



Rapport Text Mining 1:

Studie over het toegankelijk maken van methodologische gegevens in systemen voor het ontsluiten van statistische informatie.

Inleiding

Statistische data worden veelal vergezeld van methodologische noten. Methodologische noten zijn nota's die gekoppeld worden aan statistische data ter verduidelijking van deze statistische data. In deze noten staan dingen zoals de methodologie die gebruikt wordt bij het aanmaken van de statistische data, gebruikte definities en concepten, de nauwkeurigheid van de statistische data, ... Dit rapport beschrijft het gebruik van methodologische noten op statistische websites voor het toegankelijk maken van statistische data. Het beschrijft enkele veel gebruikte standaarden, die aangeven hoe je het best methodologische informatie aanbiedt en welke methodologische informatie je het best aanbiedt. Verder bespreekt en evalueert dit rapport verschillende web portals van belangrijke internationale statistische instituten.

Dit rapport zal dan gebruikt worden om een evaluatie te houden over de disseminatie van methodologische gegevens op de site van de Nationale Bank van België. Naast de evaluatie zullen ook aanbevelingen gedaan worden die moeten dienen om het gebruik van de methodologische nota's te verbeteren.

Ik zal beginnen met het beschrijven van twee internationale standaarden in het gebruik en aanbieden van statistische gegevens en hun bijhorende procedures ter controle van de kwaliteit van de statistische organisatie. Uiteraard zal ik vooral ingaan op het aanbieden van methodologische informatie en minder op de rest van de statistische data en metadata. Naast het bespreken van deze standaarden ben ik ook enkele belangrijke statistische sites gaan bezoeken om uit eigen ondervinding te gaan zien wat er goed is aan bepaalde sites, wat er ontbreekt, wat overbodig is enz...

I. Studie van bestaande standaarden

I.1 De *Special data dissemination standard* en de *Data quality reference*

Het *International Monetary Fund* (IMF) heeft een standaard ontwikkeld die kan gebruikt worden als gids ter verbetering van de disseminatie van statistische gegevens, de *special data dissemination standard* (SDDS) genoemd. Naast deze standaard hebben ze ook voor een gestandaardiseerde omkadering (*data quality assessment framework* of DQAF) gezorgd waarbinnen de kwaliteit van een statistische organisatie kan gemeten worden. In dit deel zal ik een korte uiteenzetting geven van wat deze twee standaarden te vertellen hebben omtrent de methodologische informatie.

De *special data dissemination standard* is een gids die beschrijft wat er in de methodologische nota's moet komen. Ze focust zich dus vooral op wat er in de methodologische documenten moet staan (de inhoud) en minder op hoe ze gepresenteerd moeten worden. De standaard splitst het deel over methodologische noten op in twee delen. In het eerste deel wordt vooral iets verteld over revisies en aanpassingen aan methodologische nota's. In het tweede deel wordt vooral gesproken over welke dingen zeker behandeld moeten worden in deze documenten.

Wanneer er bepaalde aanpassingen gemaakt worden, kunnen deze best opgeslagen worden als een nieuwe versie van het methodologisch document. Dit zodoende de oudere gegevens niet te verliezen. De gebruiker moet duidelijk op de hoogte gebracht worden van het feit dat er aanpassingen gebeurd zijn en de reden van aanpassing, de voordelen enz. kunnen best vermeld worden.

De standaard maakt vervolgens ook suggesties over wat er best in de standaard vermeld staat. Dit zijn echter suggesties aangezien niet alles nuttig is of zelf mogelijk is. Het hangt zeer sterk af van het soort informatie waarover de methodologische nota gaat en de statistische organisatie achter die statistische data. De volgende lijst met onderwerpen is gegeven:

- *Definities, concepten, classificaties*: Indien er een standaard is gebruikt waarop deze dingen steunen kan dit best vermeld worden. Eventuele afwijkingen moeten ook vermeld worden.
- *De draagwijdte van de data*: Gaat het over een bepaald land, de wereld, ... Indien het bij voorbeeld over een land gaat en de gegevens van enkele gebieden ontbreken, moet dit ook vermeld worden.
- *Accounting-conventies*: Informatie over wanneer de informatie genoteerd is enz.
- *De natuur van de basis data*: Hoe worden de data verzameld, hoe verloopt het samplen,...
- *Compilatieprocedure*: Dit deel beschrijft de gebruikte procedures zoals gewicht schema's die gebruikt worden, wat er gebeurt met ontbrekende waarden,...
- *Overige aspecten*: Zoals basisjaar, referentiejaar enz.

Naast deze suggestie heeft het IMF ook een framework ontwikkeld waarbinnen ze de kwaliteit van bepaalde statistische organisaties en hun bijhorende site kunnen meten. In dit deel ga ik iets meer vertellen over de *data quality assessment framework* (DQAF) en specifiek over het meten van de kwaliteit van de methodologische noten.

Het framework bestaat uit één generische framework en een paar specifiekere frameworks die zijn opgesteld voor de hoofdaggregaten gebruikt in macro economische analyses. Het generische framework wordt gebruikt als een model voor het opstellen van deze specifiekere frameworks. De verschillende deeldomeinen waarvoor IMF een framework heeft opgesteld zijn: national accounts, consumer price index, producer price index, government finance statistics, monetary statistics en balance of payment.

IMF deelt de kwaliteits controle op in vijf grote delen: integriteit (integrity), methodologische correctheid (methodological soundness), nauwkeurigheid en betrouwbaarheid (accuracy and reliability), bruikbaarheid (serviceability) en toegankelijkheid (accessibility). Voor elk van deze delen zijn er indicatoren opgesteld die gebruikt worden bij de metingen. Uiteraard is het deel 'methodological soundness' het belangrijke gedeelte voor ons onderzoek. Methodologische correctheid wordt door IMF omschreven als:

The dimension of quality control that covers the idea that the methodological basis for the production of statistics should be sound and that this could be attained by following international standards, guidelines and agreed practices. In application, this dimension will necessarily be dataset-specific, reflecting differing methodologies for different datasets (for example the 1993 SNA for national accounts and the fifth edition of the Funds 'Balance of payments manual' for balance of payments) [1].

Zoals je in het schematisch overzicht van het methodological soundness gedeelte van het generische framework (zie tabel 1) kunt zien is dit gedeelte zeer afhankelijk van de specifieke dataset. Dit komt omdat de kwaliteit van de gebruikte methodologie gedeeltelijk af hangt van hoe goed een internationale standaard wordt gevolgd. Uiteraard is het gebruik van een standaard zeer specifiek voor de dataset.

Quality Dimensions	Elements	Indicators
Methodologische correctheid De methodologische basis voor de statistieken volgen internationaal aanvaarde standaarden.	2.1 Concepten en definities Concepten en definities zijn in overeenstemming met internationale statistische frameworks. 2.2 Draagwijdte De draagwijdte is in overeenstemming met internationaal aanvaarde standaarden. 2.3 Classificatie/sectorisatie De classificatie en sectorisatie zijn in overeenstemming met internationaal aanvaarde standaarden. 2.4 De basis van het opmeten <i>Flows</i> en <i>stocks</i> worden opgemeten volgens de internationaal aanvaarde standaarden.	2.1.1 De structuur onder de concepten en de definities volgen de internationale standaarden. 2.2.1 De draagwijdte is consistent met de internationaal aanvaarde standaarden. 2.3.1 Classificatie en sectorisatie systemen zijn consistent met de internationaal aanvaarde standaarden. 2.4.1 De marktprijzen worden gebruikt voor het waarderen van <i>flows</i> en <i>stocks</i> . 2.4.2 De metingen gebeuren op een nauwkeurige basis. 2.4.3 <i>Grossing/netting</i> procedures volgen internationaal aanvaarde standaarden.

Tabel 1: Methodologische gezondheid

Wat de dataset specifieke frameworks op hun beurt dan nagaan is dus of de vier bovenstaande punten in overeenstemming zijn met de standaard gebruikt in die bepaalde sector. Bij voorbeeld in de *balance of payments statistics* wordt er nagegaan of de vier bovenstaande punten conform zijn met de *Balance of payments manual (BMP5)* opgesteld in '95.

I.2 Eurostat

Eurostat is een organisatie die statistische informatie van Europa verzamelt. Ze gebruiken de richtlijnen voor statistische metadata opgesteld door de Verenigde Naties. Ze zorgen er anderzijds ook wel voor dat hun site conform is met de *special data dissemination standard*. De gids is opgesteld voor metadata maar aangezien de methodologische nota's een onderdeel is van de metadata kunnen we ze ook gebruiken. De metadata worden in de richtlijnen onderverdeeld in 3 categorieën:

- (1) Metadata gebruikt voor het zoeken en navigeren
- (2) Metadata gebruikt voor het interpreteren
- (3) Metadata gebruikt in het post-processen

Er is een belangrijk verschil in metadata en methodologische noten. Metadata bevat alle informatie over de statistische data. Dus de methodologische nota's zijn een deel van de metadata. De belangrijkste categorie van metadata waaronder de methodologische documenten onder vallen is categorie twee aangezien het belangrijkste doel van een methodologie het duidelijk maken is van welke procedures er gebruikt zijn enz. Volgende dingen worden vermeld in de tweede categorie: de classificatie, standaarden die gebruikt worden, wat te doen met ontbrekende data, de vergelijkbaarheid met andere alternatieve bronnen, beschrijving van de gebruikte methoden voor het verzamelen van de gegevens, de berekeningen op de data, de gebruikte afrondingen, benaderingsmethoden, ... In deel drie wordt er verteld dat wanneer bepaalde statistische gegevens worden gedownload, de bijhorende methodologische noten ook moeten meegestuurd worden. Dit vermijdt problemen in het post-processen van de informatie.

