

Vooronderzoek Samenwerkingsproject UKB, IWI, Uitgevers

Joost Kircz

KRA-Publishing Research

Office@kra.nl



Versie
6 april 2002

Vooronderzoek samenwerkingsproject UKB, IWI, Uitgevers

Joost Kircz

Motto: *Wij weten niet wat wij willen weten*

geanonimiseerd citaat van een UB-directeur

Korte Samenvatting & Conclusies

- Het gebied van elektronisch publiceren is aan het ontluiken. De vertaalslag van medium, van het traditionele papieren document naar een elektronische versie, is nu goed onderweg. De eerste contouren van hoe een werkelijk elektronisch document/ tijdschrift eruit ziet krijgen langzaam vorm. Serieus onderzoek naar de veranderingen die plaatsvinden in dit gebied is dan ook noodzakelijk.
- Helaas is het vak informatiewetenschap in Nederland slecht ontwikkeld. Dit in tegenstelling tot de VS en het Verenigd Koninkrijk waar de combinatie bibliotheek- en computerwetenschap traditioneel een grotere vlucht heeft genomen. De meeste Nederlandse groepen neigen sterk naar de formele informatica. Gezien het feit dat het al direct duidelijk werd dat de articulatie van vraag & doel bij de partners in dit project nog niet eenduidig is, is er in dit rapport afgezien van een overzicht van lopend Nederlands onderzoek op het brede gebied van kennismanagement, ontologiebouw, databaseonderzoek, beschrijvende talen (XML, SMIL, SVG, enz.) of data-mining. Binnen deze onderzoeken wordt veel mooi werk gedaan dat zeker in een volgende fase doelmatig kan worden toegepast op de onderzoeksvragen. Vragen die in de eerste fase van het project, dat dit rapport poogt vorm te geven, helder moeten worden.
- Gebruikersonderzoek is op dit moment van sterke flux steeds slechts een *snapshot* waarvan de resultaten al verouderd zijn bij presentatie. Om greep te krijgen op de intrinsieke ontwikkelingen die elektronisch publiceren bewerkstelligt alsmede om bij deze ontwikkelingen vooruitziend en liefst sturend op te treden, is het nodig langdurige longitudinale gebruikersstudies op te zetten.
- Hierbij moet expliciet aangetekend worden dat gebruikersonderzoeken sterk vakafhankelijk zijn. In het ene vak zoekt men vooral pertinente informatie (bijvoorbeeld getallen in de materiaalkunde), in een ander vak wil men vooral volledigheid (zoals rechten of toxicologie) en in de letteren wil men vaak flink bladeren. Gebruik van een artikel is dan ook vakafhankelijk en iets heel anders dan het daadwerkelijk opnemen van een gebruikt artikel in een referentielijst. Gebruiksanalyse is dus geen citatie-analyse.
- De eenheid van wetenschappelijke rapportage die het papieren product kent, zal zeker in de komende decennia uitgroeien naar zeer gevarieerde vakspecifieke (multimediale) presentatievormen. Daarom is het essentieel dat de materiële onderbouw van het proces op een algemeen aanvaarde gestandaardiseerde leest geschoeid wordt. XML, Z39.50/ZING, ed. zijn hier de trefwoorden.

- De elektronische productietechnieken zullen de traditionele functies van bibliotheek en uitgever in elkaar doen groeien terwijl ook bij de schrijvende-lezer en de lezende-schrijver de nodige veranderingen plaatsvinden. Een schets van de situatie wordt gegeven hoofdstuk I. Hoofdstuk II gaat in op de veranderingen die nu plaatsvinden. Hoofdstuk III is een overzicht van elementen voor onderzoek uitgaande van respectievelijk de schrijvende onderzoeker, de lezende onderzoeker, de uitgeeffunctie en de bibliotheekfunctie. Hoofdstuk IV gaat dan verder op gebruikersonderzoek terwijl hoofdstuk V een aantal probleem/onderzoeksgebieden schetst.

In het laatste hoofdstuk VI worden implementeerbare voorstellen gedaan, waarbij in VI-1→ 4 direct uitvoerbare projecten worden voorgesteld. In hoofdstuk VII worden twee veelbelovende onderzoeksvorstellen gepresenteerd die in de loop van het schrijven van dit rapport werden ingediend. Het Twentse voorstel wil meer ingaan op de methodologie van gebruikersonderzoek. Het Tilburgs voorstel aangaande *Digital Rights Management* (DRM) systemen geeft handvatten voor de implementatie van daadwerkelijke testen. Beide onderzoeksvorstellen dienen nog verder uitgewerkt te worden.

In bijlage B1 wordt een indruk gegeven van de buitenlandse initiatieven. B2 geeft een initieel overzicht van vraagstellingen bij logboekanalyse, B3 de lijst van gesprekspartners en B4 een literatuur overzicht.

Uit alle gesprekken alsmede uit de literatuur zijn de volgende punten duidelijk geworden.

1- Bibliotheken en uitgevers willen dezelfde gebruikersinformatie hebben, de ene partij voor collectievorming en ontsluiting, de andere voor productontwikkeling. Allereerst op een zo laag mogelijk aggregatieniveau (geschoonde ruwe data) en daarnaast in verschillende hogere aggregatietoestanden. Welke informatie er precies relevant is, is echter nog niet eenduidig gearticuleerd. Dat dit helder moet worden is dus een eerste conclusie die verder uitgewerkt wordt in de voorstellen.

2- De uitgevers zijn alleszins bereid om “ongeschoonde” logboekbestanden voor experimenten beschikbaar te stellen aan de afnemende bibliotheek binnen dit project. De vraag of geaggregeerde bestanden die alle UKB-partners beslaan ook beschikbaar zijn, is nog een punt van nader overleg tussen UKB en uitgevers. De uitgevers hebben immers met alle UB's aparte afspraken en het is niet de gewoonte gegevens van één klant aan een andere klant beschikbaar te stellen. Op internationaal niveau (beschreven in bijlage B1) is er veel activiteit op het gebied van logboekbestandsanalyse, waar uitstekend bij kan worden aangesloten, zie voorstel VI-1 voor een uitwerking. Voorstel VI-2, dat onderzoek behelst over het gebruikersgedrag bij zoeken en ontsluiten, sluit hier naadloos bij aan.

