

Prof.dr. R. Brand

# Datorum Scientia Scientiae

Wetenschap van gegevens, gegeven voor de Wetenschap



Universiteit Leiden

# Datorum Scientia Scientiae

Wetenschap van gegevens, gegeven voor de Wetenschap

Oratie uitgesproken door

**Prof.dr. R. Brand**

bij de aanvaarding van het ambt van hoogleraar op het gebied van

Good Research Data Management

aan de Universiteit Leiden

vanwege de European Group for Blood and Marrow Transplantation

op vrijdag 18 juni 2010.



**Universiteit Leiden**



*Meneer de Rector Magnificus, Members of the Board of the EBMT, leden van het Curatorium van deze bijzondere leerstoel, zeer gewaardeerde toehoorders in het Groot maar ook in het Klein Auditorium....,*

**“Beter zijn, Beter worden”** is de missie van het LUMC als wetenschappelijke werkplaats voor zorg en onderzoek. Het aardige is dat deze missie op zo veel manieren uitgelegd kan worden.

Beter zijn. Dan wie? Dan vroeger?

Beter worden. De patiënt? De onderzoeker? De student? Het onderzoek? Of misschien wel de data zelf?

Voor de sectie Advanced Data Management van de afdeling Medische Statistiek betekent “Beter zijn”, beter zijn in het structureren van kostbare gegevensverzamelingen, beter in het bewaken van kwaliteit en beter in de bescherming van privacy. Maar ook beter worden dan we nu zijn; door constante vernieuwing; door opleiding van docenten en onderzoekers; door het ontwikkelen van richting gevende principes; door volharding in een bewezen succesvolle aanpak.

In de Geneeskunde worden voortdurend gegevens verzameld. In de zorg moeten gegevens van patiënten heel betrouwbaar zijn: hun behandeling hangt er direct van af. Maar dat geldt ook voor onderzoeksgegevens: zij betreffen immers de behandeling van toekomstige patiënten.

Gegevens zijn de herinneringen in het geheugen van de wetenschap. Net als herinneringen worden zij herordend en in context geplaatst; worden zij gedeeld met anderen en gekleurd door ervaringen. Op gegevens worden oordelen gebaseerd; net als herinneringen, worden zij doorgegeven aan volgende generaties. Zonder herinneringen geen toekomst, zonder gegevens geen wetenschap.

Niet toevallig trouwens dat In het oud-Grieks “anamnese” synoniem aan “herinnering” is en dat het Latijnse “data” staat voor “dat wat gegeven is” ofwel “donaties”. Vandaar ook de titel

van mijn oratie: **Datorum Scientia Scientiae**; Wetenschap van gegevens, gegeven voor de Wetenschap.

Maar hoe verwerk je die gegevens? Hoe garandeer je foutloze opslag en feilloze oproepbaarheid door de grijze cellen van de wetenschap? Talloze systemen kun je bedenken en de bruikbaarheid ervan wordt bepaald door de mensen die de gegevens vastleggen, interpreteren en rapporteren...

Een citaat uit de “Litause Klavieren” van Johannes Bobrowski: *“Professor Voigt heeft de schriften in de conferentiekamer laten liggen. Het schiet hem te binnen doordat hij vaststelt: ik heb zo’n licht gevoel. Daarvan niet geschrokken, opluchting, je hebt een licht gevoel en nu weet je ook waardoor, dus verder met dit gevoel en met de vele papiertjes in al je zakken- gelukkig maar - waarop staat wat je te zijner tijd allemaal nodig hebt, dat is niet weinig, aantekeningen van etymologische aard, lijsten met namen, met overgenomen citaten. Je draagt je kostuum als een kast, een ander aantrekken betekent inventaris opmaken: een doorbladeren van al die papiertjes, eventueel opnieuw rangschikken, het uitgezochte apart leggen, om over te schrijven, iets dergelijks is het wel. Ja, het geeft je een licht gevoel, deze volle zakken, ja, het maakt een mens lichter.”*

Professor Voigt heeft een uitstekend datamanagementsysteem: zijn jasjes. Voor elke gelegenheid één. Voor elke situatie is er een ander jasje; wordt de hele inventaris opnieuw opgemaakt; wordt alles opnieuw gerangschikt en overgebracht. Professor Voigt krijgt er een licht en gelukkig gevoel van; hij wéét waar hij het kan vinden! Hij heeft alles onder controle...

Fijn dat professor Voigt alle tijd van de wereld heeft en dat hij alles kwijt kan in al die zakken. Mooi dat hij zo precies is dat er nooit een papiertje meegaat naar de stomerij om onleesbaar terug te komen; het treft dat er geen zakkenrollers zijn en er geen gaten in zijn zakken zitten. Toevallig of niet, het datamanagementsysteem zit professor Voigt als gegoten. Hoewel iedereen weet hoe dol ik op nette pakken ben, toch zou ik het niet graag zo zien bij mijn afdeling, al die datamanagers rondlopend met modieuze, dure data jasjes van Armani!

Om data te kunnen analyseren, moeten ze eerst nauwkeurig verzameld, feilloos geordend en veilig opgeslagen worden. Dit gebeurt in computer bestanden die we “databases” noemen. Helaas wordt deze term wat losjes gebruikt zodat wij anno 2010 nog steeds onderzoeksgegevens tegenkomen die in Excel of Word argeloos hun ongestructureerde, ongedocumenteerde en kwetsbare bestaantje leiden.

Deze oratie gaat dieper in op het structureren, bestuderen en beveiligen van onderzoeksgegevens die vaak gedurende vele jaren met moeite en tegen hoge kosten verkregen zijn. Maar laat ik beginnen de professorale Baret af en de Pet van de biostatisticus op te zetten. Althans in gedachten... Nu herkent u mij vast beter: 30 jaar aan het LUMC verbonden als medisch statisticus en het voorrecht genietend om met zovelen van u samen te werken en mijn steentje bij te dragen aan medisch onderzoek, zowel binnen het LUMC - bijvoorbeeld op het gebied van veel te vroeg geboren kinderen - als daarbuiten - bijvoorbeeld op het terrein van de beenmergtransplantaties.

In die statistische consultaties ligt de kiem van mijn huidige leeropdracht. Na een dus bepaald niet prematuur geëindigde zwangerschap van bijna 30 jaar, wordt het vakgebied van Good Research Data Management heden met enige trots aan familie, vrienden en collega's getoond... "Ars longa vita brevis" (dit voor de kenners van Hippocrates onder u).

Eerst moet ik u uitleggen wat het betekent om met medical research datamanagement bezig te zijn. Ik doe dat nog even met de vertrouwde pet van een biostatisticus op.

Veel klinisch-wetenschappelijk onderzoek begint met een simpele vraag:

“Zou behandeling A beter kunnen zijn dan B?”

“Is een bepaalde gewoonte een risicofactor voor het ontstaan van een ziekte?”

“Leveren alle ziekenhuizen even goede zorg en resultaten?”

Er moet dan een plan opgesteld worden om bewijsmateriaal (data) te gaan verzamelen ter ondersteuning (of juist ter

weerlegging) van een bepaald vermoeden. Net als in de rechtspraak moet bewijsmateriaal in de medische wetenschap, van onbesproken kwaliteit zijn: objectief, representatief, foutloos, ter zake doend enzovoort. Het bewijsmateriaal dient dus in overeenstemming met het doel en de redenering te zijn.

Rechter én wetenschapper zitten feitelijk in het zelfde schuifje; zij moeten - elk in hun domein - er van op aan kunnen dat de data zó goed zijn dat zij toevallig ontstane patronen inderdaad als “toevallig” kunnen herkennen. Je moet er tenslotte niet aan denken dat zij door een niet-onderkende toevalsvariantie of door verkeerde gegevens iemand veroordelen of een ziekenhuis of arts als “slechter dan anderen” kwalificeren.

Good Research Data Management of GRDM is het vakgebied dat zich bezig houdt met het ordenen, opslaan, beveiligen en gereed maken voor analyse van medische gegevens. In dit doolhof van regelgeving, privacy-bedreigingen, ICT-problemen en methodologische valkuilen dwaalt een driemanschap rond dat daar hand in hand uit probeert te komen: statisticus, datamanager en medicus. Zonder het overzicht van de datamanager over het doolhof aan gegevens, tast de statisticus hulpeloos rond. Als de moeizaam verkregen data niet (adequaat) geanalyseerd (kunnen) worden, raakt de datamanager meestal in een soort van postdatale depressie.

Het trio wordt gecompleteerd door de medicus, die weliswaar weet waar hij vandaan komt en waar hij heen wil, heeft maar geen flauw idee heeft hoe hij er zonder zijn maatjes ook echt komt. In de geneeskunde staat Wetenschap voor Samenwerking. Wetenschap berust op respect voor de kennis, kunde en intuïtie van de leden van dit driemanschap dat zijn weg moet vinden in een doolhof vol bedreigingen van hun data.

De weg door het doolhof is slechts te vinden door ervaring, slimheid, kennis en bundeling van krachten (en natuurlijk een espressoapparaat). Samenwerking tussen mensen, maar ook tussen computersystemen, als dé drijvende kracht. En wat is een beter uitgangspunt dan een afdeling Medische Statistiek

waar alvast datamanagement en statistiek bijeen zijn en het ook handig wisselen is tussen pet en baret? Het motto achterop de pet is overigens “The Power of Good Research Data”. Een doordenkertje ...