Categorie één is echter ook een heel belangrijke categorie. Je kunt niet enkel metadata van de statistische informatie hebben maar je kunt ook metadata van de methodologische nota's hebben. Om het zoeken en navigeren naar methodologische noten gemakkelijker te maken kun je metadata toekennen aan de methodologische noten. Dit kan bijvoorbeeld het domein zijn, de naam van de auteur,... Als je dan een bepaalde nota zoekt kun je dan niet enkel full-text search gebruiken maar kun je ook gebruik maken van deze metadata wat meestal het zoeken efficiënter maakt. De metadata die de methodologische noten vervoegen kunnen ook gebruikt worden om een boomstructuur van alle methodologische noten op te bouwen.

Naast de bovenstaande onderverdeling van de metadata, vermeldt de standaard ook iets over de opdeling van de gebruiker in verschillende types. Ze zeggen dat er ten eerste een onderscheid moet gemaakt worden tussen de gebruikertypes en ten tweede dat het aanbieden van de methodologische informatie moet afhangen van de kennis en noden van die gebruiker. In de VN standaard wordt een onderscheid gemaakt tussen volgende types:

- (1) De gebruiker met een zeer beperkte kennis van statistiek
- (2) De gebruiker met de nodige kennis van statistiek maar met een beperkte interesse in de methodologische nota's
- (3) Experts.

Net zoals IMF heeft Eurostat ook een framework ontwikkeld dat kan gebruikt worden voor een kwaliteitscontrole van de statistische informatie en de bijhorende organisatie. Het grote verschil is dat het IMF framework zich concentreert op de inhoud van de methodologische documenten terwijl Eurostat zich (baserend op de ISO definitie van kwaliteit) eerder richt op de gebruiker. Eurostat verdeelt de kwaliteit op in zeven categorieën:

- (1) Relevantie
- (2) Nauwkeurigheid
- (3) Timing en stiptheid
- (4) Toegankelijkheid en duidelijkheid
- (5) Vergelijkbaarheid
- (6) Samenhangendheid
- (7) Volledigheid

Voor elk van deze zeven categorieën zijn er enkele vragen opgesteld die beantwoord moeten worden in een rapport over de kwaliteit. Enkele van deze categorieën kunnen ook gebruikt worden bij het expliciet evalueren van de kwaliteit van methodologische noten.

Relevantie: Dit onderdeel onderzoekt of er aan de noden van de gebruiker voldaan wordt. Indien mogelijk, moet er ook een opsplitsing gemaakt worden tussen de verschillende types gebruiker, en de hoeveelheid en type methodologische informatie die worden aangeboden, moet hieraan gekoppeld worden.

Nauwkeurigheid: Iets dat zeker in een methodologische nota moet komen is de nauwkeurigheid van de verkregen data. De nauwkeurigheid hangt onder meer af van de manier van samplen, de meetfouten die kunnen voorkomen, de procedures die de ontbrekende informatie behandelen, ...

Timing en stiptheid: Niet echt van toepassing aangezien dit deel gaat over het al dan niet tijdig uitbrengen van bepaalde statistische informatie.

Toegankelijkheid en duidelijkheid: In dit deel wordt de toegankelijkheid van de methodologische documenten besproken. De toegankelijkheid wordt verbeterd door bij voorbeeld het integreren van een zoekmachine. Verder moeten de methodologische documenten duidelijk zijn en niet verschillend kunnen geïnterpreteerd worden.

Vergelijkbaarheid: Dit is één van de belangrijkste categorieën voor de methodologische noten. Eén van de redenen waarom methodologische noten zo belangrijk zijn, is de vergelijkbaarheid van data van een bron met data van een andere bron. De manier om deze data te vergelijken is gaan zien welke verschillende procedures er gebruikt zijn. De vergelijkbaarheid kan nog eens onderverdeeld worden in 3 delen:

- a. geografische vergelijkbaarheid
- b. vergelijkbaarheid in tijd
- c. vergelijkbaarheid tussen domeinen

De methodologische nota's moeten deze drie delen ondersteunen. Als er bijvoorbeeld in verschillende landen een andere standaard gebruikt wordt, of als je een standaard gebruikt die afwijkt van de internationaal aanvaarde standaard, moet dit ergens vermeld worden. Als de gebruikte methodologie aangepast is tegenover de vorige (oudere) versie, moet dit vermeld worden in de methodologische nota, ...

Samenhangendheid: Sommige tabellen worden geconstrueerd vanuit verschillende bronnen. Deze verschillende bronnen kunnen verschillende technieken en procedures gebruiken. Deze verschillen moeten ook duidelijk uitgelegd worden aan de gebruiker van die tabel.

Volledigheid: Bij elke tabel zou een methodologische nota moeten hangen.

II. Onderzoeken van verschillende sites

Op basis van een eerste verkenning van volgende statistische websites:

- (1) Statistics Canada (<http://www.statcan.ca/>)
- (2) Eurostat (<http://europa.eu.int/comm/eurostat/>)
- (3) Statec of Statistics Luxemburg (<http://statec.gouvernement.lu/>)
- (4) Statistics Finland (http://www.stat.fi/index_en.html)

heb ik een lijst met evaluatieparameters opgesteld.

III. Lijst met evaluatieparameters

Ik geef nu een overzicht met een kort beschrijving van de gebruikte parameters. Deze worden opgedeeld in verschillende categorieën naar gelang hun onderwerp. Bij elke parameter staat een maatstaf, deze geeft aan hoe het resultaat wordt weergegeven.

A. Algemeen

(A.1) Layout

A.1.1 Uniformiteit van de layout: Zien de methodologische nota's er allemaal hetzelfde uit? Dit zorgt ervoor dat de gebruiker gewend geraakt aan waar bepaalde informatie in een methodologisch document zit.

Maatstaf: Ja/Nee

A.1.2 Duidelijkheid van de layout: Is het een overzichtelijke layout waar je gemakkelijk bepaalde informatie in terug vindt? Soms heb je van die lange doorlopende teksten die niet zijn onderverdeeld in hoofdstukken of die slecht omgaan met paragrafen. Deze zijn uiteraard minder leesbaar dan goed gestructureerde teksten.

Maatstaf: Ja/Nee

A.1.3 Zijn de methodologische nota's opgedeeld in logische eenheden: Zijn de methodologische noten gewoon tekstfiles zonder een duidelijke onderverdeling of zijn ze onderverdeeld in logische blokken? Dit laatste zorgt er ook weer voor dat de gebruiker bepaalde delen sneller terug zal vinden.

Maatstaf: Ja/Nee/Soms

A.1.4 Uniformiteit van de onderverdeling: Indien ze inderdaad zijn onderverdeeld in logische eenheden, gebeurt dit dan voor elk document op dezelfde manier, hebben ze dezelfde titel, ...? Ook dit bevordert de leesbaarheid van het documenten, de snelheid waarmee je bepaalde informatie terugvindt, enz.

Maatstaf: Ja/Nee

(A.2) Gebruikersgericht

A.2.1 Onderscheid van verschillende types gebruiker: Wordt er op de site een onderscheid gemaakt tussen de verschillende types gebruiker? Indien er onderscheid gemaakt wordt kan er ingespeeld worden op de specifieke eisen en noden van de verschillende gebruikerstypes (zie volgende criteria).

Maatstaf: Ja/Nee

A.2.2 Wordt de methodologische informatie in verschillende stappen aangeboden: Het kan zijn dat de methodologische informatie in verschillende stappen wordt aangeboden. In dat geval krijg je dan eerst niets te zien, dan een korte samenvatting en dan de volledig gedetailleerde beschrijving. In andere gevallen zit je in een alles of niets situatie waar je ofwel alle methodologische informatie te zien krijgt ofwel krijg je geen methodologische informatie te zien.

Maatstaf: Ja/Nee/Soms

A.2.3 Wordt het type en/of de hoeveelheid methodologische informatie gekoppeld aan het type gebruiker: Elk type gebruiker heeft zijn eigen noden en de aangeboden informatie moet hieraan aangepast worden. Zo zijn er gebruikers die geen interesse hebben in methodologische informatie of niet genoeg kennis hebben van statistische gegevens om zich bezig te houden met methodologische nota's, deze gebruikers moeten niet overstelpt worden met methodologische gegevens, dit in tegenstelling tot experts die deze informatie enorm belangrijk vinden om de statistische informatie te kunnen inschatten.

Maatstaf: Ja/Nee/Soms

B. Zoeken en navigeren

(B.1) Zoeken

B.1.1 Is het mogelijk om te zoeken naar methodologische informatie? Meestal kan er naar statistische informatie gezocht worden, kunnen bepaalde tabellen opgevraagd worden, kunnen publicaties met statistische informatie opgezocht worden,... Soms is het ook mogelijk om via zoekoperaties methodologische noten op te vragen. Dit is vooral interessant voor mensen die geïnteresseerd zijn in deze nota's.

Maatstaf: Ja/Nee

B.1.2 Wordt bij het zoeken full-text search gebruikt: Er zijn verschillende manieren van zoeken. Eén van de manieren houdt in dat je alle woorden van de methodologische tekst kunt gebruiken.

Maatstaf: Ja/Nee

B.1.3 *Kan er gezocht worden op meta-informatie:* Soms hebben bepaalde methodologische documenten ook metadata zoals de auteur, de datum, ... Is het mogelijk om op deze attributen te zoeken?

Maatstaf: Ja/Nee

B.1.4. *Indien er kan gezocht worden op metadata, bevatten zij ook vaste trefwoorden die de inhoud beschrijven?*

Maatstaf: Ja/Nee/Niet van toepassing (NVT)

B.1.5 *Is het een goede zoekmachine:* Dit geeft een beoordeling van de nauwkeurigheid en performantie van de zoekmachine.