3- In alle gesprekken bleek dat er wel veel behoefte aan serieus onderzoek is, maar er nog nauwelijks gearticuleerde vraagstellingen zijn waarop empirisch onderzoek uitsluitsel kan geven. Het is daarom niet aan te raden om de planperiode van 2002-2004 nu al volledig in te vullen. De eenduidige conclusie die uit alle gesprekken volgt, is dat er eerst een analysefase nodig is om de termen, waarop en waarbinnen zinvol longitudinaal gebruikersonderzoek vorm kan krijgen, helder te benoemen. Concreet moet boven tafel komen welke gebruiks- en gebruikersinformatie de bibliotheek nodig heeft om het beleid van de bibliotheek te onderbouwen op de diverse niveaus zoals Instituut/faculteit, UB-bestuur, College van Bestuur, UKB.

4- De resultaten van dit vooronderzoek beslaan het voorstel tot het inrichten van een serie bijeenkomsten over:

a- logboekanalyse om een beter inzicht te krijgen in wat redelijkerwijs op dat vlak mogelijk is (VI-1),

b- informatie-ontsluiting en portaalinrichting om na te gaan hoe het gebruik verbetert en verandert in een elektronische bibliotheekomgeving en wat dat betekent voor de organisatie (VI-2),

c- redactionele aspecten om na te gaan hoe redacteuren van elektronische tijdschriften zich de toekomst van hun gebruikers (auteurs, referenten, auteurs) voorstellen (VI-3).

Het voorstel is dat deze workshops een basis leggen voor:

d- een werkgroep die expliciete onderzoeksvragen moet gaan benoemen (VI-4).

Op basis daarvan is aan de stuurgroep om welomschreven onderzoeksvragen vast te stellen die vervolgens aan de academische gemeenschap kunnen worden aangeboden.

5- Als het in dit vooronderzoek voorgestelde traject gevolgd wordt is het mogelijk aan het einde van dit jaar tot onderzoeksvorstellen te komen.

6- In hoofdstuk VII zijn twee reeds ingediende voorstellen genoemd. Het is interessant op te merken dat zowel KAP als ES expliciet interesse uitspraken in verder onderzoek naar de mogelijkheden die *Digital Rights Management (DRM)* systemen zouden kunnen opleveren. De bibliotheek van de faculteit geneeskunde in Maastricht zou graag als object van studie participeren, waardoor samen met de KUB dan ook twee verschillende vakgebieden vergeleken kunnen worden. Verdere uitwerking is hier noodzakelijk.

Het voorstel van de UT wil eveneens aan de eindgebruikerskant het daadwerkelijk gebruik onderzoeken. De insteek daar is van een meer methodologisch karakter. Ook hier is verdere uitwerking noodzakelijk.

7- De consequentie van dit alles is waarschijnlijk dat de beschikbare gelden niet noodzakelijk ultimo 2004 alle doelmatig besteed zijn. Het advies is dan ook de effectieve looptijd te laten bepalen door de kwaliteit van de projecten. Het moeilijke karakter van de sterk veranderende omgeving waarin de partners opereren, dwingt immers tot langdurige gezamenlijke inzet.

8- Een van de gesprekspartners in dit vooronderzoek stelde de vraag hoe de activiteiten van de stuurgroep op blijvende wijze gecontroleerd en beheerd worden. Een rol voor IWI werd hierbij als mogelijkheid genoemd.

Inhoudsopgave

Korte samenvatting en conclusies.....	2
Inhoudsopgave	5
Inleiding op dit rapport.....	6
I) Schets van de situatie in algemene termen	8
1) Fasen in het proces.....	8
2) De uitgeeffunctie	9
3) De bibliotheekfunctie.....	9
II) Wat er verandert	10
1) Concreet zien wij de volgende tendensen.....	11
III) Elementen voor onderzoek	13
1- De schrijvende onderzoeker	13
2- De lezende onderzoeker	14
3- De veranderende rol van de uitgeeffunctie	15
4- De veranderende rol van de bibliotheekfunctie	16
IV) Gebruikersonderzoek	18
V) Onderzoeksmogelijkheden	19
1- De auteursomgeving	19
2- De lezersomgeving en de ontsluiting van informatie	20
3- De meting van gebruik en wat gebruik is.	20
4- Domein specifieke kenmerken van informatiebehandeling	21
5- Kennisdiffusie en padanalyse	22
6- Niet bibliografische links.....	22
VI) Aan te vatten projecten.....	23
1- Workshop logboekanalyse en statistiek	26
2- Workshop ontsluiting.....	28
3- Redacteursconferentie.....	31
4- Werkgroep onderzoeksvragen	32
VII) Ingediende onderzoeksvorstellen	33
1- Experimenten met DRM (KUB)	33
2- Voorstel Universiteit Twente	36
Bijlagen:	
B1- Internationale initiatieven	38
B2- Initiële lijst van te stellen vragen aan een logboekanalyse	40
B3- Lijst van gesprekspartners	42
B4- Literatuurreferenties.....	43

Inleiding op dit rapport

In januari 2001 begon de discussie hoe de samenwerking tussen de UKB partners, de wetenschappelijke uitgeverijen Wolters-Kluwer en Elsevier Science en het IWI vorm gegeven kan worden.

Na enkele voorontwerpen is de tekst *Onderzoekssamenwerking UKB, IWI, Uitgevers versie 3, februari 2001* op de eerste stuurgroepvergadering van 27 juni 2001 besproken. Hierop is de tekst *UKB-Uitgevers project dd 4 september 2001* besproken op de tweede stuurgroepvergadering van 21 november. Op deze vergadering werd een opdracht tot vooronderzoek aan opsteller van dit rapport gegeven, dat contractueel op 10 december 2001 werd geëffectueerd. Het was expliciet de bedoeling een inventarisatie te maken en niet over te gaan tot een “tender” voor onderzoeksprojecten.

Het rapport *Vooronderzoek Samenwerkingsproject UKB, IWI, Uitgevers versie 2 maart 2002*, is besproken op de stuurgroepvergadering van 8 maart 2002. Verzocht werd om de voorgestelde 3 workshops nader uit te werken almede nog even de stofkam door de hele tekst te halen. In deze versie is e.e.a. geïmplementeerd. Voor alle drie workshops zijn concrete plannen en organisatie comités. Separaat aan dit rapport wordt een financieel plaatje gepresenteerd.

Ter informatie herhaal ik hieronder de essentie van de oorspronkelijke werkopdracht.

