

BIBLIOTHEKEN VOOR STATISTISCHE PROGRAMMATUUR

(C.J.A.Sprenger)

Inleiding.

Het aanbod van programmatuur voor sociaal-wetenschappelijk data-analyse vertoont een grote diversiteit, die evenwel niet alleen voort-spruit uit methodologische verschillen. De gebruiker dient zijn weg te vinden in een doolhof van applicatietalen, data-entry mogelijkheden, programmabeschrijvingen en vindplaatsen van programmatuur.

Hoewel de statistische programmatuur beschikbaar wordt gesteld met als doel het vereenvoudigen van althans de meer machinale aspecten van het sociaal-wetenschappelijk onderzoek, werkt deze verscheidenheid eerder bemoeilijkend en tijderspillend.

Aangezien de aan de Nederlandse universitaire rekencentra voorhanden zijnde programmatuur voor een deel is ondergebracht in pakketten en series als SPSS, OSIRIS, BMD, BMDP, CLUSTAN en MDS ¹⁾ terwijl er daarnaast uit een veelheid van bronnen afkomstige losse programma's zijn voor minder "ingeburgerde" en nieuwe methoden, wordt deze diversiteit aanmerkelijk in de hand gewerkt.

Voorals de groep losse programma's vertoont grote verschillen in applicatietalen, foutbehandeling, uitvoer en documentatie. ²⁾ Het standaardiseren van deze programma's en het vervolgens opnemen in een bibliotheek is dan een veel gekozen oplossing.

De bibliotheken KUNST, LISTOR en TCLIB

Omdat projecten voor dergelijke stroomlijning zowel in Nijmegen (KUNST) als in Groningen (LISTOR) als in Amsterdam (TCLIB) vrijwel onafhankelijk van elkaar gaande zijn, leek het belangrijk om een bespreking te organiseren tussen deelnemers aan deze ontwikkelingen.

1) Coxon's "Edinburgh-Cardiff series" voor multidimensional scaling, cf. van Schuur (1977)

2) Diversiteit op deze punten is binnen een pakket of serie doorgaans opgeheven.

Op deze 21 juni jl. te Nijmegen gehouden bijeenkomst werd allereerst uitvoerig aandacht besteed aan de opzet van KUNST³⁾ (Katholieke Universiteit Nijmegen Statistische Toepassingsprogramma's). Deze programmabibliotheek omvat programma's voor sociaal-wetenschappelijk en/of statistisch onderzoek, ontwikkeld dan wel aangepast voor computerverwerking in Nijmegen. De eerste stap in de richting van KUNST werd gezet dmv. bundeling van de know-how, aanwezig in de verschillende, onafhankelijk van elkaar opererende afdelingen, door een participatie van deze afdelingen in het Programmeursoverleg Sociale Fakulteit (PSF). Door het PSF werd in de loop van 1975 een grootscheepse reorganisatie van de standaard-programmaverzameling voorbereid uitmondend in het ontwerp van de bibliotheek KUNST. De uitgangspunten van deze reorganisatie waren:

1. Opname van programma's in KUNST is afhankelijk van beschikbaarheid in pakketten.
2. Oude standaardprogramma's dienen in nieuwe vorm beter op elkaar afgestemd te zijn.
3. De kwaliteit van de programma's dient goed te zijn.
4. Er moet een goede documentatie komen.

ad 1. KUNST is niet bedoeld als allesomvattend pakket maar als aanvulling op de faciliteiten die standaardpakketten als SPSS, OSIRIS, BMD, e.d. bieden. Dus nauwelijks datamanagement programma's maar daarentegen wel een groot aantal meer specialistische analyse technieken.

ad 2. Het op elkaar afstemmen van programma's behelst o.m. het elimineren van overlaps, het stroomlijnen van invoerspecificaties en het mogelijk maken van "chaining" (de uitvoer van het ene programma dient als invoer voor een tweede)

³⁾Zie ook Bendermacher et al (1977)

- ad 3. Ten aanzien van de kwaliteit van de programma's werden de volgende regels afgesproken.
- a) de programma's dienen eenvoudig te besturen zijn.
 - b) ze dienen controleerbaar te zijn, d.w.z. dat de gebruiker aan de hand van de uitvoer de korrekte verwerking van stuurkaarten en data en de juistheid van de berekeningen moet kunnen nagaan.
 - c) de programma's dienen alle informatie te verschaffen die nodig is om de resultaten te interpreteren.
 - d) de programma's dienen zoveel mogelijk systeem onafhankelijk te zijn.
 - e) de programma's zijn (vooralsnog) niet voor interactief gebruik maar wel moeten ze toegankelijk zijn via terminals.
- ad 4. De documentatie van KUNST valt uiteen in drie delen, te weten een algemene documentatie, voor elk programma een programmaspecifieke documentatie en tenslotte een programmeursgids. Wat betreft de programmaspecifieke documentatie wordt geregeld dat deze bestaat uit een zgn. Input/Output Description, waarin een korte inhoudelijke beschrijving, en beschrijvingen van besturing en van de uitvoer, verder is er een aparte Inhoudelijke Beschrijving indien deze te uitvoerig wordt voor de Input/Output Description, terwijl er tenslotte voor elk programma (in principe) een Programmers Notes is met uitvoerige informatie over algoritme, opmerkingen over uitwisselbaarheid, e.d.