De term GRDM, “Good Research Data Management” (GRDM), is gevormd naar analogie van gevestigde begrippen als “Good Research Practice” en “Good Clinical Practice”. Deze term is bewust gekozen omdat het wezenlijk gaat om het “goede te doen”. Het heet dus niet “Better Research Data Management” of “Perfect Research Data Management”. De middelen zijn beperkt; de organisatie is groot; het onderzoek veelzijdig en omvangrijk en de kennis van datamanagement in alle eerlijkheid nog vaak te mager. Goed is- voorlopig - goed genoeg.

Wat zijn de kernbegrippen die we bij het ontwikkelen en implementeren van GRDM hanteren? Dat zijn “generaliseerbaarheid”, “eenvoud” en “bruikbaarheid” maar ook zaken als “kostenbesparing”, “kwaliteitsborging” en “privacybewaking”. Het eigen onderzoek naar methoden van GRDM is onlosmakelijk verbonden met de toepassing van die methoden. We ontwerpen niet zozeer een concreet project maar een set heel speciale gereedschappen. Daarmee bouwen we dan die projecten, infrastructuren voor het verzamelen en analyseren van onderzoeksgegevens. Tijdens dat bouwproces komen nieuwe ideeën op over die gereedschappen, meteen geabstraheerd tot algemene werkwijzen en toepassingen.

Dit betekent dat we bij ADM nooit voor een ad-hocoplossing van een nieuw concreet probleem kiezen: als er nog geen adequaat stuk gereedschap voorhanden is, ontwikkelen we altijd eerst dat nieuwe gereedschap en daarmee lossen we pas het concrete probleem op. Gereedschap dat meteen ten dienste staat van alle toekomstige maar ook van reeds vervaardigde projecten. Het kost even tijd maar het verdient zich altijd ruim terug.

Een goed voorbeeld van het rendement hiervan is “prototyping”: hiermee kunnen we binnen een paar weken nationale of internationale projecten realiseren. Zo kunnen onderzoekers hun project gewoon uitproberen. Eenvoudigweg omdat het

gereedschap en dus alle mogelijkheden, van het begin af aan aanwezig zijn.

Hoe realiseren we nu zo’n “project”? Er zijn altijd 4 belangrijke aspecten aan het ontwerp van een database:

- Statistisch: wat zijn het studie -design, de hypothesen en de analyses?
- Menselijk: wie zijn de gebruikers?
- Procesmatig: hoe worden de data verzameld?
- Kwalitatief: welke eisen moeten we aan datakwaliteit en privacy bewaking stellen?

U hoort het goed: mijn visie op datamanagement is niet die van de informatica maar die van de statistiek. Niet de theorie van de informatica en het databasedesign is uitgangspunt maar een goed gebruik van structuren en gegevens door onderzoekers, datamanagers en statistici is leidend bij ontwerpbeslissingen.

Dit maakt dat de wijze waarop ik vorm wil geven aan datamanagement vaak haaks staat op informatietheoretische beschouwingen, zoals het normaliseren van databases.

Het ontwerpen van een data base is in mijn ogen niets anders dan het “terug vertalen” van de uiteindelijke analyse naar een daarvoor geoptimaliseerde opslag en verwerking.

Het ontwerpproces dwingt als het ware de onderzoeker en passant de onderliggende vraagstellingen te concretiseren en zich op voorhand rekenschap te geven van het studie ontwerp en de analyses, die soms pas jaren later plaatsvinden.

Pragmatisme als theoretisch vertrekpunt; eenvoud en herhaalde bruikbaarheid als *conditio sine qua non*. Maar datamanagement is meer dan dataopslag. Data hebben pas betekenis in de context waarin ze verzameld zijn. Deze context noemen we metadata (ook wel Dictionary of CodeBoek).

Daartoe behoren

- alle definities van begrippen,
- alle checks op juistheid en consistentie,
- alle instructies bij het selecteren van de gegevens.

Alle logica dient namelijk in een goed managementsysteem geïntegreerd te zijn.

Projecten die de Zorg en Onderzoek domeinen pogen te versmelten - zoals Cura Rata en het landelijk Parelsnoer-initiatief - waarvoor wij een infrastructuur mogen leveren - zouden niet mogelijk zijn als de codes van beide domeinen niet in elkaar te vertalen zouden zijn. Problemen genoeg, overigens, want een goede en nuttige meting in het ene domein kan makkelijk onbruikbaar zijn in het andere.

Koppeling van zorg en onderzoek maakt het bijvoorbeeld mogelijk ziekenhuizen op kwaliteit te vergelijken. Daar is wel integratie van datamanagement en biostatistiek voor nodig. Patiënten, artsen en ziekenhuizen dienen beschermd te worden tegen methodologisch incorrecte conclusies, gebaseerd op onjuiste of onvergelykbare gegevens. Alle partijen hebben recht op een in rust en vrijheid tot stand gekomen, gewogen wetenschappelijk oordeel. Hier is haastige spoed nooit goed!

Zo gaan datamanager en statisticus hand-in-hand en kan de afdeling Medische Statistiek van het LUMC een bijdrage leveren aan het zuiver en eerlijk vergelijken van ziekenhuizen op kwaliteit van zorg juist door onze kennis van zowel datamanagement als biostatistiek.

Die statistische inbreng in de vergelijking van ziekenhuizen (en natuurlijk de consultatie in zijn algemeenheid) stoelt op de kennis die Hans van Houwelingen in de afgelopen jaren rotsvast in onze afdeling heeft verankerd. Daar ben ik hem nog steeds erkentelijk voor.

Al deze mooie principes hebben echter een gereedschapskist nodig. Een toolkit ontworpen met moderne ICT hulpmiddelen en conform inzichten uit de informatica. Mijn vakgebied behoort tot de methodologie van wetenschappelijk onderzoek; de gereedschapskist is gebaseerd op kennis van Informatie Technologie. En onze gereedschapskist heet ProMISe.

ProMISe is de toolkit waarmee data-analisten projecten bouwen; datamanagers de kwaliteit, privacy en logistiek bewaken; en waarmee het eerder genoemde driemanschap zijn gegevens tevoorschijn haalt voor verdere analyse. Maar het is ook een hulpmiddel om een visie op datamanagement te kunnen verwezenlijken.

En dat brengt mij dan op de rol van de organisatie die mijn leerstoel hier in Leiden heeft gevestigd, de European Group for Blood and Marrow Transplantation.

Het is 1987. Mijn collega Jo Hermans is al een tijd als biostatisticus betrokken bij de EBMT. Op de Hematologie in het Academisch Ziekenhuis Leiden werkt Ferry Zwaan, één van de oprichters van de EBMT rond 1975. Hij wil voor heel Europa één databasesysteem waarin alle beenmergtransplantaties verzameld worden om zodoende een solide basis te hebben voor onderzoek naar nieuwe behandelmethoden en inzichten in het ziekteproces. Jo antwoordt dat hij een collega heeft die al iets handigs doet met personal computers en gegevens. Hij belooft Ferry dat die collega wel even wat zal maken voor die paar honderd centra.

In 1988 draait de productielijn van de Project Manager op volle toeren. Duizenden floppy disks worden over heel Europa verstuurd met daarin de eerste Dictionary van de EBMT (eenheid van taal!). Data worden in Leiden gecombineerd (nog geen bruikbaar internet toen!) en Europese analyses waren rond 1990 een feit.

Rond 1998 werd de roep om een centrale, op internet technologie gebaseerde database groter en tijdens de Board meeting van de EBMT in Courmayeur in 1998 mocht ik aan de Board een prototype tonen van wat ter plekke in de wandelgang ProMISe werd gedoopt (Project Manager Internet Server). De Board onder de bezielende leiding van Andrea Bacigalupo gaf toestemming tot de verdere ontwikkeling van wat heden ten dage nog steeds het Europese datamanagementsysteem van de EBMT is: ProMISe.

De EBMT, waar samenwerking en respect voor elkaars vakgebied de gewoonste zaak van de wereld is. Waar vind je een specialisten-organisatie die toestemming geeft voor en veel geld investeert in software die toch volledig “generiek” is; op klinische gebieden buiten de BMT toepasbaar; waarvan ook andere organisaties gebruik mogen en kunnen maken? Waar je als statisticus en als datamanager volwaardig lid van het wetenschappelijk team bent en je mede vorm kunt geven aan toekomstige behandelingen en inzichten?

Wederom samenwerking als inspiratiebron. De EBMT: tientallen ziektebeelden, uniform en webbased verzameld in heel Europa in honderden centra met uniforme kwaliteitscriteria. Tientallen universiteiten die de gecombineerde gegevens van honderdduizenden transplantaties dagelijks analyseren. Parel-snoer avant la lettre, maar dan Europees en anno 1998!

ProMISe, Project Manager Internet Server, is een toolkit met een filosofie, geboren in 1998, met liefde grootgebracht binnen de EBMT en volwassen geworden binnen het LUMC. Maar evenzeer een filosofie met een toolkit.