Maatstaf: N=Nauwkeurig, P=Performant, S=Slecht → N/P/NP/S

(B.2) Navigeren

B.2.1 *Is het gemakkelijk om bepaalde methodologische documenten te vinden:* Als je bezig bent met een bepaalde tabel, kun je dan gemakkelijk de bijhorende methodologische nota vinden. Je moet ervoor zorgen dat de gebruiker niet gefrustreerd geraakt doordat hij te hard moet zoeken naar de juiste informatie.

Maatstaf: Ja/Nee

B.2.2 *Is de link tussen een tabel en zijn methodologische nota duidelijk en uniform:* In de meeste gevallen zit er bij een tabel wel een link naar de bijhorende methodologische nota. Dit zou best steeds op de zelfde manier moeten gebeuren zodat de gebruiker het gewoon wordt en gemakkelijk de bijhorende methodologische nota terug vindt.

Maatstaf: Ja/Nee

B.2.3 *Is er een overzicht van alle methodologische nota's:* Soms kun je alle methodologische nota's terug vinden op een gecentraliseerde plaats. Dit is gemakkelijk voor gebruikers die enkel geïnteresseerd zijn in de methodologische nota's.

Maatstaf: Ja/Nee

B.2.4 *Is dit overzicht duidelijk en goed georganiseerd:* Is het gemakkelijk om een bepaald document terug te vinden in het overzicht? Dit kan bij voorbeeld verbeterd worden door ze te ordenen per domein.

Maatstaf: Ja/Nee

C. Revisies

(C.1) Revisies

C.1.1 *Worden de methodologische noten up-to-date gehouden:* Indien de gebruikte methodologie verandert, moet de methodologische nota ook aangepast worden.

Maatstaf: Ja/Nee

C.1.2 *Worden deze aanpassingen als versies bewaard:* Dit gaat het verlies van oudere methodologische informatie tegen. Het bijhouden van oudere methodologische informatie is belangrijk voor een paar redenen. Ten eerste moet oudere data, data die berekend is met een oudere methodologie, voorzien worden van de juiste methodologische informatie. De gebruikers moeten weten hoe die specifieke data berekend is en hebben dus geen boodschap aan het meest recente methodologische document. Een tweede belangrijke reden voor het bijhouden van methodologische informatie is dat je gemakkelijker de vergelijking kunt maken tussen oudere en nieuwere methodologieën. Voor de gebruikers die echt geïnteresseerd zijn in methodologische informatie is dit toch belangrijk.

Maatstaf: Ja/Nee

C.1.3 *Is er een connectie tussen de data en de juiste versie van methodologie:* Tabellen die een oudere versie van een methodologie gebruiken, moeten gelinkt worden met die oudere versie en niet met de laatste nieuwe.

Maatstaf: Ja/Nee/Soms

C.1.4 *Wordt de gebruiker voldoende op de hoogte gebracht van een nieuwe versie:* De gebruiker moet gewaarschuwd worden dat hij met een nieuwe versie bezig is en dat er oudere versies bestaan.

Maatstaf: Ja/Nee/Soms

D. Inhoud

(D.1) Standaard

D.1.1 *Wordt er een standaard gebruikt:* Indien er een bepaalde standaard wordt gebruikt moet deze zeker vermeld worden in de methodologische nota. Met een standaard bedoelen we een standaard methodologie. Verschillende statistische tabellen worden opgesteld door gebruik te maken van internationaal aanvaarde formules. Zo kan het zijn dat je bij voorbeeld bij het maken van de statistische gegevens van de '*Balance of payments*' gebruik maakt van BPM5 ('the fifth Balance of payments manual'). In de methodologische nota moet dan vermeld worden dat je je methoden hierop baseert.

Maatstaf: Ja/Nee

D.1.2 *Verschillen met de gebruikte standaard:* Als je een standaard gebruikt, maar je hebt toch enkele verschillen, moeten deze duidelijk vermeld worden.

Maatstaf: Ja/Nee/Soms

D.1.3 *Redenen van verschil met de gebruikte standaard:* Indien er verschillen zijn met de gebruikte standaard, kun je best ook vermelden waarom ze er zijn.

Maatstaf: Ja/nee/Soms

D.1.4 *Vergelijkbaarheid:* Kun je gemakkelijk vergelijkingen maken met andere standaarden gebruikt in andere landen, oudere gegevens, een ander domein,... In sommige methodologische documenten staan zelf de gemaakt vergelijkingen. Dit zorgt ervoor dat gebruikers die twee statistische tabellen proberen te vergelijken duidelijker te zien krijgen wat de eventuele verschillen in methode zijn.

Maatstaf: Ja/Nee

(D.2) Inhoud

D.2.1 *Staat de nauwkeurigheid e.d. vermeld in de methodologische nota's:* De nauwkeurigheid van bepaalde methoden moeten zeker vermeld worden. Samenhangende gegevens zoals de sample procedures, de meetfouten die kunnen optreden en wat er gebeurt met ontbrekende data, kunnen hierbij vermeld worden.

Maatstaf: Een opsomming van wat er allemaal in vermeld wordt → N (nauwkeurigheid) / S (sample methode) / M (de meetfouten) / O (wat er gebeurt met ontbrekende data).

D2.2 *Wordt de bron vermeld?*

Maatstaf: Ja/Nee

E. Extra

(E.1) Extra

E.1.1 *Is er één methodologische nota per domein, per tabel of per tijdsreeks?*
Soms staat de methodologische nota bij het domein zodoende dat alle gebruikte methodologieën die in dat domein thuis horen, vermeld staan in dat document. Andere keren heb je 1 methodologische nota per tabel of tijdsreeks. Alles heeft zijn voor- en nadelen. Zo is het gemakkelijk om 1 document te lezen voor heel het domein in plaats van alle afzonderlijke documenten te lezen. Anderzijds is het soms handiger dat je een methodologische nota per

tabel hebt zodanig dat je direct de relevante informatie over die tabel terug vindt en niet in heel het document moet gaan zoeken.

Maatstaf: Domein/Tabel/Tijdsreeks

E.1.2 *Heeft elke tabel een methodologisch document (analoog domein, tijdsreeks als dat meer van toepassing is):* Dit gaat over de volledigheid van de methodologische nota's. In principe moet elke tabel toch zijn eigen methodologische nota hebben. Dit is spijtig genoeg (voor de gebruiker althans) niet altijd mogelijk doordat je bepaalde informatie niet hebt, doordat diegenen die de tabellen opstelt niet altijd de tijd of de motivatie ervoor vindt,...

Maatstaf: Ja/Nee

E.1.3 *Indien de gebruiker zelf tabellen kan samenstellen, worden de methodologische nota's dan ook aangepast:* Soms is het mogelijk om door het selecteren van enkele parameters een nieuwe tabel te construeren. Indien dit mogelijk is moet er eigenlijk een aangepaste versie van de methodologische nota komen (een combinatie van de methodologische nota's die horen bij de gecombineerde tabellen).

Maatstaf: Ja/Nee/Niet van toepassing (NVT)

E.1.4 *Worden er verschillende talen ondersteund:* Aangezien we toch spreken over de disseminatie van statistische informatie, is het belangrijk dat de methodologische documenten in verschillende talen aangeboden worden. Buiten de officiële talen van het land kan best ook het Engels ondersteund worden (aangezien dit toch algemeen aanvaard wordt als de wereldtaal).

Maatstaf: De lijst met alle ondersteunde talen. → N (Nederlands) / D (Duits) / E (Engels) / F (Frans) / It (Italiaans) / Fin (Fins)

E.1.5 *Als je statistische gegevens download, worden de methodologische documenten dan mee gedownload:* Om consistentie redenen en om het post-processen te verbeteren zou je dit best doen. Op zijn minst moet de gebruiker de vraag gesteld worden of hij de methodologische nota's mee wil downloaden.

Maatstaf: Ja/Nee/Soms/Niet van toepassing (NVT)

E.1.6 *Zijn de populaire bestandstypes beschikbaar:* Zijn de methodologische documenten beschikbaar als pdf, word document, html file,...

Maatstaf: Opsomming van de ondersteunde types → Html / Txt (tekst bestand) / Excel (rekenblad) / pdf (portable document format) / ps (postscript)

E.1.7 *Is alle informatie gratis te verkrijgen:* Soms heb je bepaalde informatie op een site waarvoor je moet betalen.

Maatstaf: Ja/Nee

Opmerking: Indien het niet meetbaar is, doordat je niet aan die gegevens kunt, zal ik een vraagteken zetten.

IV. Evaluatie van verschillende sites

In de onderstaande tabellen vind je een overzicht van de evaluaties van de verschillende onderzocht websites gegroepeerd per parameter categorie. De onderzochte websites zijn (opgedeeld in 2 groepen):

1. Statistische organisaties

. *Eurostat* (<http://europa.eu.int/comm/eurostat>): Deze site biedt economische en sociale statistische informatie aan van de Europese Gemeenschap.

. *Statistics Canada* (<http://www.statcan.ca>): Deze site verzamelt allerlei statistische gegevens over Canada.

. *Statec of Statistics Luxemburg* (<http://statec.gouvernement.lu>): De economische en sociale statistieken van Luxemburg

. *Statistics Finland* (<http://statfin.stat.fi/statweb/start.asp?LA=en&lp=home&DM=SLEN>): De statistische site van Finland.

. *Fedstat* (<http://www.fedstats.gov>): De statistische site van de Verenigde Staten.

2. Nationale banken

. *Bank of England* (<http://www.bankofengland.co.uk>) : De site van de Nationale Bank van Engeland.

. *Banca D' Italia* (<http://www.bancaditalia.it>): De site van de Italiaanse Nationale Bank.

. *De Nederlandse Nationale Bank* (<http://www.dnb.nl/dnb/homepage.jsp>): De site met statistische gegevens beheert door de Nationale Bank van Nederland.

