
CLAMOUR (Classifications MOdelling and Utilities Research)

De economie van de Europese landen bevindt zich in een stroomversnelling. Nieuwe eenheden met nieuwe activiteiten volgen elkaar snel op. Dit stelt eisen aan de flexibiliteit waarmee eenheden en hun activiteiten worden vastgesteld en het is een uitdaging aan de statistiek om deze ontwikkelingen in beeld te brengen. Het spreekt vanzelf dat dit op een geharmoniseerde wijze dient te gebeuren voor alle lidstaten van de Europese Unie. Vanwege de internationalisering van economische activiteiten groeit daarnaast ook de behoefte aan harmonisering op wereldniveau.¹

CLAMOUR

Het Europese samenwerkingsproject² CLAMOUR (classifications modelling and utilities research) leverde een modelmatige beschrijving op van de structuur en de activiteiten van bedrijven. Het model maakt een flexibel en geharmoniseerd gebruik van eenheden en activiteitenclassificaties in economische statistieken mogelijk. Op basis van informatie uit het model kan men enerzijds statistische eenheden formeren en classificeren en anderzijds bestaande activiteitenclassificaties onderhouden of alternatieve classificaties produceren.

Clamour bestond uit de volgende vier deelprojecten:

1. De organisatie van een Europees gebruikersonderzoek waarin is onderzocht wat het gebruik is van activiteitenclassificaties en definities van statistische eenheden en wat de wensen van de gebruikers zijn.
2. Het verder uitbouwen van een eerder ontwikkeld model om de structuur en de activiteiten van bedrijven te beschrijven.

3. Het ontwikkelen en testen van een elektronische vragenlijst om de benodigde informatie van bedrijven te verkrijgen.
4. Het ontwikkelen van algoritmes om statistische eenheden te construeren en te classificeren naar hun activiteit op grond van de informatie in het model. Daarnaast ook het ontwikkelen van theorieën op het gebied van actualisatie van bestaande activiteitenclassificaties en het bouwen van nieuwe activiteitenclassificaties.

Onderzoeksgebieden

User Needs

Nationale statistische bureaus verzamelen en classificeren informatie om een zo accuraat en actueel mogelijk beeld te geven van de nationale economische ontwikkelingen. Statistische bureaus van internationale organisaties, zoals Eurostat van de Europese Unie en het United Nations Statistics Division van de Verenigde Naties, combineren deze data voor de beschrijving van internationale ontwikkelingen. Overheden kunnen hiermee rekening houden bij het uitstippelen van hun politiek en het bedrijfsleven kan daarop zijn beleid baseren. Ook voor particulieren is het van belang de vinger aan de economische pols te houden, bijvoorbeeld bij de beslissing voor belangrijke aankopen. Zulke statistische informatie wordt daarom zo breed mogelijk en vaak gratis beschikbaar gesteld.

Om een zicht te krijgen op het gebruik van deze informatie heeft elk aan Clamour deelnemend nationaal statistisch bureau in eigen land een aantal belangrijke gebruikers van bedrijfsstatistiek met een vragenlijst benaderd.³ De vragen gingen over het huidige gebruik van informatie en welke mogelijke toekomstige veranderingen in de te verzamelen gegevens en de presentatie daarvan gewenst zijn. Sommige vragen waren toegespitst op wensen voor de komende herziening van de internationale activiteitenclassificaties in 2007.

Dit 'klantenonderzoek' gaf aan dat bij toekomstige wijzigingen van standaarden gezocht moeten worden naar een evenwicht tussen de behoefte aan in de tijd vergelijkbare statistieken en de behoefte aan actuele classificaties. Verder moet gedacht worden aan de enquêtedruk op de veelal tot het bedrijfsleven behorende dataleveranciers. Detailleringen brengen immers extra enquête-vragen met zich mee, niettegenstaande de vraag naar detailleringen vaak uit het bedrijfsleven komt.