Een visie op datamanagement om medisch-wetenschappelijk onderzoek te ondersteunen en waar nodig actief te structureren.

Maar ook weer niet meer dan een schil rond de kern van de zaak: de kostbare onderzoeksgegevens blijven te allen tijde in een goed gedocumenteerd en internationaal ondersteund formaat ter beschikking, zelfs als ProMISe zou wegvallen. Bij veel datamanagementprogramma's is helaas van zo'n “overleving” van de data zelf, geen sprake.

Samenwerking vormt het hart van de ontwerpideeën in ProMISe; niet alleen tussen mensen en disciplines maar ook tussen computersystemen. Cruciaal wordt dit als gegevens uit Ziekenhuis Informatie Systemen uitgewisseld moeten worden met een onderzoeksdatabase.

Om veel verschillende “connecties” te kunnen ondersteunen, kun je er als kleine organisatie gewoon niet aan beginnen om alle “verbindingen” dan zelf te programmeren of te implementeren. Het moet generiek, anders ben je verloren.

In onze gereedschapskist hebben we een set “interfaces”, algemene aansluitpunten die zodanig gedocumenteerd en state-of-the-art zijn, dat elke zich zelf respecterende ICT-afdeling daaraan zou moeten kunnen koppelen. Hiermee wordt de werklast gelegd bij het ziekenhuis dat aansluiting wil, maar wordt die koppelingsmogelijkheid toch als een reëel en haalbaar ICT instrument aangeboden.

Hebben we bewijzen dat de hierboven geschetste aanpak werkt? Dat de theoretische onderbouwing correct is? Dat de toolkit inderdaad het juiste gereedschap levert? Gezien de tijd die mij is toegemeten, kan ik slechts de volgende argumentatie aanvoeren:

sinds het ontstaan van ProMISe heeft ADM meer dan 150 infrastructuren voor klinisch onderzoek gebouwd voor 30 research organisaties, in binnen- en buitenland en is het in gebruik in meer dan 100 Nederlandse en bijna 1000 buitenlandse centra.

Het pure feit dat vele zeer gewaardeerde en hooggeleerde collega's hun kostbare Leidse onderzoeksgegevens, vaak in het kader van mede door hen opgerichte landelijk verbanden, toevertrouwen aan ónze manier van datamanagement, is veelzeggend. Zo verzamelen en analyseren wij voor Rob Nelissen protheses; voor Wilco Peul hernia's; voor Martin Schalijs pacemakers; voor Michel Ferrari migraines; voor Gé Luyten melanomen; voor Bob van Hilten Parkinson patiënten; voor Daan Hommes Parels; voor Robert Vermeiren eetstoornissen; voor Frans Helmerhorst in-vitro fertilisaties; voor Cock van der Velde naadlekkages; voor Rob Tollenaar coloncarinomen en voor Anneke Brand bloedtransfusies, om maar een paar onderwerpen te noemen. Hun vertrouwen is een enorme stimulans voor onze sectie. De samenwerking met hun superprofessionele medewerkers een waar genoegen.



Maar wij voeren ook beleidsondersteunende registraties zoals een sterfte-audit in samenwerking met het College van Zorgverzekeraars (CvZ), het landelijk patiëntveiligheidsonderzoek, incidentenmeldingen, ondernemingsraadverkiezingen, kwaliteit van arbeid enquêtes, kwaliteitsvergelijkingen van ziekenhuizen en, als vingeroefening, de organisatie van een Europees congres van fysici inclusief banktransacties en abstract management.

En niet te vergeten: u zit hier dankzij het ProMISe Party Management System!

Tussen 2003 en 2010 hebben ongeveer 10.000 gebruikers bijna 300.000 keer een ProMISe sessie gestart. In het totaal zitten meer dan een miljoen “patiënten” records in de data bases en zijn er meer dan 50 miljoen data elementen ingevoerd of gewijzigd via het eigen data entry interface.

The proof of the pudding is in the eating..... (met dank aan Cervantes hoewel ik hoop dat ik niet te veel weg heb van Don Quixote).

Maar natuurlijk dekken wij maar een fractie van alle onderzoeken die er in het LUMC lopen. Wie hebben er eigenlijk toegang tot al die andere bestanden met persoonlijke gegevens van patiënten en vrijwilligers? Worden de wachtwoorden goed bewaard of zitten de geeltjes op de beeldbuis geplakt? Worden de onderzoeksgegevens wel ontdaan van persoonlijke informatie voordat ze naar de statisticus reizen?

In het recent verschenen, uiterst informatieve Preadvies genaamd “Wetenschappelijk onderzoek in de Zorg”, analyseert de deskundige gezondheidsrecht mr.dr. Ploem de huidige ethische en juridische situatie ten aanzien van het gebruik van gegevens, lichaamsmateriaal en biobanken. De titel? “Gegeven voor de wetenschap”; ergo: “Data Scientiae”.

Eén illustratief onderdeel uit dit fraaie betoog wil ik hier uitlichten: haar analyse van het huidig juridisch vacuüm ten

aanzien van de ongecodeerde opslag van het Burger Service Nummer (afgekort BSN) voor research doeleinden: dat is gewoon niet geregeld.

Sluitende wetgeving hoeven we echter niet af te wachten. Straks zijn we namelijk in staat via een Trusted Third Party in real time versleutelde BSN’s op te slaan: een zorgverlener die naar zijn beeldscherm kijkt en daar gegevens van zijn of haar patiënt oproept, ziet wél alle persoonsgegevens en het BSN-nummer. Elke andere gebruiker die niet geautoriseerd is, ziet echter alleen maar onbegrijpelijke tekentjes. Slimme pixelletjes op je beeldscherm die informatie alleen maar tonen aan wie daar toegang toe heeft!

Er is duidelijk behoefte aan een heldere formulering van alle spelregels, zowel juridisch, wetenschappelijk als moreel; of het nu wetten zijn als de WMO of zelfregulering als de “Code Goed Gedrag”. Voor het LUMC als instelling is dat bijvoorbeeld GRP. Bij de sectie ADM is gekozen voor een NEN7511 certificering die wij met gepaste trots als eerste academische afdeling in Nederland van Lloyds in ontvangst hebben mogen nemen.

We moeten onderzoekers een gereedschapskist bieden die dan misschien geen “maatwerk” levert of “fashion design” maar toch alles bezit om volgens de spelregels te kunnen werken.

In de aanbieding, dames en heren: gereedschap met gebruiksaanwijzing, voor een vriendenprijsje! Voor dat vriendenprijsje dient het datamanagement in het LUMC dan wel georganiseerd te worden zoals de statistische consultatie: laagdrempelig voor wie behoefte heeft aan een sparringpartner; met gereedschap om de database vervolgens zelf te realiseren.

Een gedwongen gang langs de afdeling Medische Statistiek om de zegen van de Hogepriester der Biostatistiek of het Orakel van ProMISe te verkrijgen, is niet alleen onwerkbaar maar ook te kostbaar en contraproductief; het doet ook onrecht aan het principe van de volwassen arbeidsrelatie.

Maar als - terecht - toolkit noch expert dwingend worden voorgeschreven, dan moet de achterliggende kennis makkelijk toegankelijk zijn! Een schone bestuurlijke taak om dit gigantische kennisdomein te doen ordenen en ook daadwerkelijk voor onderzoekers open te stellen.

Bedenk daarbij echter wel dat sommige ándere onderzoeksinstellingen en met name de farmaceutische industrie, wel degelijk de accordering van statisticus en data-analist als een *conditio sine qua non* voor al het onderzoek hanteren bij wijze van kwaliteitsborging. Als wij niet zo ver willen gaan, dan moeten we wel de randvoorwaarden en spelregels heel strikt formuleren en implementeren.

Maar mag ik toch nog even een open deur intrappen? Men kan nog zo'n goede timmerman zijn, zonder goed gereedschap ben je nergens. Maar het goede gereedschap maakt je nog niet tot timmerman. Een arts is toch ook geen statisticus omdat hij SPSS aan de praat krijgt? Denk dus niet dat het ter beschikking stellen van ProMISe vanzelf leidt tot een optimale infrastructuur. Daar heb je echt altijd een dataspecialist voor nodig.

En mag ik dan ook nog één goede raad geven aan de managers onder u die verantwoordelijk zijn voor de implementatie van datamanagement projecten? Houd het eenvoudig en onafhankelijk van andere systemen; haal niet het onderste uit de kan!

Eer boven alles het motto van onze grote Boerhaave: *Simplex sigillum veri* - Eenvoud is het kenmerk van het ware.

### **De toekomst van de sectie ADM en de dienstverlening aan het LUMC**

Ik begin met een kort citaat van Marten Toonder: "Kijk, sprak heer Ollie, ik was uw verjaardag natuurlijk niet echt vergeten, want ik had het op een papiertje geschreven, dat ik even had weggelegd zodat het door mijn hoofd gegaan is". Ik heb iets degelikers voor ogen dan het papiertje van Ollie B. Bommel. Ik ga kort door de bocht met mijn "ideaalbeeld" maar de Raad van Bestuur zal mij dat voor deze éne keer wel

vergeven. U hoeft nu niets op te schrijven: de notulen verschijnen op DVD.