(A.1) Layout:

Website	A.1.1	A.1.2	A.1.3	A.1.4
Eurostat	Ja	Ja	Ja	Ja
Statistics Canada	Ja	Ja	Ja	Ja
Statec	Nee	Nee	Ja	Nee
Statistics Finland	Ja	Nee	Ja	Ja
Fedstat	Ja	Ja	Ja	Ja
National Bank of England	Ja	Ja	Ja	Nee
Banca d'Italia	Ja	Ja	Nee	Nee
Nederlandse Nationale Bank	Ja	Nee	Nee	Nee

Conclusies tabel: De meeste websites hebben wel een min of meer uniformiteit in hun methodologische noten. Ze zijn niet altijd even overzichtelijk maar dit kan verbeterd worden door het opdelen van de noten in logische delen (hoofdstukken, onderwerpen,...). Dit wordt dan ook door de meeste organisaties gedaan. Indien je er dus voor wilt zorgen dat de gebruikers gemakkelijk de gezochte informatie kunnen terugvinden kan je best een onderverdeling maken. Het helpt om deze onderverdeling uniform (qua lay-out) te maken (zelfde titels enz...), maar dit is niet in alle gevallen mogelijk. Soms verschillen de tabellen/data te veel om de methodologische nota's op dezelfde manier in te delen.

(A.2) Gebruikersgericht:

Website	A.2.1	A.2.2	A.2.3
Eurostat	Nee	Ja	Nee
Statistics Canada	Nee	Nee	Nee
Statec	Nee	Nee	Nee
Statistics Finland	Nee	Ja	Nee
Fedstat	Ja	Ja	Ja
National Bank of England	Nee	Ja	Nee
Banca d'Italia	Nee	Nee	Nee
Nederlandse Nationale Bank	Nee	Nee	Nee

Conclusies tabel: Op de meeste sites wordt er geen onderscheid gemaakt tussen de verschillende types van gebruiker. Het aanbieden van de methodologische informatie gebeurt echter wel in de helft van de gevallen in verschillende stappen (eerst een korte inhoud om vervolgens gedetailleerder de methodologie te beschrijven). Het is in veel gevallen bij voorbeeld enkel nodig om snel even te zien hoe de informatie uitgedrukt en bekomen wordt zonder alle informatie over de nauwkeurigheid en manier van samplen te nemen. Aangezien er meestal toch geen onderscheid wordt gemaakt tussen de verschillende types gebruiker kan de indeling in stappen hier ook niet op toegespitst worden. De enige uitzondering op dit alles is Fedstat. Deze hebben een aparte pagina voorzien voor geïnteresseerde kinderen. Deze pagina is qua lay-out en inhoud aangepast voor de kinderen en staat vol met plezante statistieken en ideeën voor projecten. Naast dit gebruikerstype heb je dan de gewone gebruiker die op de rest van de website is gericht.

(B.1) Zoeken:

Website	B.1.1	B.1.2	B.1.3	B.1.4	B.1.5
Eurostat	Nee	Nee	Nee	NVT	NVT
Statistics Canada	Ja	Ja	Ja	Ja	NP
Statec	Nee	Nee	Nee	NVT	NVT
Statistics Finland	Ja	Ja	Nee	NVT	N
Fedstat	Ja	Ja	Nee	NVT	NP
National Bank of England	Nee	Nee	Nee	NVT	NVT
Banca d'Italia	Ja	Ja	Nee	NVT	NP
Nederlandse Nationale Bank	Nee	Nee	Nee	NVT	NVT

Conclusies tabel: Zeker niet op alle sites is het mogelijk om te zoeken naar methodologische nota's. Let wel op in vele gevallen is het wel mogelijk om op zoek te gaan naar statistische gegevens. Zo is het bijvoorbeeld op de site van de Nationale Bank van Engeland wel mogelijk om op zoek te gaan naar tabellen door gebruik te maken van een zoekoperatie maar het zoekresultaat zal altijd een lijst met tabellen zijn en dus nooit een bepaalde methodologische nota. Indien het toch mogelijk is om naar methodologische noten op zoek te gaan, gebeurt dit meestal door een full-text search. Of dit nu het werkelijk doorzoeken van heel de tekst is of het koppelen van sleutelwoorden aan een document en op basis van deze sleutelwoorden te zoeken valt uiteraard moeilijk na te gaan. Het zoeken door gebruik te maken van extra metadata wordt echter in minder gevallen ondersteund (zie B.1.3). Indien toch de mogelijkheid wordt gegeven om te zoeken via metadata kun je ook metadata gebruiken die betrekking hebben tot de inhoud van het document. De beste manier vind ik persoonlijk die van Statistics Canada. Hierbij kun je een woord opgeven, dan voor dat woord kiezen of het in de titel moet voorkomen, in de *body* van de tekst, in de *abstract*, in het onderwerp, in de lijst met sleutelwoorden, in de naam van de auteur,... Meestal zijn de zoekmachines wel nauwkeurig en performant. De performantie van de zoekmachines heb ik getest door verschillende queries uit voeren en te zien of het al dan niet lang duurt vooraleer je een resultaat krijgt. Om te zien of een zoekmachine nauwkeurig is, heb ik telkens voor een paar zoekopdrachten twee dingen nagegaan. Ten eerste heb ik enkele termen uit verschillende documenten genomen en die opgegeven als zoekstring om te zien of de documenten er terug uit komen. Verder ben ik dan ook gaan zien of er geen documenten tussen zitten die de zoekstring niet bevatten. Het enige probleem met de tweede manier van meten is dat je moet opletten met het gebruik van een thesaurus. Indien je woorden neemt die een synoniem kunnen hebben dat eventueel in de thesaurus zit kan het zijn dat je documenten terug krijgt die het synoniem van het woord bevatten zonder dat ze heel de zoekstring bevatten. Bijna alle zoekmachines waar je naar de methodologische noten kunt zoeken geven de documenten terug die de woorden bevatten die je opgegeven hebt. En de documenten die de zoekstring niet bevatten worden meestal ook niet terug gegeven. Een uitzondering is de statistische site van Finland. Hier heb je bij voorbeeld dat als je zoekt op de zoekstring *European Monetary Union*, dat je documenten terugkrijgt die deze zoekstring niet bevatten maar die wel EMU bevatten. Hierdoor vermoed ik dat ze op deze site gebruik maken van een thesaurus, een lijst van afkortingen, of een tool dat afkortingen detecteert.

(B.2) Navigeren

Website	B.2.1	B.2.2	B.2.3	B.2.4
Eurostat	Ja	Ja	Ja	Ja
Statistics Canada	Ja	Ja	Ja	Ja
Statec	Nee	Nee	Ja	Nee
Statistics Finland	Ja	Ja	Nee	Nee
Fedstat	Ja	Ja	Ja	Nee
National Bank of England	Ja	Ja	Nee	Nee
Banca d'Italia	Ja	Ja	Ja	Ja
Nederlandse Nationale Bank	Ja	Ja	Nee	Nee

Conclusies tabel: Bij de meeste pagina's kun je de methodologische documenten redelijk gemakkelijk terug vinden. Dit komt omdat de link van de data naar de methodologische

informatie op een duidelijke en uniforme manier gemaakt wordt. Bij sommige websites kun je tevens een overzicht krijgen van alle methodologische documenten in het systeem. Deze moeten echter best opgedeeld worden in logische eenheden (hoofdstukken), anders vind je niet gemakkelijk de gezochte methodologische nota terug. Bij de meeste moet je al voldoende kennis bezitten van de statistische data vooraleer je de juiste methodologische data kunt terug vinden via zo een overzichtstabel.

(C.1) Revisies

Website	C.1.1	C.1.2	C.1.3	C.1.4
Eurostat	Ja	Ja	Nee	Ja
Statistics Canada	Ja	Ja	Ja	Soms
Statec	Ja	Nee	Nee	Nee
Statistics Finland	Ja	Nee	Nee	Ja
Fedstat	Ja	Ja	Ja	Ja
National Bank of England	Ja	Ja	Ja	Ja
Banca d'Italia	Ja	Ja	Ja	Soms
Nederlandse Nationale Bank	Ja	Ja	Ja	Ja

Conclusies tabel: In alle gevallen worden de methodologische noten up-to-date gehouden. Ze worden niet altijd als aparte versies bijgehouden. Dit zorgt ervoor dat je oudere versies niet altijd kunt raadplegen. Meestal wordt enkel de laatste versie getoond, maar dan wordt er meestal wel vermeld dat er aanpassingen zijn gebeurd en wanneer deze zijn gebeurd (in het document zelf). Dit gebeurt ofwel door de aanpassingen op te sommen ofwel door de laatste update datum van het document er bij te zetten. De link tussen oudere gegevens en hun bijhorende methodologische gegevens worden dus niet in alle gevallen bijgehouden (enkel de meest recente versie wordt getoond).

(D.1) Gebruik van een standaard

Website	D.1.1	D.1.2	D.1.3	D.1.4
Eurostat	Ja	Ja	Ja	Ja
Statistics Canada	Ja	Nee	Nee	Ja
Statec	Ja	Ja	Ja	Nee
Statistics Finland	Ja	Ja	Ja	Ja
Fedstat	Ja	Ja	Ja	Ja
National Bank of England	Ja	Ja	Ja	Ja
Banca d'Italia	Ja	Ja	Ja	Nee
Nederlandse Nationale Bank	Ja	Nee	Nee	Nee

Conclusies tabel: Alle websites gebruiken en vermelden de standaard die ze toepassen bij het berekenen en opstellen van hun tabellen. In de meeste gevallen worden de verschillen en aanpassingen (met betrekking tot de gebruikte standaard) beschreven met de bijhorende uitleg van waarom ze het anders doen. Niet in alle methodologische documenten worden vergelijkingen gemaakt met andere standaarden (uit andere landen, toegepast op een ander domein of oudere standaarden).