Uitgangspunt

Besloten is dat van de notitie d.d. 4 september punt 2.1.B “Analyse van de Huidige Trends. Kennis van de (domein)specifieke creatie en consumptie van wetenschappelijke informatie” het belangrijkste thema wordt. Almede dat er drie themaconferenties worden georganiseerd.

Doel van het uiteindelijke onderzoek

Inzicht geven in

- Het lezen en gebruiken van wetenschappelijke artikelen (alle stadia), E. versus P.
- De veranderende rol van literatuur nu zij on-line permanent aanwezig is.
- De mate waarin er meer interdisciplinair gelezen wordt.
- De mate waarin de kwaliteitsperceptie van artikelen of tijdschriften geschraagd wordt door daadwerkelijk gebruik.
- De mate waarin niet-tekstuele elementen een grotere rol gaan spelen in de wetenschappelijke communicatie.
- Inzicht in de aard en beschikbaarheid van gebruikersgegevens.
- Inzicht in hun manipuleerbaarheid en mate van bruikbaarheid.
- Voltooide, lopende en geprojecteerde onderzoeken naar gebruikersgedrag, verlangens/verwachtingen

Dit om aanzetten te geven voor het ontwikkelen van nieuwe economische modellen voor het wetenschappelijk tijdschrift.

Leveranties vooronderzoek

Implementatieplan voor onderzoek op het gebied van boven gespecificeerd gebied met als looptijd 2002-2004, inclusief duidelijke fasering, structuur van samenwerkingsverbanden, eerste kostenplaatje, rapportage momenten en themaconferenties.

Dit implementatieplan zal gebaseerd zijn op onderstaande elementen en considerata

- 1.1 Overzicht van de beschikbaarheid en het toekomstige beschikbaar komen van gebruikersgegevens van bibliotheken en uitgeverijen en de vorm waarin dit gebeurd/ zal gebeuren.
- 1.2 Overzicht van de mogelijkheden en ambities van de deelnemende organisaties voor concreet onderzoek dat leidt tot het bereiken van eerder genoemde doelen.
- 1.3 Eerste overzicht van gelijkaardige onderzoekingen en plannen in het buitenland.
- 2.1 Het aanscherpen van de doelen van het project opdat binnen de bemeeten tijd resultaten geboekt kunnen worden.
- 2.2 Een overzicht van mogelijke uit te voeren projecten in Nederlanddoelen, tijdspaden en kosten.
3. De opzet tot drie conferenties/workshops over dit thema.

In een uitgebreide serie gesprekken en brede literatuurstudie werd het snel duidelijk dat er wel bij alle partijen een grote behoefte bestaat aan transparante en overzichtelijke gebruikersinformatie als voorwaarde tot bestuurlijke stappen, maar dat het verre van duidelijk is om welke informatie het gaat, hoe die te meten is, en hoe dan te interpreteren?

Om die reden is er in dit rapport gekozen tot een getrapte opstelling waarbij eerst geprobeerd wordt het probleemgebied duidelijk te beschrijven. Van daaruit worden dan de veranderingen in de diverse functies en rollen beschreven. Vervolgens wordt ingegaan op mogelijke onderzoeksthema's om af te sluiten met concrete voorstellen.

In alle gesprekken bleek dat er wel veel behoefte aan serieus onderzoek is, maar nog nauwelijks gearticuleerde vraagstellingen waarop empirisch onderzoek uitsluitsel kan geven. Het is daarom ook niet zinvol om de planperiode van 2002-2004 nu al volledig in te vullen. De eenduidige conclusie die uit alle gesprekken volgt is dat er eerst een analysefase nodig is om de termen, waarop en waarbinnen zinvol longitudinaal gebruikersonderzoek vorm kan krijgen, helder te benoemen.

De conclusie van dit vooronderzoek is dat het zinvol zou zijn een serie bijeenkomsten (logboekanalyse, informatie ontsluiting en redactionele aspecten) te starten, die de basis moeten leggen voor expliciete onderzoeksvragen.

Het voorstel is dat dan een werkgroep onderzoeksvragen de resultaten van de andere bijeenkomsten zou moeten vertalen naar voorstellen. Dit zou dan de basis kunnen vormen waarop de stuurgroep welomschreven onderzoeksvragen kan formuleren. Deze vragen zouden dan aan de academische gemeenschap kunnen worden aangeboden. Vanuit de KUB en de UT zijn in de loop van dit vooronderzoek reeds suggesties gegeven.

In hoofdstuk VII zijn deze voorstellen ter informatie opgenomen. Het is interessant op te merken dat zowel KAP als ES expliciet interesse uitspraken in verder onderzoek naar de mogelijkheden die *Digital Rights Management* systemen zouden kunnen opleveren. De bibliotheek van de faculteit geneeskunde in Maastricht zou graag als object van studie participeren, waardoor samen met de KUB dan ook twee verschillende vakgebieden vergeleken kunnen worden. Verdere uitwerking is hier noodzakelijk. Het voorstel van de UT wil eveneens aan de eindgebruikerskant het daadwerkelijk gebruik onderzoeken. De insteek daar is van een meer methodologisch karakter. Ook hier is verdere uitwerking noodzakelijk.

De consequentie van dit alles is waarschijnlijk dat de beschikbare gelden niet noodzakelijk ultimo 2004 alle doelmatig besteed zijn. Het advies is dan ook de effectieve looptijd te laten bepalen door de kwaliteit van de projecten. Het moeilijke karakter van de sterk veranderende omgeving waarin de partners opereren, dwingt immers tot langdurige gezamenlijke inzet.

Een positief bijkomend verschijnsel van een voorzichtige start is dat de uitkomsten van de voorgestelde workshops het mogelijk maken om daadwerkelijk de termen waarin gebruikersonderzoek gevat moet worden te formuleren. Dit is op zich al belangrijk voor iedere discussie over de rollen, taken en financiële consequenties tussen bibliotheken en uitgevers.

In dit rapport gaat het expliciet niet om de discussie over publicatiemodellen zoals bijvoorbeeld: traditioneel (Mea1988), een keten van pre-print servers (Har1999), of experimenten met locale uitgeefdiensten (Arn2002, Kli2002). Over dit -vaak emotionele- aspect is er inmiddels een enorme literatuur waaruit in ieder geval blijkt dat de meningen vaak ingegeven zijn door de specifieke vakgebieden van de auteurs (een redelijk overzicht is: Pee1996).

I) Schets van de situatie in algemene termen

I-1) Fasen in het proces

In de waardeketen van auteur naar lezer zijn verschillende fasen te onderkennen. Voor de volledigheid noemen wij die nog eens.