Ik breng mijn plannen onder in de voor een biostatisticus klasieke verdeling in consultatie, onderwijs en onderzoek maar voeg daar uitdrukkelijk de categorieën "HRM" (personeelsmanagement), "samenwerking" en "financiën" aan toe.

#### *Algemeen*

De sectie ADM blijft ingebed in de afdeling Medische Statistiek & BioInformatica. Methodologische ondersteuning beslaat daar het volledige spectrum van proefopzet, logistiek, data-basedesign, kwaliteitscontrole, analyse en rapportage.

ADM zal zich gaandeweg ook profileren als kennis- en opleidingscentrum. Huisvesting van medewerkers van andere afdelingen gedurende de opzet van ProMISe projecten, is mogelijk. Vaste krachten van ADM kunnen tijdelijk een werkplek elders krijgen om snelle kennisverspreiding in de organisatie te bevorderen. Ontdek je FlexPlekle!

#### *HRM*

Wij zijn op een punt aangekomen dat een solide, veilige, generieke en eenvoudige infrastructuur voor wetenschappelijk datamanagement LUMC-breed niet alleen kán, maar ook móet worden gerealiseerd. Er komen steeds meer signalen uit het LUMC dat zo'n voorziening gewenst en noodzakelijk is; een voorziening die betaalbaar is binnen de beperking van onze krimpende academische budgetten.

Ik denk dat te kunnen realiseren als ik kan komen tot een bescheiden maar vaste formatie van een aantal gepromoveerde en ervaren data-analisten op docent/onderzoeker niveau, die een brede ondersteuning van het eigen LUMC onderzoek mogelijk maken en, net als in de statistische consultatie, de rol van sparringpartner voor onze onderzoekers kunnen spelen. Dit garandeert de opbouw en instandhouding van een laagdrempelige, "kostenloze" toegang tot een basisvoorziening in veilig

datamanagement. Het biedt tevens medewerkers een gezond perspectief op een goede academische carrière en bindt hen aan ADM en het LUMC.

#### *Samenwerking*

Gezien de omvang en complexiteit van op ProMISe gebaseerde landelijke audits en kwaliteitsregistraties, zoals de Dutch Surgical Colorectal Audit, zoekt ADM een hechte samenwerking met de Stichting Informatie Voorziening Zorg (SIVZ), geheel in overeenstemming met de missie van het LUMC een duidelijk herkenbare maatschappelijke rol te willen spelen op het gebied van kwaliteitsbevordering.

Dankzij mijn lidmaatschap van de CME heb ik inzicht mogen krijgen in vele aspecten van al het mensgebonden onderzoek in het LUMC. Fundamentele discussies over bescherming van proefpersonen en privacy hebben de ontwikkelingen gestuurd in het denken over het spanningsveld tussen privacy bescherming en de noodzaak gegevens over ziekteprocessen langdurig te verzamelen. De leden en oud-leden van de Commissie Medische Ethiek en met name Dick Engberts, ben ik erkentelijk voor hun inzichten en overwegingen op dit gebied, die aantoonbaar hebben doorgewerkt in onze beveiligingsstrategieën. De samenwerking met ZorgTTP, een Trusted Third Party, in de ontwikkeling en implementatie van een door ons bedacht encryptie systeem, zal het bijvoorbeeld mogelijk maken de privacy nog verder te beschermen zonder het wetenschappelijk onderzoek te belemmeren.

Samenwerking met het Directoraat ICT van het LUMC was, is en blijft van groot belang voor de duurzaamheid van GRDM en ProMISe in het LUMC en scheidt vertrouwen bij klanten van ADM.

De directeur van het Directoraat ICT, Karel van Lambalgen, stimuleert actief de samenwerking tussen de DICT en ADM. Zijn uiterst collegiale steun en belangstelling voor ADM en die van vele DICT-ers in de afgelopen jaren wordt door mij zeer gewaardeerd. Natuurlijk hebben beide domeinen, Zorg en On-

derzoek, elk hun eigen problemen, mogelijkheden, beperkingen en doelstellingen. Ze zullen volgens mij ook nooit geheel samen kunnen vallen. Dat laat onverlet dat veel rendement uit een goede samenwerking en kennisuitwisseling te halen is nu ICT in Zorg en Onderzoek beginnen te vervloeien.

Een simpel maar prachtig voorbeeld is het feit dat het sinds gisteren mogelijk is om binnen de zorg, in Mirador, rechtstreeks research data van patiënten in ProMISe te bekijken en indien gewenst aan het zorg dossier toe te voegen.

#### *Onderwijs*

ADM zal actief in het onderwijs participeren. Het studenten onderwijs zal opnieuw ontwikkeld worden maar langs andere lijnen. Het richt zich op het verweven van methodologie, statistiek, datamanagement, beveiliging, documentatie en efficiency. De basis filosofie achter ProMISe zal leidend zijn; de nadruk zal echter niet op ICT-aspecten liggen.

Postacademisch wordt een bijdrage geleverd in het kader van basis cursussen biostatistiek en clinical trials en wordt aansluiting bij GCP- en GRP-onderwijs gezocht. Initiatieven, met name vanuit de Heelkunde, die zich richten op het tot stand brengen van een professionele opleiding tot datamanager, zullen van harte gesteund worden en delen van het academisch onderwijs hiermee vervlochten worden.

#### *Onderzoek*

Natuurlijk is er ook promotieonderzoek. Henk Jan van der Wijk brengt de principes achter de ontwikkeling van ProMISe in kaart; ontwikkelt een visie op dynamische encryptie en op ondersteuning van de samenwerking tussen medici en biostatistici door dynamische visualisering van survival modellen.

Ik prijs mij gelukkig dat de ervaren klinisch informaticus Pieter de Vries-Robbé uit Nijmegen bereid is als medepromotor op te treden. Hij sprak in een vroeg stadium al zijn overtuiging van het potentieel van ProMISe uit. Dat was een niet te

onderschatten stimulans en heeft uiteindelijk geleid tot een mooie samenwerking met het Radboud in Nijmegen in veel onderzoeksprojecten.

Het promotieonderzoek van Anne Marieke Schiere van de Stichting Perinatale Registraties bij Jan van Lith en mijzelf concentreert zich op de totstandkoming van de Landelijke Perinatale Registratie, de gevolgen daarvan voor de statistische analyses alsmede de processen die het ontstaan van zulke landelijke registratiesystemen bevorderen dan wel tegenwerken.

#### *Consultatie*

Er zal een leidraad komen om op simpele wijze gegevens veilig te kunnen opslaan en hun aanwezigheid in het LUMC transparant vast te leggen (DataSafe project). Een eenvoudig ProMISE meldingssysteem van data verzamelingen zal de privacy functionaris én de onderzoekers behulpzaam zijn.

We zullen streven naar een simpel, LUMC-breed meldingssysteem van Serious Adverse Events en SUSAR's.

Flowcharts en basisprincipes voor veilig en verantwoord datamanagement zullen onderzoekers helpen om met de middelen van hún keuze veilig gegevens voor onderzoek te verzamelen.

Dames en Heren. Eigenlijk is een oratie iets heel onwetenschappelijks. Halsmisdrijf nummer 1 in de wetenschap is het niet vermelden van de bronnen van je kennis; het doen-alsof-je-het-zelf-verzonnen-hebt. Maar in deze 45 minuten kan ik zelfs niet beginnen met recht te doen aan al diegenen die door hun houding, kennis, vragen, eisen, suggesties, voorbeeld en eruditie mij gevormd hebben.

De Leidse mores laten echter geen tijdsoverschrijding toe. In feite hebt u dus een synthese beluisterd van wat mij door al die mensen die ik hier niet kan noemen, direct of indirect is aangeleerd. Niet een gebrek aan dankbaarheid of erkenning maar de dwang van de tijd, laat velen in deze rede ongenoemd.

In the first place I would like to honour **all** EBMT members and personnel by mentioning just a few: Andrea Bacigalupo, Peter Ernst, Claude Gorin, Jane Apperley, Alois Gratwohl, Theo de Witte, Dietger Niederwieser, Per Ljungman and Carmen Ruiz whom I want to thank for their longtime cooperation and support in establishing this Chair for GRDM and the incredibly interesting and inspiring years among true friends. A very big “thank you” to all EBMT datamanagers who really shaped ProMISE over the years.

I'd like to express my sincere gratitude to the Board of the EBMT for the establishment of my Chair at the LUMC.

Leden en oud-leden van de Raad van Bestuur van het LUMC, leden van het Bestuur van Divisie 5. U hebt mij van meet af aan het gevoel gegeven dat ik kon rekenen op uw steun, op een organisatie met oprechte en directe belangstelling voor mijn vak en voor menselijke verhoudingen. Dat noem ik geen Management maar Leiderschap en dat waardeer ik zeer. Ik ben u daar zeer dankbaar voor.

Hooggeleerde Stijnen, beste Theo. Als afdelingshoofd heb jij vanaf dag 1 je volledige steun gegeven aan mijn benoeming en de inbedding van het datamanagement in de afdeling Medische Statistiek. Jouw brede kijk op wat Medische Statistiek is en wat datamanagement daaraan kan bijdragen, is cruciaal geweest.

Hooggeleerde Zwetsloot, beste Bertie. Bedankt dat je mij met raad en daad terzijde hebt gestaan bij het schrijven van mijn voorstel voor deze leerstoel.