(D.2) Inhoud

Website	D.2.1	D.2.2
Eurostat	N/S/M	Ja
Statistics Canada	N/S/M/O	Ja
Statec	N/M	Ja
Statistics Finland	S/M	Ja
Fedstat	N/S/M/O	Ja
National Bank of England	-	Ja
Banca d'Italia	N/S	Ja
Nederlandse Nationale Bank	-	Ja

Conclusies tabel: Alle websites vermelden de bron van de statistische gegevens in hun methodologische gegevens. Maar zeker niet alle websites vermelden de nauwkeurigheid, de sample methode, de meetfouten, en wat er gebeurt bij ontbrekende waarden. Wat er in vele gevallen wel wordt gedaan, is verwijzen naar de gebruikte standaard voor meer informatie. In deze standaarden zijn deze gegevens dan wel terug te vinden.

(E.1) Extra

Website	E.1.1	E.1.2	E.1.3	E.1.4	E.1.5	E.1.6	E.1.7
Eurostat	Dom	Nee	NVT	E,F,D	NVT	Html	Ja
Statistics Canada	Tab	Ja	NVT	E,F	NVT	Html	Nee
Statec	Dom	Nee	NVT	F	Ja	Html/Pdf	Ja
Statistics Finland	Tab	Ja	Ja	E,Fin	Ja	Html/Txt/Excel/ps	Ja
Fedstat	Tab	Ja	NVT	E	NVT	Html	Ja
National Bank of England	Dom	Ja	Nee	E	Nee	Html	Ja
Banca d'Italia	Tab	Ja	NVT	E, It	Ja	Pdf	Ja
Nederlandse Nationale Bank	Tab	Ja	NVT	E,N	Ja	Excel	Ja

Conclusies tabel: Op geen enkele website worden er methodologische noten aan tijdreeksen gekoppeld. Zij zijn ofwel aan het domein gekoppeld (een algemene methodologische nota), ofwel aan een tabel. Over het algemeen zijn de websites redelijk volledig. Dit wil zeggen dat als je een methodologische nota per domein hebt, dat elk domein ook daadwerkelijk een methodologisch nota heeft. Als de methodologische noten aan de tabellen worden gekoppeld, heb je meestal wel bij elke tabel een methodologische nota. Soms gaat het wel alleen over een verwijzing naar de gebruikte methodologische standaard maar dit is nog altijd beter dan helemaal niets. In de meeste gevallen kun je niet zelf je tabellen samenstellen, dus is het derde criteria niet van de orde. De websites zijn meestal opgesteld in het Engels en de taal van herkomst. Als je documenten echt kunt downloaden wordt er meestal voor gezorgd dat de methodologische informatie mee wordt gedownload. Het meest ondersteunde documenttype zijn gewone html- files. Toch is denk ik een documenttype zoals pdf aan te raden aangezien je deze gemakkelijker kunt opslaan op je harde schijf en je die ook kunt openen met een veel gebruikt (en bovendien gratis te verkrijgen) programma.

V. Evaluatie van de site van de Nationale Bank van België:

Nu ga ik de site van de Nationale Bank van België onderwerpen aan de evaluatieparameters. In de eerste plaats geef ik de bovenstaande tabellen maar dan ingevuld voor de website van de Nationale bank van België.

(A.1) Layout:

Website	A.1.1	A.1.2	A.1.3	A.1.4
Nationale Bank van België	Nee	Ja	Nee	Nee

De methodologische noten van de Nationale bank van België zijn zeker niet gestandaardiseerd qua lay-out. De ene keer zijn het overzichtspagina's in html waar dan links op staan naar de verschillende delen van de methodologische nota's, de andere keer is het één groot pdf-document met de methodologische nota in. Soms staat er effectief een link naar een methodologische nota en andere keren zijn het referenties naar publicaties waar een woordje uitleg over de gevolgde methodologie te vinden is. De methodologische noten zijn ook niet opgedeeld in standaardblokken. Sommige documenten lijken op dezelfde manier onderverdeeld te zijn in:

- Inhoud
- Basisgegevens
- Verwerking
- Referenties

Maar andere methodologische documenten zijn dan weer helemaal niet opgedeeld volgens deze onderdelen.

(A.2) Gebruikersgericht:

Website	A.2.1	A.2.2	A.2.3
Nationale Bank van België	Nee	Soms	Nee

Er wordt nergens op de site onderscheid gemaakt tussen de verschillende types gebruiker. Aangezien dit niet gedaan wordt, kan de hoeveelheid aangeboden methodologische informatie en het type methodologische informatie niet aangepast worden aan het soort gebruiker. Soms wordt de informatie echter in verschillende stappen aangeboden. Dit is zo bij voorbeeld bij de “niet financiële rekeningen”. Hier wordt er begonnen met kort iets te zeggen over de gebruikte standaard. Dan kun je links volgen naar publicaties met een summier weergave van de gebruikte methodologieën. Voor het bruto binnenlands product en het bruto nationaal inkomen kun je zelf de gedetailleerde informatie bekomen. Er zijn dus verschillende niveaus van aanbieden maar de onderlinge relaties, verhoudingen zijn niet zo duidelijk of worden alleszins niet zo duidelijk aangebracht. Het zou beter zijn dat je begint met het aanbieden van een korte uitleg over de gebruikte standaard en dat je dan 1 link moet volgen indien je iets meer uitleg wilt over de verschillende delen van de rubriek “niet-financiële rekeningen” om daar een link te hebben naar een gedetailleerde uitleg over het bruto binnenlands product en het bruto nationaal inkomen.

(B.1) Zoeken:

Website	B.1.1	B.1.2	B.1.3	B.1.4
Nationale Bank van België	Ja	Ja	Nee	P

Je kan zoeken naar methodologische noten op de site van de Nationale Bank van België. Je kunt echter niet aangeven dat je enkel de methodologische noten wilt doorzoeken. Het beste dat je kan doen is aangeven dat je binnen het domein statistieken wilt zoeken. Het is een full tekst search. Je kan niet zoeken op meta-informatie zoals de titel, auteur of soort informatie (zoals hierboven reeds vermeld, zou je kunnen aangeven dat je op zoek bent naar een methodologische nota). Het enige van meta-informatie dat je wel kunt gebruiken bij het zoeken is het domein.

(B.2) Navigeren

Website	B.2.1	B.2.2	B.2.3	B.2.4
Nationale Bank van België	Ja	Nee	Nee	Niet van toepassing

Je moet meestal niet lang zoeken naar de methodologische documenten en zeker niet op de Belgostat website. Hier heb je in het overzicht van de tabellen een link naar hun methodologische informatie. Op de site van de nationale bank is er geen uniforme link tussen de statistische data en de methodologische documenten. Er is geen overzicht van alle methodologische documenten (dit in tegenstelling tot de CD-rom waar je dit wel hebt) waarin je gemakkelijk zou kunnen browsen naar de gewenste methodologische informatie. Op de Belgostat website vind ik dit niet per se nodig aangezien je in het overzicht van de statistische tabellen per domein een goede link hebt naar de methodologische documenten.

(C.1) Revisies

Website	C.1.1	C.1.2	C.1.3	C.1.4
Nationale Bank van België	Ja	Ja	Nee	Nee

Er is niet veel informatie terug te vinden over het aanpassen en up-to-date houden van de methodologische noten. In de methodologische noten zelf staan er af en toe verwijzingen naar oudere versies maar je moet dus al in de documenten zelf op zoek gaan naar de informatie over revisies. De gebruiker wordt dus zeker niet genoeg op de hoogte gebracht van het feit dat er een nieuwe versie is.

(D.1) Gebruik van een standaard

Website	D.1.1	D.1.2	D.1.3	D.1.4
Nationale Bank van België	Ja	Ja	Ja	Soms

Er wordt telkens vermeld welke standaard er gebruikt wordt. Eventuele afwijkingen worden soms vermeld met de reden waarom ze het zo doen. Vergelijkingen met andere standaarden van andere landen of andere domeinen worden niet gemaakt. Een vergelijking met oudere standaarden wordt soms wel gemaakt.

(D.2) Inhoud

Website	D.2.1	D.2.2
Nationale Bank van België	Soms het één soms het andere	Ja

Aangezien de methodologische documenten van de Nationale Bank van België afkomstig zijn van verschillende bronnen en opgesteld door verschillende mensen is de inhoud zeker niet uniform. In sommige documenten wordt er iets gezegd over de manier waarop het verzamelen van de gegevens gebeurt in andere wordt er weer niet op in gegaan maar wordt er gewoon verwezen naar de standaard die ze gebruiken.

(E.1) Extra

Website	E.1.1	E.1.2	E.1.3	E.1.4	E.1.5	E.1.6	E.1.7
Nationale Bank van België	Dom	Nee	Nee	N,F,E	Nee	Html, pdf	Ja

De methodologische documenten worden gekoppeld aan een domein en niet aan een tabel. Toch hebben zeker niet alle domeinen een methodologisch document. De gebruiker kan wel tabellen samenstellen maar aangezien je niet 1 methodologisch document per tabel maar per domein hebt wordt deze niet aangepast aan de nieuw gevormde tabel. In dit geval krijg je niet de mogelijkheid om methodologische informatie van die tabel op te vragen dus je krijg allezinds geen verkeerde informatie. Stel echter dat je per tabel een methodologische nota zou hebben met de informatie in die enkel relevant is voor die tabel en je zou dan van twee tabellen enkele variabelen combineren tot een nieuwe tabel zou het handig zijn om ook bij deze nieuwe tabel een methodologisch document te hebben. Ofwel zijn dit gewoon verwijzingen naar de twee methodologische documenten die bij de twee tabellen staan ofwel ga je de informatie die betrekking heeft tot de geselecteerde variabelen eruit filteren en samen voegen in 1 nieuw methodologisch document. De tweede manier zou uiteraard de handigste manier zijn maar ze is echter niet altijd realiseerbaar. Dit hangt af van hoe de documenten zijn opgedeeld en dan nog moet het allemaal nog technisch realiseerbaar zijn. De ondersteunde talen zijn Nederlands, Frans en Engels alhoewel de site ook in het Duits te bezoeken valt. Het gedeelte met de statistieken is echter niet in het Duits te bezichtigen. Je kunt bepaalde tabellen downloaden maar de methodologische informatie gaat niet mee. De meeste methodologische documenten zijn pdfs maar een deel (de inleiding vooral) is ook in html geschreven. Alle informatie is bij mijn weten gratis te verkrijgen.