1) **De onderzoeksfase**, waarbij de auteur informatieconsument is (lezer) van reeds bestaande documentaire informatie (rapporten, getallen- en beeldbestanden, artikelen, collegedictaten, (les)boeken, enz).

2) **De schrijffase**, waarbij de auteur zijn/haar kennis in een, voor het onderhavige vakgebied, maatschappelijk erkende wijze rapporteert. De auteur gebruikt hiervoor steeds meer (semi) geautomatiseerde hulpmiddelen (tekstverwerkers (WordPerfect, MSWord, LaTeX, enz), spellingcontrole- en beeldbewerkingsprogramma's en referentie hulpmiddelen (bijv. Procite, Endnotes), enz.). In deze fase gaat het niet meer om algemene stavende informatie, maar om pertinente verwijzingen en/of citaten.

3) **De aanbiedingsfase**, waarbij de auteur de rapportage ofwel direct openbaar maakt zoals bij een collegedictaat of een "invited speech", ofwel aanbiedt aan een verificatie/ certificatie gremium in de vorm van een (promotie)commissie, een (tijdschrift)redactie of een commerciële opdrachtgever.

4) **De authenticatie/ validatie/ certificatiefase**, waarbij de validerende instantie de bron (auteur) authentiseert, de rapportage beoordeelt, zo nodig wijzigingen voorstelt en al dan niet certificeert (accepteert voor publicatie in het door haar gecontroleerde kanaal). In deze fase wordt onder meer aan de hand van de reeds beschikbare gecertificeerde (en in een enkele geval ook nog niet gecertificeerde) literatuur de kwaliteit en relevantie beoordeeld. Ook worden de gegeven referenties naar de literatuur en/of gegevensbestanden gecontroleerd.

5) **De productiefase**, waarin de (al dan niet) gevalideerde en gecertificeerde rapportage in een gestandaardiseerde vorm op een medium publiek wordt gemaakt. In deze fase wordt er aan de rapportage een voldoende verzameling (bibliografische) metadata meegegeven, waardoor het eindproduct een-eenduidig via verschillende routes aanspreekbaar wordt.

6) **De distributiefase**, waarin het (gecertificeerde en van metadata voorziene) eindproduct haar weg vindt naar de consument.

7) **De indexeringsfase**, waarin, binnen een specialistische verzameling van rapportages, domein specifieke trefwoorden en/of classificatiecodes worden toegevoegd. Let wel, reeds in de fasen 3-5 zijn er indextermen toegevoegd in de vorm van o.m. bibliografische informatie en sectie-informatie van bijvoorbeeld een tijdschrift.

8) **De archiveringsfase**, waarin de rapportage op een procedureel vastgelegde wijze toegevoegd wordt aan "het wetenschappelijk en cultureel erfgoed".

9) **De ontsluitingsfase**, waarin de beschikbare literatuur op basis van de verschillende indexen daadwerkelijk aan een eindgebruiker ter hand kan worden gesteld.

10) **De leesfase**, waarin onafhankelijk van tijd of plaats een geïnteresseerde op één of andere

manier kennis neemt van de rapportage. Deze fase gaat binnen de academische wereld weer over in een nieuwe fase 1.

Het vervolg van dit rapport zal het slechts gaan om die rapportages die uiteindelijk, volgens de huidige standaarden, als artikelen gepubliceerd worden in reguliere wetenschappelijke tijdschriften.

Een wetenschappelijk tijdschrift definiëren we hier als een bewust bepaalde verzameling wetenschappelijke artikelen die zich kenmerkt door een eigen (niet noodzakelijk uniek) wetenschappelijk niveau, domein specificiteit en cultuur (school).

Uitgaande van bovenstaande kunnen we naast de rol van de schrijver en de lezer, traditioneel twee groepen functies onderscheiden die aan het proces auteurs/lezer onafhankelijke waarde toevoegen.

I-2) De uitgeeffunctie

De functie beslaat de organisatie van alle veredelingsstaken tot en met de geboorte van het gevalideerde, gecertificeerde en geauthentiseerde eindproduct. Een product dat als een duidelijke eenheid (informatie-object) een intellectuele eigenaar heeft en als onafhankelijke eenheid reproduceerbaar is.

Deze functie komt in fase 1 voor als bezorger van noodzakelijk bronnenmateriaal, in fase 2 als richtlijngever in de vorm van instructies aan auteurs ter begeleiding van het schrijfproces, in fase 4 als verzorger van het validatie/ certificatieproces (*Peer-review*) en in fase 5 en 6 als opdrachtgever/ uitvoerder. Het gaat hier om intellectuele productontwikkeling.

I-3) De bibliotheekfunctie

Deze functie kenmerkt zich door het kant en klare artikel als uitgangspunt te nemen. In fase 1, 2 en 10 functioneert zij als facilitator van de consumerende auteur/ lezer. In de fasen 7, 8 en 9 wordt gebruikersgemeenschap afhankelijk geordend, geselecteerd, opgeschoond en beschikbaar gesteld. Het gaat hier om intellectueel productbeheer en collectievorming.

Men kan in deze indeling nog twisten of de indexeringsfase (fase 7 op de vorige pagina) een uitgeef- of een bibliotheekfunctie is; duidelijk is dat juist hier een belangrijk draaipunt in de keten ligt, tussen de door de uitgever verzorgde auteurs gestuurde processen en de door de bibliotheek gestuurde lezersprocessen.

II) Wat er verandert

Naast de inhoudelijke functiever verschillen tussen bibliotheek en uitgever zijn veel van de huidige verschillen in organisatie, maatschappelijke positie en traditie voortgekomen uit de materiele productiefactoren. De uitgever heeft zijn kantoren in plaats A, laat het werk drukken en binden en stuurt het vervolgens de wereld rond waar dan lokaal handelaren en bibliotheken de verzorging van de uitlevering en beschikbaarstelling regelen.

De uitgever produceert en de bibliotheek beheert en archiveert. Tot voor kort was deze scheiding de normale gang van zaken.

Een essentieel verschil tussen de rollen was steeds dat de uitgever over het algemeen de legale rechten heeft op een werk maar geen kopieën en de bibliotheek geen rechten heeft maar wel kopieën.