Hooggeleerde Gratwohl, dear Alois. Our last paper together where we prove that JACIE accreditation and quality management do indeed substantially reduce mortality on a European scale, is the perfect apotheosis of a long, long cooperation between EBMT and LUMC. You were the very first one to support me in the process of attaining this great honour, to be EBMT professor at the Leiden University Medical Center.

Zeergeleerde Hermans, beste Jo. Jij en De Jonge, de oprichter van onze afdeling, hebben mij opgeleid tot medisch statisticus. Jij was een inspiratiebron van inlevingsvermogen en gedrevenheid tot samenwerken. Jij hebt altijd mijn datamanagement “hobby” gewaardeerd en gestimuleerd en daarmee een onmisbare schakel in mijn loopbaan gevormd.

Hooggeleerde van Bockel, beste Hajo. Jij bent al tientallen jaren een van die “true believers” in mijn aanpak van datamanagement. Het is een prachtige uitkomst dat nu de Heelkunde voor zoveel registraties en trials met enthousiasme ProMISe gebruikt.

Dames en heren studenten. Ik hoop dat ik naast mijn liefde voor de biostatistiek nu ook mijn ideeën over datamanagement in het onderwijs kwijt kan. Ik ben er van overtuigd dat simpele methoden u het leven aangenamer kunnen maken en meer tijd gunnen voor waar het echt om gaat: de optimale interpretatie van gegevens.

Dierbare medewerkers van ADM. Jullie vormen een fantastische en slagvaardige sectie en zijn met jullie ongekend enthousiasme en kundigheid in feite de sleutel tot het succes van GRDM in het LUMC. Jullie input is cruciaal; jullie humor geweldig. De manier van samenwerken, kennis overdragen en elkaar helpen, overtreft mijn stoutste dromen. De werklust spat er af: nijvere werkbijen, echte data beesjes. Ik ben er van overtuigd dat mijn sabbatical rimpelloos zal verlopen.

Weedelgeleerde van der Wijk, beste Henk Jan. Zonder jou als sparringpartner en mede-ontwikkelaar geen ProMISe. Lang geleden gaf je me als student al cruciale informatie waardoor baby ProMISe levensvatbaar werd. Je bent een zeldzaam voorbeeld van een informaticus-pragmaticus met inlevingsvermogen in het academisch onderzoek. Samen blijven we bouwen aan generieke en flexibele low-cost toepassingen in een academische wereld geplaagd door bezuinigingen.

Lieve familieleden en dierbare vrienden. Velen van jullie heb ik door de liefde voor mijn werk de afgelopen jaren schandelijk verwaarloosd. Dat jullie hier toch met zovelen heen gekomen zijn, bewijst dat vriendschap sterker is dan wat dan ook.

Lieve Anneke. Het is heel jammer mijn fantastische schoonmoeder nu hier niet voor me te zien zitten. Ik hoop dat je eerste SKYPE sessie op 82-jarige leeftijd vanuit de Cardiologie in Apeldoorn goed gelukt is!

Mijn ouders maken deze dag niet mee. Herinneringen aan wie zij waren en waar zij voor stonden, hun liefde en belangstelling zijn de fundamenten onder het bouwwerk van mijn bestaan. Mijn vader zou apetrots geweest zijn; mijn moeder zou zich daarnaast ook erg druk gemaakt hebben over de hoeveelheid werk. Ze wist dat deze dag er zou komen.

Lieve Gert Jan. Voor jou de laatste zinnen van deze rede. Achtentwintig jaar zijn we al dol op elkaar. Het is heerlijk om samen te leven met iemand die zo’n brede belangstelling voor wetenschap en kunst heeft als jij; en zo lief is; en dan ook nog zo mooi cello speelt. Natuurlijk heb ik te veel gecomputerd en werd je af en toe horendol van de om aandacht schreeuwende baby ProMISe, maar de komende 6 maanden sabbatical levert toch al gauw 180 dagen echt samen op pad naar Italië; dat is dus 15 compensatiedagen per ProMISe jaar. Kortom, ik ga op zorgverlof!

Ik heb gezegd.

## Ronald Brand



Na zijn promotie in de wiskunde aan het toenmalig Mathematisch Instituut in Leiden (1980) kwam Ronald Brand te werken bij de afdeling Medische Statistiek van het Academisch Ziekenhuis te Leiden. Tijdens zijn opleiding tot biostatisticus kreeg hij al gauw - in de praktijk van de statistische consultatie

- belangstelling voor de vele aspecten van data management. Rond 1985 ontwikkelde hij al de eerste PC applicaties die het verzamelen van gegevens voor onderzoeker en statisticus gemakkelijker en betrouwbaar maakten. In de vele jaren die hij als biostatisticus bij de afdeling Medische Statistiek werkte, hield hij zich enerzijds bezig met de opzet en analyse van klinische studies (consultatie), het onderwijs in de biostatistiek en de introductie van nieuwe ICT technieken in de biostatistische consultatie en anderzijds met de ontwikkeling van datamanagementmodellen en -implementaties als ondersteuning van wetenschappelijk onderzoek vanuit statistisch perspectief.

Wat als een soort van hobby-binnen-het-werk is begonnen, is in de loop van de jaren uitgegroeid tot het vakgebied Good Research Data Management dat ingebed is binnen de afdeling Medische Statistiek en ook onlosmakelijk met de biostatistische consultatie is verbonden. De sectie Advanced Data Management, waarvan Ronald Brand thans het hoofd is, is uitgegroeid tot een groep van zo'n 25 medewerkers die met de in de loop der jaren in ADM ontwikkelde software ProMISe zeer vele (inter)nationale studies en registraties van een data management infrastructuur voorzien hebben. De rode draad bij deze ontwerpen vormen begrippen als veiligheid (ADM is NEN7511 gecertificeerd) en methodologische correctheid.

Op het onderwijs terrein heeft hij in de loop der jaren het onderwijs in de biostatistiek voor studenten Gezondheidswetenschappen gegeven en in het afgelopen decennium heeft hij

in het post-academisch onderwijs de basis cursus Biostatistiek voor AIO's (en medewerkers) gegeven. Op dit moment wordt het datamanagement in het studentenonderwijs opnieuw ontwikkeld en geïntegreerd met het biostatistiek onderwijs.

Het onderzoek van zijn sectie richt zich op de relatie tussen statistiek en data management, de beveiliging van gegevens in de context van klinisch-wetenschappelijk onderzoek, de invoering en acceptatie van landelijke registraties, het ontwikkelen van strategieën voor het ontwerpen van onderzoeksdatabases en de aansluiting daarvan aan Ziekenhuis Informatie Systemen alsmede het gebruik van webbased-applicaties in de statistische consultatie.

Ronald Brand is als methodoloog sinds het begin van de jaren '90 lid van de Commissie Medische Ethiek van het LUMC en heeft daarmee het voorrecht kennis te nemen van al het mensgebonden wetenschappelijk onderzoek in het LUMC.

In de loop der jaren heeft Ronald Brand als UD/UHD Biostatistiek 13 artsen en klinisch onderzoekers als copromotor begeleid. Sinds 2009 liggen de onderwerpen zowel op het terrein van het medisch onderzoek als het klinisch datamanagement.

Per 1 juni 2009 is hij benoemd tot hoogleraar Good Research Data Management vanwege de European Group for Blood and Marrow Transplantation. Binnen de EBMT heeft hij als statisticus en als datamanagementspecialist 25 jaar kunnen bijdragen aan internationale onderzoeken op het gebied van de beenmergtransplantaties. De EBMT heeft de ontwikkeling van ProMISe op een unieke wijze gesteund en gestuurd en heeft er voor gezorgd heeft dat de opgedane kennis en ontwikkelde software ter beschikking van het LUMC is gekomen.

is the perfect apotheosis of a long, long cooperation between EBMt and LUMC. You were the very first to support me in the process of attaining this great honour of EBMt professor at the Leiden University Medical Center.

Most esteemed Hermans, dear Jo, You and de Jonge, the founder of our department, have trained me to be a medical statistician. You have been a source of inspiration in terms of your ability to understand the world of clinical research and drive to collaborate. You introduced me to the EBMt. You have always appreciated and stimulated my data management 'hobby' and by doing so you formed an essential link in my career.

Professor van Bockel, dear Hajó. For many years now, you have been one of the "true believers" in my approach to data management. It is wonderful to now see ProMISe being used with such enthusiasm for so many registers and trials in the field of surgery.

Ladies and Gentlemen students. I hope that, as well as my love for biostatistics, I can now teach you my thoughts on data management. I am convinced that simple methods can make life easier for you and will allow you time for what it's really all about: the optimum interpretation of data.

Dear ADM members of staff. You are a fantastic and skilled division and, with your unmatched enthusiasm and expertise, are the true key to the success of GRDM in the LUMC. Your input is crucial and your sense of humour is fantastic. The way in which you work together, exchange knowledge and help each other exceeds my wildest dreams. You are busy bees dripping with enthusiasm for your work; real little 'data Bs'. I'm convinced that everything will go smoothly during my sabbatical.