VI. Conclusies en aanbevelingen:

In de eerste plaats moet er meer uniformiteit komen onder de methodologische documenten. Er is nu al een kleine opsplitsing van sommige documenten in logische onderdelen. Zo zijn verschillende methodologische noten onderverdeeld in: inhoud, basisgegevens, verwerking, referenties. In de eerste plaats moet gezien worden of alle documenten een bepaalde onderverdeling kunnen ondergaan. Hiervoor moeten we met mensen spreken die een groot deel van de inhoud van de methodologische nota's bepalen. We kunnen ook vertrekken van goede indelingen die gebruikt worden door bij voorbeeld de Eurostat website. De onderverdeling die Eurostat gebruikt is de volgende:

- Analytical framework, concepts, definitions, and classifications
- Scope of the data

- Accounting conventions
- Nature of the basic data
- Compilation practices
- Other aspects

Alle onderdelen zijn in elke methodologische nota ingevuld. Een tweede goed voorbeeld dat we kunnen gebruiken is Statistics Canada. Zij delen hun methodologische informatie ook op in verschillende blokken maar niet elke blok komt overal terug. Zo zijn er bepaalde blokken die bij een bepaald soort statistische informatie hoort. Maar toch is een groot deel van de onderverdeling overlappend omdat deze ingevuld kunnen worden voor alle statistische data. Enkele gebruikte onderverdelingen:

- Type of survey
- Conceptual universe and target population
- Estimation
- Quality evaluation
- Disclosure control
- Data accuracy (komt bijvoorbeeld ongeveer overal voor)
- Time series

Het voordeel van het opdelen van de methodologische documenten is dat je gemakkelijker informatie terug vindt en dat het geheel overzichtelijker, duidelijker en properder wordt. Een tweede belangrijk voordeel is dat je gemakkelijker gebruikte methodologieën kunt vergelijken doordat je de overeenkomstige blokken moet vergelijken. Als je enkele methodologieën wilt vergelijken op gebied van nauwkeurigheid moet je enkel de blokken 'data accuracy' gaan lezen en vergelijken en moet je niet telkens heel het document gaan lezen om te zien wat de nauwkeurigheid van de verschillende methodologieën zijn.

Het grootste probleem dat zal optreden bij het uniform maken van de methodologieën en het opdelen van de methodologieën is dat deze van verschillende bronnen afkomstig zijn. En deze verschillende bronnen hebben telkens een andere lay-out en een andere inhoud (of structuur van inhoud). Indien het niet mogelijk is om de oorspronkelijke methodologische noten in een andere (de eventuele nieuwe) lay-out te gieten moet er door middel van het automatisch extraheren van de verschillende informatieblokken de opdeling gemaakt worden. Het ideale zou natuurlijk zijn dat je in de toekomst de mensen die de methodologische nota's opstellen een soort template laten invullen die al opgedeeld is in de logische blokken.

Er wordt momenteel geen onderscheid gemaakt op de site van de Nationale Bank van België tussen de verschillende types gebruiker. De bedoeling is dat dit wel zou meespelen bij de disseminatie van de statistische gegevens en vooral de methodologische noten. Een uitgebreide studie van welke gebruikerstypes er zijn en welke noden ze hebben (bij voorbeeld op gebied van zoeken) volgt later in het project. Een eerste opsplitsing van verschillende gebruikers zou kunnen zijn:

1. academici: studenten, docenten;
2. ambtenaren van openbare en andere nationale en internationale instellingen;
3. ambtenaren van de Nationale Bank;
4. journalisten;
5. gewone burger.

Uiteraard kan het zijn dat bepaalde groepen grotendeels de zelfde informatie noden hebben en dus niet apart behandeld moeten worden.

Ongeacht van hoe je het aanbieden van de informatie afstemt op de gebruikerstypen zou ik sowieso het aanbieden van methodologische informatie in verschillende stappen laten gebeuren. Weer baserend op de manier dat Eurostat het doet zou ik bij voorbeeld bij de tabellen zelf (onderaan bijvoorbeeld) al een korte beschrijving geven van de tabellen en de gebruikte methodologie. Dit kan bij voorbeeld gewoon het aangeven zijn van welke standaard er gevolgd wordt. Een andere plaats waar het misschien goed zou staan is bij de metagegevens. Op de Belgostat website kan je bij elke tabel de metagegevens raadplegen. Hier zou je ook een verwijzing kunnen geven naar de gebruikte methodologieën. Op de pagina die te voorschijn komt als je op de knop *metagegevens* drukt zou je onderaan een kort uitleg kunnen zetten over de gebruikte methodologie. Bij deze korte beschrijving zou dan een link kunnen gevolgd worden *more details* waarbij je het huidige methodologisch document te zien krijgt.

Er zijn op het eerste zicht twee mogelijkheden bij het aanbieden van een korte beschrijving. Ofwel, en dit zou veel makkelijker verlopen als je de methodologische noten al in logische blokken verdeeld hebt, bied je enkel de belangrijkste blokken aan. Bijvoorbeeld

- Standaard:
- Sample methode:
- Nauwkeurigheid:

En laat je de minder belangrijke onderwerpen weg. In deze aanpak moet er onderzocht worden wat in de eerste plaats de gebruiker belangrijk vindt en in de tweede plaats wat de Nationale Bank zelf belangrijk vindt. Zo kan het bij voorbeeld zijn dat de gebruiker niet echt geïnteresseerd is in de voorwaarde waarop je de statistische informatie mag gebruiken terwijl dit voor de Nationale Bank wel belangrijk kan zijn. Je kunt niet enkel het geen je laat zien afhangen van het gebruikertype en de wensen van de Nationale Bank maar ook de volgorde en de manier van tonen. Zo kan blijken dat de gebruikersgroep *journalisten* zich weinig aantrekt van de voorwaarde waarop de statistische informatie getoond mag worden (terwijl ze dit wel zou moeten). Je kan er voor zorgen dat ze deze informatie niet missen door ze bovenaan de lijst met logische blokken te plaatsen en in een andere kleur, in het vet,...

Een andere mogelijkheid is een beschrijvende tekst opstellen die gaat samenvatten wat er in het methodologische document staat. De vraag is uiteraard hoe je de twee manieren gaat verwezenlijken. Ofwel moet de maker van het methodologisch document zelf de opsplitsing maken van het document in logische blokken of moet hij zelf een korte beschrijving opstellen van de nota, ofwel moet je automatisch de data extraheren uit de nota's. Dit kan je doen door verschillende text mining technieken te gebruiken (text summarization,...).

Eén van de belangrijke zaken die de disseminatie van statistische data en hun bijhorende methodologische noten efficiënter maakt is het kunnen zoeken via een zoekmachine naar bepaalde methodologische informatie. Aangezien dit één van de hoofdzaken is van het project en er een uitgebreidere studie komt van wat de mogelijkheden zijn die je kunt voorzien in je zoekmachine ga ik er hier niet heel gedetailleerd op ingaan. Wat ik wel ga opsommen zijn de mogelijkheden die je kunt voorzien. Er zijn eigenlijk twee grote manieren waarop je kan zoeken en deze kunnen eventueel gecombineerd worden. Ofwel zoek je op woorden uit de tekst (full-text search), ofwel ga je zoeken aan de hand van metadata toegekend aan de metadata.

Momenteel is er op de site van de Nationale Bank van België een algemene zoekmachine voorzien. Hier wordt gewerkt met full text search en het is niet mogelijk om te zoeken op metadata (buiten het domein). Het is ook niet mogelijk om specifiek naar methodologische informatie te zoeken. Er zou misschien bij de zoekmachine een optie moeten staan van naar wat je juist zoekt (publicaties, statistische informatie, methodologische noten,...). Eens een methodologische nota gevonden wordt, moet ook duidelijk gemaakt bij welke data ze juist behoort (want in het overzicht van de gevonden documenten staan gewoon de gegevens van het document zelf). Als in het document zelf duidelijk staat bij welke statistische informatie ze behoort, is er geen probleem natuurlijk, maar momenteel is dit niet het geval.

In zowel de full-text search optie als de metadata optie moet er een indexering gebeuren. In geval van full-text search worden alle nuttige woorden (dus geen woorden zoals 'een', 'de', 'het',...) verbonden met het bijhorende document. Meestal worden nog andere text mining technieken gebruikt om de woorden te transformeren naar een bruikbare vorm. Zo kun je bijvoorbeeld stemming gebruiken om de woorden naar hun basisvorm om te zetten, er kunnen formules gebruikt worden die bepaalde gewichten (geven de belangrijkheid van het woord weer) aan woorden toekennen,... Bij het werken met metadata ga je niet alle woorden uit de tekst koppelen met het document maar ga je bepaalde meta-informatie toekennen aan het document. Dit is belangrijke informatie zoals de titel van het document, de auteur, de gebruikte standaard,... Ook dit kan door gebruik te maken van text mining.

