteni. Ezért egy statisztikai adatbank kisebb adatbankok (alrendszerek) rendszeréből áll.

Egy alrendszer sokféle katonai statisztikai anyag regionális tagolását öleli fel. Legcélszerűbb egy ilyen regionális adatbank bázisának az ezredet választani. Hogy milyen ismérveket kell tartalmaznia, annak megállapítására pontos elemzésre van szükség. Mindenesetre a keretet viszonylag széleskörűen kell megállapítani.

További alrendszerek az anyag szakmai felosztására vonatkozhatnak. Különösen fontosnak látszik egy adatbank a katonai egységekről, területek szerinti bontásban. Itt az ismérvek széles skáláját kell tárolni.

Kiseb adatbankokat lehetne létesíteni a katonai intézményekről és egyéb katonai szervekről. Emellett egy alrendszer speciális idősorokat tartalmazna a gazdasági és szociális indikátorokról.

További lehetőségek, amelyek természetesen jelentősen nagyobb rá-

fordítást kívánnak, személyek vagy egyes személyi csoportok nyilvántartására létesítendő adatbankok.

6. Az adatbank megvalósítása

Egy adatbank gyakorlati felépítésével kapcsolatban három kérdés merül fel:

- az adatgyűjtés kezdetének időpontja,
- az adattárolás és hozzáférés módja, és a
- költségek.

Önmagában kívánatos, hogy az elmúlt időszakból lehetőleg sok információt vegyünk fel egy adatbankba. Mivel azonban egy ilyen bank felépítése jelentős munkát okoz és az eddigi adatok nagyrészt csak lyukkártyákon állnak rendelkezésre, a felépítésnél a régebbi időkből csak a legszükségesebb adatokat kell felvenni és elvileg egy új fordulópontból kell kiindulni. Sok esetben a nagy számlálások ajánlhatók erre.

Az adatok tárolására a mágnesszalag, a mágneslemez és a mágnesfilm látszik alkalmasnak. Nagyobbrészt a statisztikai adatállományt nem kell azonnali lehívásra készen tartani. Maga a közvetlen hozzáférés sem szükséges minden esetben. Ebből az következik, hogy a nagyobb állományokat mágnesszalagon, a kisebbeket és a gyakran használtakat lemezen kell elhelyezni. Amennyiben egy nagyobb állomány részére a közvetlen hozzáférés szükségessé válik, mágnesfilm-tárolóval dolgozhatunk.

Egy adatbank továbbvezetésénél felmerülő ráfordítást különösen figyelembe kell venni, ugyanis egy adatbank nem jelentheti csupán a statisztikai anyag egyszerű archiválását.

KÖVETKEZTETÉSEK

A tanulmány befejezéseképpen rá kell világítanunk az adatbankok koncepciója sikerének okaira, és el kell helyeznünk az adatbankokat a klasszikus vertikális vagy horizontális integrált rendszerek között.

Véleményem szerint az integrált rendszerek esetében a feldolgozások racionalizálásán van a hangsúly; ugyanakkor mindenki gyorsan számot adhat magának arról, hogy rendkívül nagyszabású célkitűzésekről van szó, amellyeknek megvalósítása az adatok előzetes összefüggésbe hozását teszi szükségessé.

Az adatbankok segítségével történő megoldás a feldolgozások kihangsúlyozása helyett igyekszik egycsapásra biztosítani az adatok kompatibilitását, és azután fokozatosan kialakítani egy adatbázist a feldolgozások harmonizálásával.

Az adatbankok koncepciójának sikere azzal magyarázható, hogy ez a megoldási mód pozitívnak bizonyult és, hogy a feldolgozások és az ügyvitel integrálása előzetesen megkívánja az adatok kompatibilitása problémájának megoldását.