Over de problemen met rechten zal er in onderstaande niet verder worden ingegaan. Dat sommige problemen van alle tijden zijn mag blijken uit het volgende voorbeeld. Toen Oxford University Press, die de rechten op een nieuwe bijbelvertaling had, haar boeken in 1881 voor de Amerikaanse markt verscheepte, zag tijdens de overtocht een Amerikaans concurrent kans het geheel te zetten (in lood dus), om het bij aankomst direct ter perse te brengen. Bij aankomst in New York werd tevens een belangrijk deel naar the Chicago Tribune getelegrafeerd (een vroege vorm van elektronisch publiceren) die e.e.a. terstond in de krant afdruckte.

Voldoende is om op te merken dat er op vele internationale fronten (WTO, EU) gewerkt wordt aan een afstemming van de zeer verschillende tradities en wetgeving op het gebied van intellectueel eigendom.

In het vervolg van dit rapport gaat het dus niet om persistente problemen van alle tijden zoals plagiaat, vervalsingen en inbreuk op wettelijk vastgelegde rechten, maar wel om greep te krijgen op de uitoefening van de functies en de daaraan gerelateerde rechten en hun economisch verkeer in een veranderende technologische omgeving.

Elektronisch publiceren dicht de kloof tussen productie, distributie, opslag en ontsluiting. Het is daarom duidelijk dat, hoewel in principe de functiever verschillen blijven bestaan, deze functies niet meer op dezelfde wijze over de eerder genoemde 10 fasen verdeeld zijn.

Immers:

- 1) na finale productie kan er eerst verzonden worden en dan pas gevalideerd, gecertificeerd en gedrukt, in plaats van andersom,
- 2) elektronische productie maakt het aanhouden van een compleet archief door de uitgever relatief makkelijk,
- 3) aan elektronische bestanden kan altijd nog extra indexering worden toegevoegd zonder tot “herdruk” te hoeven over te gaan,
- 4) het begrip versie kan “vloeibare” karakteristieken krijgen,
- 5) de verhouding tussen beschikbaar archief en ambulante collectievorming voor een omschreven doelgroep verandert,
- 6) de accent-verschillen tussen uitgeversimprint, tijdschrifttitel, en auteursnaam bij het zoeken naar literatuur gaat verschuiven,
- 7) de bibliothecaire en de uitgeeffuncties vervlechten en kunnen (gedeeltelijk) wisselen van uitvoerder (de bibliotheek als uitgever en de uitgever als bibliotheek).

II-1 Concreet zien wij de volgende tendensen

- 1) Auteurs kunnen hun werk met een druk-op-de-knop vermenigvuldigen en verspreiden. Hierdoor verandert de procedure van acceptatie van een artikel. Experimenten in diverse vakgebieden geven hier sterk verschillende reacties. In formele vakken is immers “een som” goed of fout en dan maakt het niet zoveel uit hoe de exacte procedure van uiteindelijke publicatie is, in minder exacte gevallen als geneeskunde is een validatieprocedure essentieel voor de kredietwaardigheid (en volksgezondheid).
- 2) In alle vakgebieden blijft de roep om serieuze Peer-review procedures bestaan. De discussies over de te voeren vormen is in volle gang en sterk vakgebied afhankelijk (Wel2001). Veel experimenten worden besproken op conferenties over elektronisch publiceren (CCS2001, ISC1969, ISC2000).
- 3) Financiers van onderzoek blijven voorlopig zoeken naar schaalbare en “neutrale” indicatoren voor productiviteit en effectiviteit van onderzoekers in de vorm van bibliometrische indicatoren. De naam van een uitgever en tijdschrift, met de daaraan gehechte kwaliteitsperceptie blijft daarbij waardevol.
- 4) In het papieren tijdperk was het uitgeefarchief verspreid over een aantal bibliotheken met overlappende collecties. Het contact met de eindgebruiker liep geheel via de bibliotheek. Feedback naar de uitgeverij was indirect door het abonnementsgedrag van de bibliotheken. Voor het eerst in de geschiedenis zien uitgevers nu een directe band met haar eindgebruikers mogelijk worden. Het gebruik van het uitgegeven materiaal kan nu beter gevolgd worden. Het verspreiden van elektronische versies van artikelen dwingt de uitgevers een archief aan te houden. Immers een uitgevers web-pagina heeft nauwelijks nut als alleen de laatste informatie gegeven wordt. Wij krijgen hierdoor de nieuwe situatie waarin uitgevers gebruikersgegevens hebben, onafhankelijk van de bibliotheek.
- 5) De vraag is echter of deze verschuiving ook gaat betekenen dat de uitgever de collectie/archieffunctie zal gaan krijgen. Hoe groot de uitgever ook kan zijn, haar fonds blijft immer een deelverzameling van de noodzakelijk te ontsluiten en te archiveren collectie.
- 6) In de digitale bibliotheek probeert de bibliotheek voor haar gebruikers een zo volledig mogelijk aanbod te realiseren. Dit betekent dat de digitale bibliotheek een verzameling wordt van (deel) archieven van de uitgevers. In een elektronische omgeving wordt de mogelijkheid geboren om daadwerkelijk synchroon meerdere (uitgevers) bestanden te doorzoeken. Aan de ene kant wordt de bibliotheek dus afhankelijker van de uitgever doordat de artikelen meestal op de servers van de uitgever blijven, aan de andere kant wordt de uitgever ook afhankelijker van de bibliotheek omdat zij alleen via de bibliotheek eindgebruikers kan bereiken. Het is technisch-administratief onmogelijk om miljoenen individuele abonnementen/licenties bij te houden.
- 7) De eigen collectievorming van een bibliotheek kan sneller en effectiever worden bijgesteld als beter begrepen wordt hoe het materiaal wordt gebruikt. De vraag doet zich dan voor wat de verhouding wordt tussen bewuste collectievorming en -ontsluiting voor een bepaalde gebruikersgroep en het formeel beschikbaarstellen -als doorgeefluik- van de wereldliteratuur. De bibliotheek wil een optimalisering voor haar gebruikersgroep(en).
- 8) Bibliotheken ontwikkelen zich als digitale productiecentra voor de eigen locale productie.

9) Met het *en gros* elektronisch beschikbaar hebben van haar productie is het voor de uitgever mogelijk en zeer gewenst om deze productie zo simpel en ruim mogelijk aan de eindgebruiker beschikbaar te stellen. De uitgever tendert dus naar een liefst zo breedmogelijke licentieovereenkomst met de bibliotheekorganisaties.