Bij het zoeken zelf zijn er bepaalde dingen die je in het oog moet houden. Ten eerste moet de zoekmachine gebruiksvriendelijk zijn. Niet iedere persoon die geïnteresseerd is in statistische informatie is een specialist in het zoeken van bepaalde informatie door gebruik te maken van zoekmachines. De gebruiksvriendelijkheid kan verbeterd worden door het aanbieden van de gevonden resultaten goed te structureren. Zo is het niet altijd genoeg om een lijst te geven met links naar de gevonden documenten maar indien mogelijk moet je ook informatie als de titel van het document, het domein waarbinnen het zich bevindt enz. toevoegen. Voor de gebruikers die wel getraind zijn in het zoeken naar informatie moet je dan ook alle functies aanbieden die hij wenst te gebruiken (zoals het gebruik van de termen 'and', 'or',...).

Een tweede belangrijke eigenschap van een goede zoekmachine is de performantie. Een gebruiker wil geen volle minuut moeten wachten voor hij resultaten te zien krijgt. Een derde belangrijk iets om in het oog te houden is de nauwkeurigheid van de zoekmachine. Je wil natuurlijk enkel de relevante documenten terugkrijgen zodat je niet nog eens een uur kwijt bent met het doorbladeren van de resultaten. De nauwkeurigheid wordt meestal ook sterk verbeterd indien je bijvoorbeeld het toekennen van metadata laat gebeuren door mensen (en bij voorkeur de auteur van het document) en niet automatisch laat gebeuren. Anderzijds is niet iedereen bereid om zijn methodologische documenten te annoteren, en is het meestal voordeliger om het door de computer zelf te laten zijn (en de resultaten zijn toch behoorlijk).

Bij het zoeken kan ook gebruik gemaakt worden van de thesaurus om zoekoperaties te verbeteren. Als je zoekt op een bepaald woord zouden ook alle synoniemen van dat woord kunnen opgezocht worden. De Nationale Bank van België maakt al gebruik van een thesaurus voor het zoeken naar statistische data en publicaties en deze zou dus ook geïntegreerd kunnen worden in het zoeken naar methodologische noten.

Het navigeren naar een bepaald methodologisch document kan ook veel beter. Wat ik ten eerste al aanbeveel is om bij de statistische gegevens (de tabellen) ergens een link te zetten naar de methodologische documenten. Dit verduidelijkt de relatie tussen de statistische

gegevens en de methodologische nota. Nu heb je enkel een link naar de nota's in de boomstructuur van de tabellen. De link moet ook uniformer worden. Het enige probleem is dat de statistische gegevens al niet uniform worden aangebracht aan de gebruiker. Zo kun je gegevens bekijken via de Belgostat website, anderzijds kun je de gegevens via de site van de Nationale Bank raadplegen. Binnen de Belgostat website gebeurt het aanbieden van de statistische informatie wel op een uniforme manier maar zoals reeds gezegd staat de link naar de methodologische informatie op de verkeerde plaats (nl. in de boomstructuur).

Wat je wel terug vindt op de CD-rom maar niet op de website is een overzicht van de methodologische documenten geordend in hoofdstukken. Op de meeste goede statistische websites vind je dit terug (Eurostat, Statistics Canada,...). Zo een overzicht kan twee doelen hebben. Het kan gebruikt worden om gemakkelijk informatie terug te vinden, ondanks ik denk dat dit niet de efficiëntste manier is om te zoeken naar informatie. Als je een goede zoekmachine hebt of als de link structuur tussen de methodologische noten en de bijhorende statistische informatie duidelijk is, is dit veel efficiënter. Indien dit niet zo is, kan het inderdaad wel nuttig zijn om een methodologische manier via een soort boomstructuur terug te vinden. Een tweede reden is, dat sommige gebruikers echt geïnteresseerd zijn in de verschillende methodologieën en dan hebben ze via dit overzicht een goede manier om al deze noten te doorbladeren en te onderzoeken. Dit is uiteraard zeer afhankelijk van de gebruikerstypes (niet alle mensen zijn geïnteresseerd in de noten).

Er zijn verschillende manieren om zo een overzicht te ordenen. De eerste manier is de ordening van de statistische tabellen te volgen. In bijna alle gevallen zijn de statistische tabellen ook geordend volgens één of andere ordening (bijvoorbeeld volgens domein, subdomein,...). Als je in zo een overzicht dezelfde ordening gebruikt, is de connectie tussen de statistische tabellen en de nota's duidelijker. Je moet natuurlijk niet altijd deze ordening gebruiken, je kunt bijvoorbeeld de documenten ordenen op tijdstip,... Een laatste mogelijkheid (waar nog een diepgaander onderzoek naar gedaan moet worden) is door gebruik te maken van text mining technieken zoals clustering een hiërarchie te genereren. Hier kun je misschien ook weer de verschillende noden van de verschillende gebruikertypes in laten meespelen.

De organisatie van de revisies van de methodologische noten kan ook op een efficiëntere manier. Ten eerste is het al niet gemakkelijk om te weten waar je oudere gegevens kunt raadplegen. Op de Belgostat website bijvoorbeeld zijn er geen oudere gegevens te vinden. Voor de methodologische documenten moet je de tekst al lezen voordat je eventuele verwijzingen naar andere versies van een methodologisch document kunt terug vinden. Het zou misschien beter zijn om duidelijk te maken dat er andere versies zijn door een link naar deze oudere versie te voorzien bij het huidige methodologisch document. Er zou ook duidelijk vermeld moeten worden wanneer het huidige methodologisch document opgesteld is geweest. Indien oudere data (die opgemeten zijn met een oudere versie) gemixt worden met nieuwere data, moet de gebruiker gewaarschuwd worden dat een deel van de data is berekend met een oudere methodologie. Je zou de datum van inwerkingtreding van de methodologie bij voorbeeld als metadata kunnen bijhouden. Als je bijvoorbeeld een overzichtstabel hebt die een bepaald statistisch gegeven elk jaar berekent, en door de jaren heen is de gebruikte methode voor dat statistisch gegeven veranderd, kan het zijn dat je verschillende methodologische noten moet koppelen aan 1 tabel. Een andere manier is de oudere methodologische gegevens ook in de huidige methodologische nota opnemen. Indien je dit niet doet kan bepaalde statistische informatie verkeerd geïnterpreteerd, beoordeeld of vergeleken worden.

Het vermelden van op welke standaard de methodologie gebaseerd is, gebeurt consequent en op een goede manier. Het vermelden van de bron van het methodologisch document gebeurt eveneens op een goede manier. Er moet alleen zeker voor gezorgd worden dat indien je met verschillende stappen werkt en je eerst een korte beschrijving geeft, je zeker de standaard vermeldt aangezien dat dat toch één van de belangrijke zaken zijn. De gebruikte standaard zou eventueel ook als metadata kunnen bijvoegen (zodat er gemakkelijk op gezocht kan worden).

Het is zeker belangrijk om iets te vertellen over nauwkeurigheid, de manier van samplen en iets over de meetfouten die kunnen optreden. Eventueel indien mogelijk, wat er gebeurt met ontbrekende data. Aangezien ik toch aanraad om de methodologische documenten op te splitsen in logische blokken, zou ik zeker drie blokken toevoegen: nauwkeurigheid, samplen en meetfouten.

Momenteel kun je de methodologische documenten in drie talen lezen waaronder het Engels. Dit is naar mijn weten zeker voldoende en het is misschien niet nodig (mogelijk) om het ook nog in het Duits te vertalen. Nu heb je een methodologisch document per domein maar het zou misschien beter zijn om een methodologische nota aan een tabel te koppelen. De methodologische informatie is ook redelijk onvolledig aangezien de helft van de tabellen/domeinen zelf geen methodologisch document hebben. Dit zou toch moeten verbeteren. Desnoods geef je enkel een verwijzing naar de standaard die gebruikt is voor het opstellen van de tabellen. Dit is uiteraard niet de ideale oplossing maar als het niet realiseerbaar is om elke tabel te voorzien van een methodologisch document doordat het personeel dit niet wilt of wat de reden ook moge zijn, is dit een tussenoplossing.

De gebruiker kan via de Belgostat site zelf tabellen samen stellen door het kiezen van de juiste parameters. Aangezien de methodologische documenten gekoppeld worden aan een domein i.p.v. aan de tabellen is het momenteel onmogelijk om de methodologische noten mee aan te passen als de tabellen aangepast worden (door ze kruisgewijs te combineren). En zelf indien het aan tabellen gekoppeld zou zijn zal het moeilijk zijn om de juiste methodologische informatie samen te bundelen voor een zelf gecreëerde tabel. Wat we moeten kunnen doen is de methodologische informatie die hoort bij een specifieke parameter extraheren uit de methodologische nota van de tabel waarin de parameter zich bevindt. Dit zou eventueel kunnen gedaan worden door het manueel annoteren van de methodologische noten (aangeven van welke stukken methodologische informatie bij welke parameters horen). Anderzijds zouden misschien ook text mining technieken kunnen gebruikt worden om dit volledig of gedeeltelijk automatisch te doen.

Je kunt statistische tabellen downloaden door ze om te zetten naar excel maar de methodologische noten worden niet automatisch mee gedownload. Indien je de omzetting doet, zou er toch ergens de mogelijkheid aangeboden moeten worden om de methodologische noten mee te downloaden. Op de site van de Nationale Bank van Nederland wordt er ook gewerkt met Excel pagina's en hier voegen ze de methodologische gegevens gewoon toe in een andere tab van het Excel blad.

Er worden twee belangrijke types ondersteund: pdf en html. Het is echter niet zo dat je bij elk document kunt kiezen welk type je wilt. Sommige documenten staan in html en andere dan weer in pdf. Je zou bijvoorbeeld alle documenten in html kunnen laten aanbieden met eventueel een link naar de pdf voor de mensen die dit wensen. Dan heb je ineens ook het probleem opgelost van te grote pdf documenten (dit duurt soms heel lang voor je ze te zien krijgt).