Most esteemed van der Wijk, dear Henk Jan. Without you as sparring partner and co-developer, no ProMISe. In 1998, as a computer science student who happened to be doing a little work on the side a few rooms down the corridor, you gave me

some crucial information which breathed life into the baby

ProMISe. You are a rare example of a pragmatic computer scientist with an ability to understand academic research. Together we will continue to build generic and flexible low-cost applications in an academic world plagued by cost-cutting.

Dear family, dear friends. I have shamefully neglected many of you over the past years because of my love for my work. The fact that so many of you have, nonetheless, come here today, proves that friendship is stronger than anything else.

Dear Anneke. I'm very sorry that my wonderful mother-in-law cannot be seated here in front of me today. I hope that your first SKYPE session at 82 years of age from the Cardiology Department in Apeldoorn is successful!

My parents are unable to be with us today. My memories of who they were and what they believed in, their love and their support form the foundations of my existence. My father would have been as proud as punch; my mother would have also been very worried about the workload. She knew this day would come.

Dearest Gert Jan. The last few words of this speech are for you. We have loved each other for twenty-eight years. It's truly wonderful to live with someone who is so kind, who has such a broad interest in science and art, and who can play the cello as beautifully as you do. I know I've spent too long behind the computer and that from time to time the attention-seeking baby, ProMISe, drove you crazy, but the coming 6-month sabbatical will give use 180 days to travel around Italy together: the equivalent of about 15 days' compensation per ProMISe year. In other words, I'm finally taking carer's leave!

Dixi

English translation: Catriona Ester, Mediltingua BV, Leiden, The Netherlands

rather time constraints which have resulted in many remaining unmentioned in this speech.

In the first place I would like to honour **all** EBMT members and personnel by mentioning just a few: Andrea Bactigalupo, Peter Ernst, Claude Gorin, Jane Apperley, Alois Gratwohl, Theo de Witte, Dieter Niederwieser, Per Ljungman and Carmen Ruiz whom I want to thank for their longstanding cooperation and support in establishing this Chair for GRDM, and for the incredibly interesting and inspiring years among true friends. A very big thank you to all EBMT data managers who really shaped ProMISe over the years.

I'd like to express my sincere gratitude to the Board of the EBMT for the establishment of my Chair at the LUMC.

Members and past members of the LUMC Board of Directors, members of Division 5 Management team. I'd like to address a personal word to you. You have, from the outset, given me the feeling that I could count on your support and on an organisation with a sincere and direct interest in my field and in personal relationships. This I consider not Management but Leadership, and I greatly appreciate it.

Professor Stijnen, dear Theo. As head of department you have, from Day 1, fully supported my nomination and the integration of data management into the department of Medical Statistics. Your broad vision of what Medical Statistics entails and how data management can contribute to this has been crucial.

Professor Zwetsloot, dear Bertie. Thank you for supporting me with advice and assistance while I was writing my proposal for this Chair.

Professor Gratwohl, dear Alois. Our last paper together where we prove that JACIE accreditation and quality management do indeed substantially reduce mortality on a European scale,

PROF.DR. R. BRAND

prepared to act as joint PhD supervisor. By voicing his conviction about the potential of ProMISe at an early stage he gave us an invaluable boost and this has ultimately resulted in great collaboration on many research projects with the Radboud in Nijmegen.

The PhD research by Anne Marieke Schiere under supervision of Jan van Lith and myself, concentrates on the development of the perinatal register, the implications for scientific research and the statistical analyses, as well as the processes which promote or hamper the development of this kind of national registration system.

*Consultancy*

A guideline will be developed about how to simply, but safely, store data and register their presence in the LUMC in a transparent manner (DataSafe project).

A simple ProMISe reporting system for data collections will help both the privacy officer and the researchers.

We will strive to create a simple, LUMC-wide reporting system for Serious Adverse Events and SUSARs.

Flowcharts and basic principles for safe and responsible data management will help researchers to collect research data safely, using their own chosen means.

Ladies and Gentlemen. An inaugural speech is really highly unscientific. Capital offence number 1 in science is the failure to cite your information sources or to pass it off as your own. But in these 45 minutes I can't even start to do justice to all those who have formed me through their attitudes, knowledge, questions, demands, suggestions, examples and erudition.

Leiden mores do not, however, allow you to overturn. Thus, you have in fact just heard a summary of all I have learnt (both directly and indirectly) from all those whom I cannot mention here. It is not lack of gratitude or recognition, but



it will offer members of staff a healthy outlook in terms of a good academic career and will keep them with the ADM and the LUMC.

#### *Collaboration*

Given the size and complexity of PROMISE-based national audits and quality registers, such as the Dutch Surgical Colorectal Audit, ADM is looking to form a close collaboration with the Association of Healthcare Information Services (*Stichting Informatie Voorziening Zorg, SIVZ*). This is in line with the LUMC mission to play a well-defined social role in quality enhancement.

Being a member of the Medical Ethics Committee (MEC) has given me insight into many aspects of research involving human subjects within the LUMC. Fundamental discussions about the protection of study persons and their privacy have fuelled developments towards finding a balance between the protection of privacy and the need to collect data on disease processes. I am grateful to the past and present members of the MEC, in particular Dick Engberts, for their views and considerations in this field, which have had tangible knock-on effects on our security strategies.

The collaboration with ZorgTRF, a healthcare Trusted Third Party, in developing and implementing our concept of an encryption system, will make it possible offer even greater privacy protection without hampering scientific research.

Collaboration with the LUMC IT directorate was, and still is of great importance to ensure the sustainability of GRDM and PROMISE in the LUMC, and instils confidence in ADM's clients.

The director of the LUMC IT directorate, Karel van

Lambalgen, actively encourages collaboration between the IT directorate and ADM. I greatly value his highly collegial support and interest in ADM, as well as that of many other IT directorate members of staff over the past years.

Clearly both areas, Healthcare and Research, have their own problems, possibilities, limitations and goals. I don't believe they will ever converge completely. Nonetheless, a good collaboration and exchange of knowledge can still be achieved now that Healthcare and Research IT is beginning to merge.

One simple but beautiful example is that since yesterday it has become possible for those working in the healthcare sector to use Mirador to directly access patients' research data in PROMISE and, if required, add this information to the patients' healthcare file.

#### *Training*

ADM will actively participate in the field of education. Student education will be redeveloped, but along different lines. The education will focus on aspects of data management, data security and documentation, as well as efficiency in data processing and analysis. The basic philosophy behind PROMISE will be key, although the emphasis will not be on the IT aspects, but rather on generally applicable methods. The acquisition of methods, statistics and data management will be promoted.

In post-academic terms, we will offer basic courses in

biostatistics and clinical trials, and we will aim to be in line

with GCP and GRP training courses.

Initiatives aimed at developing a professional data manager

course will be heartily supported and will overlap with parts of

the academic courses.

#### *Research*

And of course there is also PhD research. Henk Jan van der

Wijk is mapping the principles behind the development of

PROMISE; he is examining dynamic encryption and the support

for collaboration between doctors and biostatisticians using

dynamic visualisation of interpreting survival models.

I consider myself fortunate that the experienced clinical

informatician Pieter de Vries-Robbé from Nijmegen was

But if, quite rightly, neither the toolkit nor the expert is pushed upon people, the underlying knowledge should be made easily accessible! Organising this gigantic field of knowledge and actually making it openly available to researchers is an important administrative task.

However, don't forget that, as a means of quality control some other research institutes and, in particular, the pharmaceutical industry, do require the approval of the statistician and data analyst for all research. If we choose not to take things this far, we must ensure that we formulate very tight boundary conditions and rules, and implement these religiously. But allow me to state the obvious: it doesn't matter how good the carpenter is, without good tools he's nothing. On the other hand, good tools don't necessarily make a carpenter. And thus, a doctor doesn't become a statistician simply because he or she got SPSS up and running. So don't go thinking that making ProMISE available will automatically lead to an optimal infrastructure. To achieve this, you really do always need a data specialist.

And I have just one more piece of advice for those managers amongst you who are responsible for the implementation of data management projects: keep it simple and independent of other systems. Don't try to have it all!

Honour the motto of our own great Boerhaave: *simples sigillum veri* - simplicity is the seal of the truth.

### The future of the ADM section and provision of services to the LUMC

I'd like to start with a short quote from Marten Toonder: "Look, said Mr Ollie, of course I hadn't really forgotten your birthday, because I'd written it on a piece of paper which I had put away for a while so that it would pass through my mind". I have something greater in mind than Ollie Bumble's piece of paper. I'm going to get straight to the point with my 'ideal scenario', but I'm sure the Board of Directors will forgive

me just this once. Please don't worry about writing anything down: my speech will be available on DVD.

I'd like to organise my plans according to the classic biostatistician's system of consultancy, training and research, but I also think it is important to add the categories HR (personnel management), collaboration and finance.

### General

The ADM section will remain within in the department of Medical Statistics and Bioinformatics. Here, methodological support will cover the full spectrum of study design, logistics, database design, quality control, analysis and reporting. Along the way, ADM will also be profiled as a centre for expertise and training. Furthermore, we will be able to provide working space for staff from other departments while ProMISE projects are being set up. The regular ADM staff can be moved to another work station temporarily so that we can stimulate rapid distribution of knowledge within the organisation. Welcome to the world of "Flexi Desks"!