10) Het digitaal beschikbaar stellen van artikelen dwingt tot standaardisatie. Idiosyncratische systemen worden óf uitgesloten óf tot norm verheven (zoals MSword). De noodzaak tot langdurig gebruik en interactiviteit tussen verschillende systemen en collecties stuurt het proces eenduidig in de richting van *open source* standaards zoals XML. Een eerste succes van deze tendens is de komst van DOI en Crossref, waarbij de literatuurreferenties van artikelen industriewijd gekoppeld kunnen worden. De flankerende ontwikkeling van de *Open Archive Initiative* sluit hier naadloos op aan.

11) Binnen vakgebieden gaat het publiceergedrag veranderen. Dit is in het bijzonder het geval met vakken waar veel gebruik wordt gemaakt van databanken (zoals erfelijkheidsleer) en of multimediale-componenten (zoals werktuigbouwkunde).

III) Elementen voor onderzoek

III-1) De schrijvende onderzoeker

1.1) Primair in het hele proces is de ontwikkeling van de wetenschap en de vrije uitwisseling van kennis.

1.1.1) Hierbij gaan wij uit dat de producerende wetenschapper onder alle omstandigheden het intellectueel eigendom van het werk houdt (*droite morale*), terwijl de reproductierechten in een licentie- of verkoopsvorm overdraagbaar zijn. Opgemerkt moet worden dat dit geen algemeen geldende opmerking is omdat de verschillende auteursrechttradities nogal afwijken.

Binnen de *Association Littéraire et Artistique Internationale (ALAI)* is er een initiatief voor een Internationale Auteursrecht wetgeving (zie: www.alai.org). Op dit buitengewoon interessante, maar zeer gecompliceerde aspect, wordt in de rest van dit rapport niet verder ingegaan.

1.2) De publicatie van de auteur moet dusdanig zijn dat de leesbaarheid voor eeuwen gegarandeerd blijft. Dit betekent dat tot in hoge mate de eisen die voortkomen uit de discussie over digitale preservatie van het cultureel erfgoed a priori moeten worden aangesproken.

1.3) Op dit moment hebben we nog te maken met het elektronisch aanmaken van essentieel papieren artikelen en zijn de elektronische publicaties vaak nog essentieel hetzelfde als de papieren publicatie. Reeds nu zien we een tweedeling van enerzijds het papieren product of haar elektronische vermomming (in Postscript of PDF) volledig conform het papieren paradigma en anderzijds haar werkelijke elektronische vorm (in HTML) dat voorzichtig functionele uitbreidingen biedt. Voornamelijk op het gebied van logistiek (interne en externe hyper-links), kleur en geluid, zien wij dat het traditionele artikel verrijkt wordt. Voor de auteur betekent dit een verandering van schrijfwijze.

1.4) Op kort termijn zien wij dat het artikel als zodanig zal veranderen en de elektronische vorm de primaire publicatie wordt (dit geldt al voor een blad als de *Astrophysical Journal*: <http://www.journals.uchicago.edu/ApJ>). De elektronische versie is omvattender dan de papieren versie in de zin dat daar kleur, simulaties, linking en integratie met databanken gemeengoed wordt.

Niet-tekstuele elementen worden steeds belangrijker (bijvoorbeeld voor vakken waar veel data verwerkt worden of bouwtekeningen/ontwerpen een integraal onderdeel van het werk uitmaken). We denken dan aan:

- i- presentatie van fenomenologische bevindingen (getallen, foto's, geluiden, enz.)
- ii- illustraties van een idee/concept en
- iii- simulaties en modelbouw.

De papieren versie wordt dan een projectie/ samenvatting van de primaire versie die eigenstandig minimaal moet voldoen aan de traditionele eisen voor een correct wetenschappelijk artikel.

Conclusie 1:

In het licht van de onafwendbare ontwikkeling dat de elektronische versie van een wetenschappelijke rapportage de primaire bron van kennisoverdracht wordt, is het essentieel dat de productieomgeving van de auteur aan minimale eisen voldoet. De *instructies voor auteurs* in het digitale tijdperk zijn veelomvattend en verrijkend. Binnen de universiteiten

zien we de ontwikkeling van afdelingssecretaresse naar zelftikkende auteurs, naar digitale productiecentra. Het ligt voor de hand dat deze productiecentra de technologische eisen die digitaal uitgeven stelt t.a.v. portabiliteit, transparantie en standaardisatie moeten volgen, nakomen en gedeeltelijk ook gezamenlijk ontwikkelen

. Deze moeten nauw aansluiten op de ontwikkelingen in de uitgeeffunctie en de bibliotheekfunctie. Het ontwikkelen van een technologische infrastructuur ter garantie van de eeuwigdurende leesbaarheid (gebruik) van wetenschappelijk materiaal wordt dus een taak voor de uitgeef- en de bibliotheekfunctie gezamenlijk. De ontwikkeling van -voor de auteur "plezierig te gebruiken"- sjablonen is hiervoor belangrijk.

III-2) De lezende onderzoeker

2.1) De lezende onderzoeker zoekt pertinente informatie, bladert, leest diagonaal of leest een artikel volledig. Dit gevarieerd zoek-leesproces zal een enorme verrijking ondergaan in de overgang naar elektronische artikelen.

2.1.1) Bij het zoeken naar pertinente informatie (bv. het smeltpunt van water bij 1 atmosfeer) gaat het om het vinden van wel-gedefinieerde informatie-eenheden. De beschikbare informatie moet daarom zeer goed gestructureerd aangeboden worden. Meta-data is hier het trefwoord. Zoekmachines moeten naast hun intrinsieke methoden, gestuurd kunnen worden door de structurering van de publicatie (dus zoeken naar Water als chemisch trefwoord en niet in het auteursveld van de auteur "van de Water", omdat de stopwoorden "van" en "de" weggefilterd zijn). Het gaat er dus om in het auteursproces en de daaropvolgende veredelingsstappen een helder beeld te krijgen van welke <tagging> noodzakelijk is. Het zoekgedrag van een onderzoeker is een maat voor het daadwerkelijk gebruik van de informatie.

2.1.2) Bij het bladerend zoeken naar algemene informatie gaat het de lezer erom een indruk te krijgen of er informatie van zijn/haar gading is.