VII. Samenvatting en besluit

Dit gedeelte beschrijft de belangrijkste conclusies. Ik ben begonnen met het beschrijven van de twee belangrijkste standaarden die op dit moment bestaan met betrekking tot statistische gegevens en vooral met betrekking tot de methodologische gegevens. Het voornaamste besluit bij de analyse van de twee standaarden (de standaard van het IMF en van Eurostat), is niet te opteren voor het volgen van één standaard maar zo goed als mogelijk (het moet praktisch haalbaar blijven) de twee standaarden te volgen. De reden is eenvoudigweg dat beide standaarden zich op een ander deel van de methodologische noten concentreren. De IMF standaard (*SDDS*) richt zich vooral op de inhoud van de methodologische noten en in hoeverre ze overeen stemt met de gebruikte standaard. De standaard opgesteld door Eurostat daarentegen richt zich meer op de gebruiker en zijn noden. In hoeverre is de aangeboden informatie relevant voor de gebruiker, is ze duidelijk, toegankelijk, nauwkeurig, volledig enz. Daarom heb ik mij bij het opstellen van mijn lijst met evaluatie criteria gebaseerd op beide standaarden.

Eens de lijst met evaluatieparameters was opgesteld zijn verschillende internationale statistische instelling geëvalueerd. Deze evaluatie is samengevat in een matrix waarin voor elk onderdeel van de evaluatie een cijfer (op 10) wordt gegeven. Dit cijfer (telkens op tien) zal min of meer een gewogen gemiddelde zijn van alle parameters binnen het specifieke gedeelte. Dit wil zeggen dat voor elke afzonderlijke evaluatieparameter een cijfer wordt gegeven, de uiteindelijke score voor dat gedeelte zal een gemiddelde zijn rekening houdend met de belangrijkheid van de evaluatieparameter (bijvoorbeeld de parameter *is er een overzicht van de methodologische noten* is belangrijker dan de parameter *is het overzicht duidelijk gestructureerd*, aangezien het in de eerste plaats belangrijk is om zo een overzicht te hebben en pas dan verwachten we dat dit nog eens duidelijk en gestructureerd is). In deze matrix zul je gemakkelijk de verschillende internationale instellingen kunnen vergelijken (een cijfer is een duidelijke indicator). Hierdoor kun je gemakkelijk zien voor welke delen welke sites een goed voorbeeld zijn.

Instelling	Lay-out	Gebruiker gericht	Zoeken	Navigeren	Revisies	Standaard	Inhoud	Extra
Eurostat	9	4.5	0	9	7	9	8	5
Statistics Canada	9	2	9	9	8	6	9	7
Statec	3	0	0	2.5	3	7	6.5	5.5
Statistics Finland	7	4.5	5.5	6.5	5.5	9	6.5	9
Fedstat	8	9	6.5	8	9	9	9	7
Nationale Bank of England	8	4.5	0	6.5	9	9	3.5	6.5
Banca d'Italia	5.5	0	6.5	8.5	8	7	6.5	8
Nederlandse Nationale Bank	4	0	0	6.5	9	4	3.5	8
Nationale Bank van België	4	2	5.5	3.5	6.5	8	6.5	5.5

Om de berekeningen te verduidelijken zal ik een voorbeeld geven. Als we gaan zien hoe het cijfer voor de lay-out van Statistics Finland is berekend gaan we eerst kijken naar de individuele score voor elke evaluatieparameter. Voor de lay-out zijn er 4 evaluatieparameters (uniformiteit in de layout, duidelijkheid van de lay-out, is er een opdeling in logische blokken en is deze opdeling uniform). Deze hebben respectievelijk de gewichten 2, 2, 2 en 1

meegekregen (de laatste is dus iets minder belangrijk). Dan ben ik gaan zien in de tabel van de analyse wat elke evaluatieparameter zegt voor Statistics Finland. We hebben:

Uniformiteit = ja, dus krijgt dit een 2/2,
Duidelijkheid = nee, dus krijgt dit een 0/2,
Opdeling in logische blokken = ja, dus krijgt dit een 2/2,
Uniformiteit van de opdeling = ja dus krijgt dit een 1/1.

Samen geeft dit dus een 5 op 7 wat afgerond neerkomt op 7/10.

In deze paragraaf wordt er een opsomming van de belangrijkste punten die aangepast moeten worden of waarmee rekening gehouden moet worden bij het opstellen van een goed model voor de methodologische noten op een website voor een statistische instelling.

- a. De uniformiteit, zowel op gebied van lay-out als op gebied van indeling van de methodologische noten, is zeer belangrijk. Dit zorgt ervoor dat de gebruiker beter vertrouwd geraakt met de methodologische noten en dus sneller de gewenste methodologische noten of bepaalde informatie binnen de gewenste methodologische noten zal terug vinden.
- b. Het aanbieden van de informatie moet goed afgestemd worden op de verschillende gebruikerstypen. In vele gevallen wordt er geen onderscheid gemaakt tussen verschillende gebruikerstypes terwijl dit toch belangrijk is. Elk type heeft andere noden en de informatie die aangeboden wordt aan een bepaalde gebruiker samen met de manier waarop deze informatie wordt aangeboden moet afgestemd worden op het type.
- c. De zoekmachine moet aangepast worden om specifiek te kunnen zoeken naar de methodologische noten. Aangezien methodologische documenten over het algemeen volledig verschillen van de documenten met statistische data moet er ook een specifieke ondersteuning komen voor het zoeken naar methodologische informatie. Dit kan gebeuren door full-text search te implementeren, de methodologische informatie te voorzien van de nodig metadata waarop dan gezocht kan worden of beide.
- d. Niet alleen de lay-out en indeling van de documenten moeten uniform gebeuren maar tevens het aanbieden van de methodologische informatie. Hiermee bedoel ik dan vooral het uniform maken van de link tussen de statistische gegevens en de methodologische noten. Als een gebruiker bepaalde statistische informatie aan het bezien is moet hij op een eenvoudige en uniforme manier naar de bijhorende methodologische noten geraken. Dit zorgt er weer voor dat de gebruiker beter vertrouwd geraakt met de site en dus sneller de gewenste informatie zal terugvinden.
- e. Er moet goed omgesprongen worden met verschillende versies van een methodologisch document. Je moet ten eerste duidelijk de hiërarchie terugvinden tussen de methodologische documenten (welk zijn de oudere versies, welk is de laatste versie, welk document bevat een lijst met aanpassingen aan een bepaald methodologisch document,...). Er moet ten tweede ook een duidelijke connectie zijn tussen oudere data en hun bijhorende oudere versie van het methodologisch document zodat het duidelijk is welk de gebruikte methoden zijn voor die oudere data.
- f. De inhoud moet ook uniform gemaakt worden. Door het opdelen van de methodologische noten in logische blokken zal dit fel verbeterd worden. Zo zul je dus eerst moeten zien welke

logische blokken je moet voorzien (bijvoorbeeld blokken gaande over de nauwkeurigheid, de sample methode, de gebruikte standaard enz). En iedereen die binnen de organisatie zorgt voor het opstellen van bepaalde methodologische noten zal dus moeten zorgen dat hij de nodige informatie voor elk blok kan invullen. Hierdoor worden de methodologische noten completer.

g. Het geheel moet volledig zijn waarmee ik dan bedoel dat alle statistische gegevens moeten voorzien worden van methodologische informatie. Het kan niet dat een gebruiker bepaalde statistische info kan raadplegen zonder dat hij nergens de methodologische informatie met betrekking tot die statistische data kan raadplegen.

h. Ten slotte, één van de belangrijkste zaken aan statistische data en methodologische noten is het kiezen van een goede en liefst internationaal aanvaarde standaard waarop je je compilatieprocedures afstemt. Je moet deze standaard dan ook best zo goed mogelijk volgen. Indien er bepaalde punten zijn waarvan je afwijkt, moet je dit duidelijk aangeven en ook uitleggen waarom.

Als besluit van dit rapport volgt nu een opsomming van de belangrijkste punten die specifiek aangepast moeten worden of waarmee rekening gehouden moet worden bij het opstellen van het model voor de Nationale Bank van België. Deze punten zijn gebaseerd op enerzijds de voorgaande lijst met belangrijke punten, anderzijds met de analyse van de website van de Nationale Bank waar de tekortkomingen naar boven kwamen. Een uitgebreide versie is reeds gegeven in het voorgaande gedeelte (conclusies en aanbevelingen), maar door het hier nog eens kort op te sommen krijgen we een duidelijker overzicht van de belangrijkste puntjes.

a. De uniformiteit, zowel op gebied van lay-out als op gebied van indeling van de methodologische noten, moet fel verbeterd worden.

b. Het aanbieden van de informatie moet beter afgestemd worden op de verschillende gebruikerstypen.

c. De zoekmachine moet aangepast worden om specifiek te kunnen zoeken naar de methodologische noten. Op welke manier dit moet gebeuren gaat nog uitgebreid besproken worden in een later deel van het project.

d. Niet alleen de lay-out en indeling van de documenten moeten uniformer maar tevens het aanbieden van methodologische informatie. Hiermee bedoel ik dan vooral het uniformer maken van de link tussen de statistische gegevens en de methodologische noten.

e. Er worden momenteel verschillende versies van methodologische documenten bijgehouden, alleen moet de hiërarchie tussen deze verschillende versies duidelijker gemaakt worden en moet er een link komen tussen data en de juiste versie van het bijhorende methodologische document.

f. De inhoud kan ook iets uniformer gemaakt worden. Door het opdelen van de methodologische noten in logische blokken zal dit al fel verbeterd worden.

g. Het geheel moet vollediger waarmee ik dan bedoel dat alle statistische gegevens moeten voorzien worden van methodologische informatie.

Al deze gegevens moeten nu in het achterhoofd gehouden worden bij het opstellen van de organisatorische component en het disseminatiegedeelte van het uiteindelijke model voor de Nationale Bank van België.