Foundations

In het onderzoek naar de grondslagen werden fundamentele bouwstenen (building blocks) onderscheiden waaruit diverse eenheden en de activiteitencode konden worden afgeleid. De bouwstenen werden gedefinieerd door de criteria die Eurostat stelt aan de verschillende eenheden en de criteria die ten grondslag liggen aan de NACE (de Europese standaardclassificatie voor economische activiteiten), alsmede de nationale varianten daarop van de deelnemende landen. Het model staat toe dat ook andere criteria in het model worden opgenomen om tot andersoortige eenheden of activiteitenclassificaties te komen. Foundations omvatte drie werkprogramma's:

- Het construeren van een model om de structuur en activiteiten van bedrijven te beschrijven. Hierbij ging het om een database en de algoritmen om tot de statistische eenheden en activiteitencode te komen.
- Het toepassen van het model op bestaande activiteitenclassificaties. Het model kan gebruikt worden om individuele bedrijven te coderen, maar ook om activiteitenclassificaties te actualiseren en nieuwe te genereren.

- Het toepassen van het model om statistische eenheden te vormen.

De resultaten van het onderzoek naar de User Needs hebben het onderzoek naar de Foundations gevoed. De verkregen inzichten zijn voorgelegd aan degenen die verantwoordelijk zijn voor de toepasbaarheid van de NACE. Dit zal moeten leiden tot een betere statistische beschrijving van bedrijfsstructuren en bedrijfsactiviteiten.

Linguistics

Binnen dit deelproject zijn linguïstische methoden en programmatuur ontwikkeld die zo goed mogelijk de betekenis van beschrijvingen vatten. De snelheid en kwaliteit van het bepalen van statistische eenheden en hun activiteiten op basis van door bedrijven beschikbaar gestelde informatie kan hiermee verbeterd worden. Bestaande programma's op dit gebied schieten veelal tekort in het herkennen van de betekenis van geschreven tekst.

Het zou moeten mogelijk zijn om de voor het model relevante informatie af te leiden uit het profiel van processen, input en output, structuur etc. van een bedrijf. Binnen een bepaalde taal en binnen de Europese gemeenschap bestaan nuances die verschillen in betekenis impliceren. Binnen Linguistics is nagegaan hoe deze verschillen zo goed mogelijk kunnen worden geminimaliseerd. Dit kan leiden tot interactieve systemen die intelligente hulp bieden bij het verwerken en interpreteren van beschikbare informatie. Het linguïstische werk bestond uit vier werkprogramma's:

- Het ontwikkelen van een context-gebaseerde procedure voor het opheffen van semantische vaagheid.
- Het verwerken van samengestelde termen.
- De analyse van sterk gestructureerde beschrijvingen.
- Het computerondersteund vergelijken van activiteitenclassificaties en productclassificaties onderling.

Tools & Systems

Het resultaat van de hiervoor genoemde projecten leverde informatie op voor degenen die betrokken

waren bij de ontwikkeling van de elektronische vragenlijst. Met deze vragenlijst kon de benodigde informatie bij bedrijven op een gebruiksvriendelijke wijze worden verzameld. Via routing door de elektronische vragenlijst kreeg elk bedrijf namelijk alleen de voor haar relevante vragen voorgeschoteld. Er werd zoveel mogelijk gewerkt met gesloten vragen, al konden deze honderden antwoordmogelijkheden omvatten. Alle voor het bepalen van de eenhedenstructuur en activiteitenclassificatie benodigde informatie kon zo systematisch binnengehaald worden.⁴

*Hans van Hooff
Peter Struijs
Leon Willenborg
Centraal Bureau voor de Statistiek
(Nederland – Voorburg)*

Noten

1. Op de CLAMOUR-website http://www.statistics.gov.uk/methods_quality/clamour.asp is meer uitgebreide informatie te vinden over het project en de participanten. De bevindingen zijn eveneens op deze website gepubliceerd. Voor nadere informatie kunt u contact opnemen met Hans van Hooff, Peter Struijs of Leon Willenborg van het Centraal Bureau voor de Statistiek van Nederland (HHFF@CBS.NL, PSJS@CBS.NL en LWLG@CBS.NL).
2. Het project maakte onderdeel uit van en werd gefinancierd vanuit het Fifth Framework Onderzoeksprogramma van de Europese Commissie. Het project werd in samenwerking uitgevoerd door de statistische bureaus van vijf EU-landen, namelijk Denemarken, Finland, Frankrijk, Nederland en het Verenigd Koninkrijk. Tevens waren bij het project twee bedrijven betrokken.
3. Daarnaast deden de statistische bureaus van een aantal andere landen ook mee.
4. Deze elektronische vragenlijst is op bescheiden schaal getest in Finland, Nederland en het Verenigd Koninkrijk.