### HRM

We have reached a point where it is not only possible but also imperative to build a LUMC-wide scientific data management infrastructure which is solid, safe, generic, and simple. We are increasingly being confronted with signs from within the LUMC that such a service is both desired and necessary; a service which is affordable within the limits of our shrinking academic budgets.

I believe we can achieve this if we can put together a modest but fixed group of data analysts with PhDs working at the level of lecturer/researcher who would enable us to provide broad support to the LUMC's own research and, as in the field of statistical consultancy, would act as sparring partners for our researchers. This will allow us to build up and maintain accessible, 'free access' to a basic safe data management service. Furthermore,

But our projects also involve policy-supporting registers such as a mortality audit in collaboration with the Dutch healthcare insurance board (CVZ), incident reports, employees council elections, quality of work questionnaires, hospital quality comparisons and, as an exercise in dexterity, the organisation of a European congress of physicists, including bank payments and abstract management.

And don't forget that you are all here thanks to the PROMISE Party Management System!

Between 2003 and 2010, about 10,000 users initiated a PROMISE session just under 300,000 times. In total, there are more than one million "patient" records in the databases and more than 50 million data elements have been entered or altered through own data entry interfaces.

The proof of the pudding is in the eating..... (with thanks to Cervantes, although I hope I don't resemble Don Quixote too much).

But we only cover a fraction of all the studies being undertaken at LUMC. So who actually has access to all those other files of personal data from patients and volunteers? Are the passwords properly protected or are there sticky-notes stuck to the monitors? Is all personal information properly removed from the study data before the data are sent to the statistician? In the recently published and highly informative preliminary Advice, entitled *Scientific Research in Healthcare*, Dr. Ploem (lawyer and expert in healthcare law) analyses the current ethical and legal situation regarding the use of data, body samples and biobanks. And the title of this analysis? "Data for science"; hence the "Data Scientiae" in my own title.

One interesting example from this well-argued piece is the author's analysis of the current legal vacuum regarding the uncodified storage of the Dutch citizen service number (*burger servicenummer*, BSN) for research purposes: at present there is no adequate legislation in place to protect the BSN.

However, there's no need to wait for adequate legislation since we will soon be able to store real-time-encrypted BSNs using a Trusted Third Party. When a healthcare provider sits down in front of his/her computer screen and calls up the data of his/her patient, he/she *will* see all the personal data and the BSN number, but any other unauthorised user will only see unintelligible symbols. Clever pixels on your screen which only show information to people with authorised access to this information!

It goes without saying that we must clearly outline the rules of the game, be it legal, scientific or moral, or be it a law such as the Medical Research Involving Human Subjects Act (WMO) or a self-regulation system such as the Code of Good Conduct. For the LUMC, as an institution, this is GRP: The ADM division elected to go for NEN7511 certification and became the first academic department in the Netherlands to (quite proudly) receive this certification from Lloyds.

We need to offer researchers a toolbox which may not deliver tailored services or "fashion design", but which does contain everything required to work according to the rules of the game. Ladies and Gentlemen, today's special offer: tools with a handbook, for next to nothing! Yet, despite this give-away price, data management in the LUMC does have to be organised in the same way as a statistical consultancy service: easy access for those who are looking for a sparring partner; with the tools to subsequently develop the database themselves.

Having to walk down a compulsory corridor past the department of Medical Statistics to get the blessing from the High Priest of Biostatistics or the Oracle of PROMISE is not only impractical, but also too costly and counterproductive. Furthermore, it does an injustice to the principle of adult working relationships.

To support the many different “connections”, as a small organisation you can’t even consider doing the programming or implementation yourself. It has to be generic or you’ll sink. In our toolbox we have a set of “interfaces” in the form of general connection points which are so well-documented and so state-of-the-art that any self-respecting IT person in a hospital should be able to link up to them. In this way, the work associated with ‘linking up’ is shifted to the hospital that wants to be connected, but the option to connect *is* offered in the form of a realistic and achievable IT instrument. Do we have any proof that the approach outlined above works? That the theoretical foundation is correct? That the toolbox actually provides the appropriate tools? Given the limited time available, I can only put forward the following argument: since ProMISE was developed, ADM has built over 150 clinical research infrastructures for 30 research organisations within and outside the Netherlands, and more than 100 Dutch centres and almost 1000 centres abroad use ProMISE.

Simply the fact that many highly esteemed colleagues entrust their precious Leiden research data (in many cases collected as part of national collaborations which they themselves were involved in setting up) to our data management system says enough. For example, we collect and analyse data on prostheses for Rob Nelissen; hernias for Wilco Peul; pacemakers for Martin Schallj; migraines for Michel Ferrari; melanomas for Geert Luyten; Parkinson patients for Bob van Hilten; a String of Pearls for Daan Hommes; eating disorders for Robert Vermeiren; in-vitro fertilisation for Frans Helmerhorst; suture leaks for Cock van der Velde; colon carcinomas for Rob Tollenaar and blood transfusions for Anneke Brand, to name but a few. Their trust has been a huge boost to our department. Working together with their highly professional teams has been a true pleasure.

PROF.DR. R. BRAND

else do you find a specialists’ organisation giving permission for, and investing a great deal of money in, software which is actually completely ‘generic’; applicable in clinical fields outside BMJ, and which other organisations could also make use of? Where else, as statistician and as data manager, are you a fully-fledged member of the scientific team and can you contribute to forming future treatments and ideas?

Once again, collaboration is the source of inspiration. The EBMT: dozens of different diseases, collected in a uniform and web-based system across the whole of Europe by hundreds of centres with uniform quality criteria. Dozens of universities analyse the combined data from hundreds of thousands of transplantations each day. A true “String of Pearls” before the term even existed - in Europe and in the year 1998!

ProMISE, Project Manager Internet Server, is a toolkit with a philosophy, born in 1998, lovingly raised within the EBMT, reaching adulthood at the LUMC. But it is also a philosophy with a toolkit.

A vision on data management, a collection of modules which support, and where necessary, medical scientific research. On the other hand, it is no more than an outer skin surrounding the core issue: the valuable research data remain available at all times in a well-documented and internationally supported format, even if ProMISE were to breakdown. This is in contrast with many other data management programmes where, in such cases, there would be no chance of the actual data ‘surviving’.

Collaboration lies at the core of the design concepts in ProMISE: not only between people and disciplines, but also between computer systems. The latter becomes crucial when data from hospital information systems have to be exchanged with a study database.

can retrieve their data to analyse or transfer them to other computer systems. Not only is ProMISE a toolbox, but it is also a device which has allowed us to realise our concept of data management.

And this brings me to the role of the organisation which established my Professorial Chair here in Leiden, the European Group for Blood and Marrow Transplantation. The year is 1987. My colleague Jo Hermans had been involved with the EBMT for some time as a biostatistician. Ferry Zwaan, one of the founders of the EBMT around 1975, was working in the department of Haematology at the Academic Hospital of Leiden. He was looking for a single database system for the whole of Europe in which all bone marrow transplantations could be recorded so as to develop a solid basis for research into new treatment methods and insights into the disease process. Jo told him that he had a colleague who has been doing some quite useful things with personal computers and data. He promised Ferry that this colleague would quickly produce something for the couple of hundred centres.

In 1988 the project manager's production line was working to full capacity. Thousands of floppy disks were sent all over Europe containing the first Dictionary of the EBMT (unity in language!). Data were combined in Leiden (there was no useable internet at the time), and around 1990, European analyses become a reality.

Around 1998, the call for a central database based on Internet technology became stronger, and during the EBMT board meeting in Courmayeur in 1998, I was allowed to show the Board a prototype of what had been nicknamed ProMISE (Project Manager Internet Server). The Board, under the inspiring guidance of Andrea Bacigalupo, agreed to the further development of what today is still the EBMT's European data management system: ProMISE. The EBMT, where collaboration and respect for each other's field of expertise is the most normal thing in the world. Where

In other words, all logic should be integrated in a good management system. Projects which cross the line between the practical and research fields of healthcare, such as Cura Kata and the national String of Pearls initiative (for which we provide the infrastructure), would not be possible if the codes from both fields could not be translated from one to the other. Yet there remain many problems, because a measurement which is valid and useful in one field can be entirely useless in another.

By linking the practical and research fields of healthcare, it has become possible to compare hospitals in terms of quality. This does require the integration of data management and biostatistics. Patients, doctors and hospitals should be protected against methodologically incorrect conclusions, based on incorrect or incomparable data. All parties have the right to reach a well-considered scientific opinion, calmly and freely.

And therefore the data manager and statistician are working hand in hand and the department of Medical Statistics at the LUMC is now able to contribute to the unbiased and honest comparison of hospitals in terms of quality of care thanks to our data management and biostatistical knowledge.

My contribution to the biostatistical consultation in general and the comparison of hospitals in particular, is based on the knowledge which Hans van Houwelingen has firmly established in our department over the past years. For this I am still very grateful.