Een belangrijk verschil tussen dit soort zoeken en het zoeken naar een pertinent gegeven is dat deze vorm van zoeken er vaak op gericht is om materiaal uit te sluiten. Het is een controle-mechanisme om zeker te zijn niets gemist te hebben. Een "hit" is dus niet altijd een positieve vondst, maar ook het tevreden wegzetten van niet direct relevante informatie. Soms gaat het hier om een vorm van trefwoordengymnastiek maar het kan ook door het systematisch doorlopen van inhoudsopgaven. Het spreekt voor zich dat deze belangrijke vorm van gebruik niet zo eenvoudig te meten is als het ophalen van pertinente informatie.

2.1.3) Traditioneel wordt er formeel wel uitgegaan van het proces dat eerst de titel inhoudsopgave wordt bekeken, dan een samenvatting wordt gelezen en dan eventueel het hele artikel. In dezen moeten we goed begrijpen dat de door de auteur geschreven samenvatting een relatief late innovatie is, voortgekomen uit de beperkte mogelijkheden van de eerste generatie digitale bibliografische bestanden. In de praktijk van de bibliotheek bladert de potentiële lezer eerst door een tijdschriftnummer heen, als iets zijn/haar oog treft kan er besloten worden even de abstract te lezen. Nu de digitale bibliotheek het artikel direct op het scherm zet, is het de vraag of de *abstract* nog wel haar rol als samenvatting zal blijven vervullen en niet veel meer de rol van oriëntatie draaipunt zal gaan innemen. De ontwikkeling naar z.g. *extended abstracts* is hier een eerste stap in. Voor een meer fundamentele aanpak zie: Tol2000, 2001.

2.1.4) Met het beschikbaar komen van elektronische versies uit verschillende bronnen wordt het synchroon doorzoeken van meerdere bestanden mogelijk. Binnen zogenaamde *Protals* zal dit vorm moeten krijgen. Ontwikkelingen in *Wrapper* technologie en Z39.50 zijn hier voorwaardelijk.

2.2) Vaak worden artikelen alleen gebruikt om als startpunt te dienen voor verder literatuuronderzoek. Het artikel is dus geen eindbestemming maar slechts een transmissiemedium. In een elektronische omgeving is het natuurlijk mogelijk te volgen in hoeverre een artikel slechts gebruikt wordt als bron voor verdere literatuurnaspeuring, of als primaire bron van informatie.

2.3) In een elektronisch artikel komen er nogal wat mogelijkheden tot lezen bij. Te denken valt aan het volgen van hyperlinksporen, het inspecteren van niet-tekstueel materiaal (intensief plaatjes kijken en manipuleren) en het “lezen” door te “spelen” met simulaties (bijvoorbeeld in modellen die laten zien waarom bij 1 atmosfeer water bij min 9 graden nog niet bevriest).

Conclusie 2:

Het zoekgedrag van een lezer wordt niet alleen bepaald door de aangeboden mogelijkheden, maar fundamenteel door de manier waarop een wetenschappelijke vraagstelling beantwoord zou kunnen worden. Dit is duidelijk zowel sterk vakspecifiek als dat het afhangt van de fase waarin het onderzoek zich bevindt.

Kennis over zoekgedrag zal leiden tot:

- i- inzicht in het daadwerkelijk gebruik van een artikel en
- ii- in de noodzakelijke pre-coördinatie, c.q. structurering van het elektronische artikel als informatieobject.
- iii- inzicht in het samenstellen van deelcollecties, al dan niet via eindgebruiker gestuurde *Portal Sites*.

Het is dus zaak om de “de daad van het lezen” goed te definiëren. Hoe verhoudt zich het snel doorkijken van een titellijst, het doorlezen van een gedeelte van een artikel, het doorlopen van de literatuurreferenties zich tot het min-of-meer lineair doorlezen van het gehele artikel?

Verder is het van groot belang om de hyperlink-sporen van de lezer te volgen om daarmee na te gaan hoe een lezer:

- iv- naar een van tevoren bepaald doel komt (laten we het doelgericht zoeken noemen) en
- v- interdisciplinair door de literatuur heen wandelt (laten we dat zwervend/ juttend zoeken noemen).
- vi- van informatieobject tot informatieobject gaat, z.g. pad-analyse.

III-3) De veranderende rol van de uitgeeffunctie

In hoofdstuk I is al gezegd dat de uitgeeffunctie de veredelingslag is tussen de aanlevering van de auteur en de gevalideerde, gecertificeerde en geauthentiseerde versie die als nieuwe invoer dient voor een het lopende onderzoeksproces.

Reeds opgemerkt is dat een elektronische omgeving inhoudt dat fasen in elkaar gaan schuiven. Er vindt een integratie plaats van auteursgereedschap (gestandaardiseerde regels en schablonen), de structurering van het digitaal geproduceerde artikel (de Document Type Definition (DTD) van het SGML/XML aangemaakte artikel), de transparante doorkoppeling van literatuurreferenties (hyperlinking via Crossref) en met het naast elkaar beschikbaar hebben van recent gepubliceerde en ooit eerder gepubliceerde artikelen.

Daarnaast wordt de elektronische vorm van het artikel langzaam maar zeker de primaire vorm waardoor er een zeer duidelijk redactioneel beleid gemaakt moet worden waarin vastgelegd wordt hoe deze primaire elektronische vorm zich verhoudt tot de papieren

uitdrukking (projectie) van hetzelfde werk. De papieren vorm die nodig blijft om rustig ergens te kunnen lezen wordt in feite een auteur gecertificeerde bewerking van het eigenlijke artikel. In auteursrechtelijke zin zou er zelfs sprake kunnen zijn van twee zelfstandige maar nauw verbonden werken. Een nieuw fenomeen, omdat de papieren versie eenduidig verbonden is met de elektronische (en vice versa) en niet een aparte bewerking of interpretatie is.

Steeds meer wordt verwezen naar pre-print versies van artikelen. Het is van het grootste belang dat het voor de lezer duidelijk is dat de pre-print versie (die al dan niet meer aanwezig is) gelijk is aan de “officiële” versie. Er kunnen immers op het laatste moment nog belangrijke correcties zijn doorgevoerd. We komen dan op het moeilijke gebied van versiebeheer.

Conclusie 3:

Het product van de auteur zal verschuiven van papier naar een rijkere multi-mediale presentatie, waardoor er nieuwe redactionele processen aan de orde komen. In het bijzonder voor Peer-review zal dit nieuwe vergezichten geven. Peer-review is thans immers puur tekstgericht en er zijn geen procedures om gepubliceerde data, beeldmateriaal, multimedia componenten of modellen en simulaties te beoordelen.