Na de uiteenzetting over de opzet van KUNST werd zeer uitvoerig aandacht besteed aan de voor KUNST ontwikkelde applicatietaal TATA. Deze taal is echter niet in alle KUNST programma's ingebouwd: een aantal worden namelijk bestuurd via (gestandaardiseerde) "fixed format" stuurkaarten. De achtergrond voor deze tweezijdigheid van KUNST is dat twee verschillend georiënteerde afdelingen programma's aanleveren: de groep die stuurkaart-programma's aanlevert, heeft gebruikers die zelf hun jobs draaien, de andere groep verzorgt zelf de jobs voor de gebruikers, zodat deze groep de fixed-format besturing verkiest.

Wel houden beide groepen zich aan de "eisen en aanbevelingen t.a.v. KUNST programma's" zoals die worden gegeven in de programmeursgids voor KUNST.

Omdat een samenvatting van de mogelijkheden van de taal TATA buiten het bestek van dit artikel valt, volsta ik met een voorbeeld uit Bendermacher et al. (1977)

```
set-up [ //DEMO JOB (212,180),REMBRANDT
        //STAP1 EXEC KUNST,PROGRAM=MULTAN
        TITLE='BLOEM BLOEM BLOEM BLOEM',DATA=RAW,
        NAMES=('KELK-LEN','KELK-BR','BLOEM-L',BLOEM-W'),
        GROUP NAMES=(VIRGINI,VERSCOL,SETOSA),
        FORMAT=(4F7.3),SUBJ=(50,50,50),PRINT ALL,RESIDUAL,
        VARS=4;
data [ //DATA DD *
      120.002 10.157 15.923 11.548
      40.776 31.923 22.127 49.223
```

In de discussie volgend op de behandeling van TATA kwam onder meer het probleem naar voren dat de introductie van een nieuwe applicatietaal nu eenmaal met zich meebrengt, nl. dat de gebruikers wéér een nieuwe taal moeten leren. Van Nijmeegse zijde werd daartegen ingebracht dat bij de start van KUNST de grote verscheidenheid van de programma's gepaard aan de zeer wisselende groep gebruikers voor de consultatieafdelingen tot een onhoudbare situatie leidde. Bovendien was SPSS toentertijd in Nijmegen nog niet zo populair. Het accent werd voor KUNST dan ook gelegd op een eenvoudig te gebruiken taal in plaats van op uniformiteit der beschikbare talen.

De ontwikkeling van de Groningse programmabibliotheek LISTOR (Library for Statistical and Operations research Routines) vertoont veel punten van overeenstemming met die van KUNST. Een belangrijk verschil is evenwel dat LISTOR werd opgezet in een periode van populaire pakketten met daarnaast een vraag naar losse programmatuur die evenwel her en der verspreid was.

De LISTOR groep werd vooral opgericht om deze losse programma's te verzamelen en duplicaties te vermijden. De in september 1975 geformuleerde doelstellingen waren:

1. Alle beschikbare programmatuur verzamelen
2. Voorlopig geen standaardisering
3. Zéér summiere documentatie.

Momenteel bevat LISTOR 29 programma's, elk met een korte beschrijving. Het punt van de documentatie leidde in het verleden enige stagnatie, maar momenteel is men aardig bij. De groep heeft wel plannen voor standaardisering maar vreest daar voorlopig geen tijd voor te hebben. Men komt regelmatig bijeen om beschrijvingen, verspreiding, beheer van programmatuur, publiciteit en documentatie te coördineren.

De bibliotheek TCLIB is in zijn ontwikkeling nogal belemmerd door de zware belasting die het projekt STAP vormt. Recente toevoegingen zijn enkele programma's voor datamanagement (zoals het interactief cleanen van data). Wanneer het projekt STAP gereed is, zal een deel van de programma's uit TCLIB verdwijnen omdat ze in het pakket zijn opgenomen. De huidige versie van TCLIB bestaat uit programma's in FORTRAN (het merendeel), ALCOL-60 en PASCAL (een klein aantal). Een belangrijk deel wordt bestuurd via een eenvoudige applicatietaal, maar enkele programma's verwacht nog 'fixed format' stuurkaarten (met name geldt dit voor recent geïnstalleerde programma's als LINWOOD en de optimal subset selection programma's van Boyce et al). Naast TCLIB beheert het TC pakketten als SPSS en OSIRIS (gedeeltelijk) en de series BMD (gedeeltelijk), BMDP en MDS⁴).

In de slotdiscussie werd duidelijk dat alle aanwezigen het van belang achtten om standaardisering van programmatuur en documentatie met vertegenwoordigers van alle Nederlandse groepen te bespreken. Mijns inziens zouden daarbij de "eisen en aanbevelingen" uit de KUNST programmeursgids een belangrijke leidraad kunnen vormen.

⁴) Onlangs (september 1977) is alle losse programmatuur (inclusief de series) uniform toegankelijk gemaakt via het TCLIB retrieval programma RUNLIB

Referenties

- A.Bendermacher, M.van't Hof, J.van Leeuwe, Th.van der Weegen,
KUNST - Algemene Documentatie (1977)
- W.H.van Schuur, Inleiding LISTOR (1977)
- W.H.van Schuur, Computer programs for multidimensional scaling,
MDN, 3, 45-55 (1977)
- TC-FSW, Index van TC-programmatuur (1977)
-

C.J.A.Sprenger,
Universiteit van Amsterdam
Technisch Centrum FSW
Roeterstraat 15
Amsterdam