All these great principles do, of course, need a toolbox. A toolbox developed using modern IT tools and from a computer science perspective. My own field focuses on the methodology of scientific research; the toolbox is based on Information Technology knowledge. And our toolbox is called ProMISE. ProMISE, the toolbox with which analysts can build their projects, data managers can record their logistics and protect data quality, in which clinicians can safely store their patient and research data, and from which researchers and statisticians

The term GRDM, "Good Research Data Management", was conceived to be analogous to established terms such as Good Research Practice and Good Clinical Practice. This term has been chosen deliberately because it is all about 'good' work. For this reason we haven't called it "Better Research Data Management" or "Perfect Research Data Management". Our means are limited and it's a big organisation; the research is multi-faceted and broad, and, to be honest, knowledge of data management is often on the lean side. For now, *good* is good enough.

Which key concepts do we apply when developing and implementing GRDM? Well, these are terms such as 'generalisability', 'simplicity', and 'usefulness', but also issues such as 'cost-saving', 'quality protection' and 'privacy protection'. Our own research into GRDM methods is embedded in a cycle of improvements and is inextricably bound up with the application of these methods. Using our knowledge, insights and ideas, we first design and build a set of very special tools. We can then use these tools to construct projects or infrastructures to collect and analyse research data. During this building process, new ideas develop and, although based on a concrete goal, they are immediately extrapolated to general methods and applications. This is of immediate benefit to all construction projects, including existing ones.

This means that we never choose an ad-hoc solution for a new concrete problem if we don't yet have an appropriate tool available. We always develop the new tool first and only then resolve the concrete problem. These are tools that can be used for all future *and* completed projects. Although it can be a little time consuming initially, it is always more than worth it in the end.

A good example of the cost effectiveness of this approach is 'prototyping'. Using this tool, within one week we can create complete national or international applications. In this way researchers can try them out to see what the consequences of their initial study design will be in practice. This is only possible because the tools, and thus all the possibilities, are available from the outset.

A project is developed on the basis of 4 important aspects:

- Statistics: what is the study design, what are the hypotheses and the analyses?
- People: who are the users?
- Processes: how will the data be collected?

- Quality: what are the requirements in terms of quality and privacy protection?

Yes, you heard me correctly! My vision of data management is *not* one that stems from theoretical informatics but rather from statistics. Instead of using the theory of informatics and database design as the starting point, I believe correct use of structure and data by researchers, data managers and statisticians is key in making design decisions.

This means that the way in which I want to develop data management is often at odds with the concepts of theoretical informatics, such as database normalisation.

The design of a database is, in my view, no different to 'back-translating' the final analysis into an optimized storage and processing environment.

The design process more or less forces the researcher to make the underlying hypotheses visible within the design of the database and to recognize beforehand the implications of both the study design and the analyses, which are sometimes not performed until years later.

Fragmatism is *my* theoretical foundation; simplicity and repeated applicability are the prerequisites.

However, data management involves more than data storage. Data only take on meaning in the context in which they are collected. This context is known as meta-data (also known as Dictionary or Code Book). This includes:

- all definitions of terms,
- all checks for accuracy and consistency,
- all instructions when selecting data.

objective, representative, faultless, relevant etc. The evidence should therefore be in line with both the aim and the rationale.

Both judge and scientist are in the same boat: each in their own field needs to be certain that the data are of such high quality that they can identify any random patterns. After all, imagine sentencing someone or labelling a hospital or doctor as being “worse than others” simply because of unidentified randomness or because of incorrect information.

Good Research Data Management or GRDM is the field responsible for ordering, storing, and securing medical data, and preparing these data for analysis. Within this maze of rules, privacy threats, IT problems and methodological pitfalls, there is a three-man team wandering around hand in hand, trying to find a way out: statistician, data manager and clinician. Without the data manager’s birds-eye view of the data maze, the statistician finds himself groping around in the dark. If the data, acquired through blood, sweat and tears, cannot be (adequately) analysed, the data manager might fall into a “postdata” depression.

The trio is completed by the medic, who does know where he came from and where he wants to go to, but has no idea whatsoever how to actually get there without his mates. In the field of Medicine, Science stands for Collaboration. Science is based on respect for the knowledge, skills and intuition of the threesome who have to find their way through the maze full of threats to their data.

The way through the maze can only be found through experience, intelligence, knowledge and by combining skills (and of course a decent espresso machine). Collaboration between people, but also computer systems, is *the driving force*. And where better to start than a department of Medical Statistics where data management and statistics are already under one roof and where one can easily switch between cap and beret? By the way, the motto on the back of my cap is “The Power of Good Research Data”; just something to get you thinking ....

this term is applied quite broadly and even today, in 2010, we still find research data leading an unstructured, undocumented and vulnerable life in Excel or Word files. In this lecture I will take a closer look at how to structure, study and protect research data which have often been obtained over many years, with great effort and high costs.

But let me start by taking off my professorial beret and putting on my biostatistician’s cap. Figuratively speaking, that is... Now I’m sure you’ll recognise me much better: having been associated with the LUMC for 30 years as a medical statistician and having been privileged enough to work together with so many of you and to make my contribution to medical research, both within the LUMC (for example, in the field of very premature babies), and outside the LUMC (in the area of bone marrow transplantations, for example).

In those statistical consultancy projects lie the seeds of my current teaching commitments. Today, at the end of a certainly not premature pregnancy of nearly 30 years, I proudly present to you the professional field of Good Research Data Management... “*Ars longa vita brevis*” (for those of you who are familiar with Hippocrates):

First of all, I’d like to explain what medical research data management involves. I’ll keep on my biostatistician’s cap a little longer for this.

Many clinical scientific studies start by defining a specific research question such as: “Could treatment A be better than treatment B?” or:

“Is a particular habit a risk factor for developing a certain disease?” or:

“Do all hospitals provide an equal level of care and do they achieve equally good results?”

Next, a plan is drawn up to collect evidence (data) to support (or disprove) a particular theory. However, just as in the legal setting, medical evidence has to be of indisputable quality:

Honourable Rector Magnificus, members of the EBMT Board, esteemed members of the endowment fund for this chair, most highly small auditorium.....

**Being better, Getting better.** This is the mission statement of the LUMC, as scientific centre for medical practice and research. The nice thing is that it can be explained in so many ways.

“Being better”: than whom? Than what? Than before? “Getting better”: The patient? The researcher? The student? The research? Or maybe the actual data?

For the Advanced Data Management division of the department of Medical Statistics, “being better” means being better in structuring valuable data collections, being better in maintaining quality control and being better in protecting privacy. But it also means getting better than we are now, through ongoing innovation, by training lecturers and researchers, by developing guiding principles, by persevering with a proven successful approach.

In the field of medicine, data are continually being collected. In terms of healthcare, data from patients must be highly reliable, since the treatment of these patients depends on these data. But this is also true for research data, since these affect the treatment of future patients.

Data are, in fact, the memories in the mind of science. Just like memories, data are rearranged and put in context; they are shared with others and coloured by experiences. Judgements are based on data: just like memories they are passed on to the next generations. Without memories there is no future, and without data there is no science.

It's no coincidence, by the way, that in ancient Greek “anamnesis” is synonymous with “memory”, and in Latin “data” means “facts” or “donations”. Hence the title of my lecture: **Datorum Scientia Scientiae**; Science of Data, donated to Science.

But how should these data be processed? How do you

guarantee faultless storage and unfolding recall of data by the grey matter of science? Many different systems can be used to achieve this, and how useful these systems are depends on the people who record the data, interpret them and report them...

A quote from “Lithuanian Pianos” by Johannes Bobrowski: “Professor Voigt has left the notebooks in the conference room. Suddenly he realises this because he thinks ‘I feel so light. You don't feel concern, more relief. You feel light and now you know why, so carry on with this feeling and the (luckily) many papers in your pockets which contain everything you will need, which is quite a lot: etymological notes, lists of names, with copied quotes. Your suit is like a cupboard - putting on another would mean drawing up an inventory: leafing through all those sheets of papers, possibly reordering them, laying selected ones aside to rewrite, something like that. Yes, it gives you a light feeling, these full pockets, yes it makes a person lighter.”

Professor Voigt has an excellent data management system: his jackets. One for every occasion. For each situation, he has a different jacket; for each situation the whole inventory is checked again; everything is reordered and transferred. It makes Professor Voigt feel light and blissful; he *knows* where to find things! He has everything under control...

Luckily Professor Voigt has all the time in the world and room for everything in all those pockets. Hats off to Professor Voigt for being so precise that no paper has ever ended up at the dry cleaners and come back illegible. He has the good fortune that there are no pick-pockets, and thank goodness there are no holes in his pockets or that he never forgets his jacket...

Random as it may seem, the data management system may suit Professor Voigt, but despite my love for well-cut suits, I'd rather not see this system in my department: all those data managers in fashionable Armani data jackets!

To be able to analyse data, we first have to carefully collect the data, arrange them perfectly and store them safely. In practice we do this using computer files called databases. Unfortunately



Datorum Scientia Scientiae.  
Science of Data, donated to Science.

Inaugural lecture by

Prof.dr. R. Brand

on the acceptance of his position of professor of

Good Research Data Management

at the Universiteit Leiden

on behalf of the European Group for Blood and Marrow Transplantation  
on Friday June 18 2010.



Universiteit Leiden

Prof.dr. R. Brand

# Datum Scientia Scientiae Science of Data, donated to Science



Universiteit Leiden



Universiteit Leiden. Universiteit om te ontdekken.