Voor de lezer zullen er verschillende afwijkende versies zijn die eenduidig met elkaar verbonden moeten zijn. In de mate waarin de uitgeeffunctie aan productontwikkeling doet, zullen er duidelijk discipline-afhankelijke verschillen zijn, niet alleen wat betreft de vorm van het wetenschappelijk artikel, maar ook wat betreft de manier waarop het gebruikt, gelezen en met andere artikelen doorverbonden wordt.

Voor de organisatie van de uitgeeffunctie betekent een en ander dat de uitgeverij voor de door haar verzorgde publicaties een duidelijke bibliotheekfunctie zal gaan vervullen. De vraag is of dit een bibliotheekfunctie is die zich richt op individuele eindgebruikers en/of op lokale aggregatoren in de vorm van Universiteitsbibliotheken en hun consortia.

III-4) De veranderende rol van de bibliotheekfunctie

De universiteitsbibliotheek als hoeder van dat deel van het wetenschappelijk en cultureel erfgoed dat de onderzoeksgroepen op haar universiteit nodig hebben voor de uitoefening van hun werk verandert ook sterk. In dit rapport wordt het archief aspect even buiten beschouwing gelaten en gaan we alleen uit van de bibliotheek als verzamelplaats van materiaal dat voor bestaande gebruikersgroepen mogelijkwijs van belang is. D.w.z. dat het archief van ooit-mogelijk-weer-eens-van-belang-zijnd materiaal niet meegenomen wordt. De collectievorming voor de universitaire doelgroepen in kwestie is nooit een eenvoudige zaak geweest. Er zijn enorme aantallen studies verricht waarin geprobeerd wordt inzicht te krijgen in de dynamiek van gebruikersbehoeften. Veel harde richtlijnen of conclusies kwamen daar niet uit. In een elektronische omgeving verandert er veel. Ook nu zien we in de vakliteratuur veel gebruikersonderzoeken. Helaas zijn die meestal van een erg lokaal karakter en door de snelle veranderingen van de technologie vaak niet meer dan zeer tijdsgebonden *snapshots*. Door de natuurlijke ontwikkeling waarbij uitgeverijen vanuit hun productiesysteem vanzelf tot archiefvorming overgaan, is het dus de vraag hoe bibliotheken als diensteninstelling voor een lokale gebruikersgroep daar gebruik van gaan maken. Zeker nu de bibliotheken onderling steeds transparanter worden en een onderzoeker kan zoeken in het gezamenlijke landelijk bibliotheekbestand, is het de vraag in hoeverre nationale (en eventueel zelfs internationale) regelingen tussen uitgevers en bibliotheken zich ontwikkelen. Het beschikbaar hebben van alle tijdschriften van alle uitgevers is op landelijk niveau een

goed streven om zorg te dragen voor een optimale potentiële capaciteit dat onderzoekers alle relevante werken vinden. Als dit een bereikbaar streven is en daarbij alle bibliotheken via netwerken verbonden zijn, is de volgende vraag natuurlijk in hoeverre het zinvol is om op lokaal niveau deze administratieve en technische organisatie te dupliceren.

Een simpel ideaalbeeld van een systeem van nationale academische licenties vereist echter een totale verandering van de structuur van de financiering en organisatie van het academisch bibliotheekbedrijf. Dit is een complexe materie die niet onmiddellijk de zorg van dit project is maar zich wel als een consequentie van haar activiteiten aandient. Een abonnements/licentie model moet immers een eenduidige kostenplaatsmodel hebben binnen de universiteiten. Voorwaarde voor de discussie is natuurlijk goed te weten hoe materiaal gebruikt wordt.

In feite kunnen wij twee aspecten onderscheiden.

4. 1) de formele beschikbaarheid van elektronische tijdschriften en
4. 2) de lokale wijze waarop die in gebruikersgroep gebonden eenheden, ontsluitbaar beschikbaar worden gesteld.

ad. 4.1: Dit aspect induceert de discussie over de rol die de UKB en haar deelnemende bibliotheken zich in de toekomst toedenken. De huidige licentie overeenkomst tussen de UKB en ES laat duidelijk zien waar de schoen wringt. Als op basis van de bestaande collectie betaald wordt, krijgt de ene universiteit (virtueel) meer voor haar geld dan een andere. Op het moment dat iedereen toegang tot alles heeft gaat natuurlijk het begrip zwaartepuntsbibliotheek ook schuiven. Eén model zou kunnen zijn dat op basis van een inzicht in de lokale en nationale behoeften een deel van het aangeboden pakket als een landelijke dienstverlening ter beschikking wordt gesteld, aangezien het overal wel ergens een keer gebruikt wordt, terwijl een specifiek deel slechts lokaal beschikbaar komt omdat het slechts plaatsgebonden gebruikt wordt of om reden dat het ergens in Nederland beschikbaar zou moeten zijn (bijvoorbeeld belangwekkende tijdschriften in voor de gemiddelde Nederlandse onderzoeker moeilijk te lezen talen). Het economisch model dat dit moet onderbouwen zal sterk afhankelijk zijn van inzicht in het daadwerkelijk gebruik van het aangeboden materiaal.

ad. 4.2: Op het niveau van de dienstverlening aan een lokale goed bepaalde gebruikersgroep, is het evident dat de ontsluiting toegesneden kan worden naar de gebruikersgroep. Hoewel uitgevers ook experimenteren met zogenaamde *Portal Sites* (in feite zijn dat domein specifieke bibliotheken met extra diensten) kan het nooit zo zijn dat deze eenvoudig scherp toe te snijden zijn naar de specifieke collecties die lokale onderzoeksgroepen hebben. Immers, ook al werken deze groepen samen met andere groepen op nationaal of internationaal niveau, hun actie-radius zal altijd beduidend kleiner zijn dan welke uitgevers *portal* dan ook. Een probleem hierbij is dat er nog geen goede software is om tussen verschillende databases te zoeken. De zogenaamde “functionaliteit” van de bibliotheek *Portal* is niet die van de uitgevers web-site (zie ook VI-2).

IV) Gebruikersonderzoek

In de gevoerde gesprekken (zie bijlage B3 met de namen van alle gesprekspartners) komt duidelijk naar voren dat er geen duidelijke doelen gedefinieerd worden. Betere en betrouwbare informatie, waar mogelijk, is natuurlijk de eerste eis, maar wat daar dan mee gedaan zou moeten kunnen worden is moeilijk te articuleren. De ene UB kent een systeem waarbij alleen naar het hele universitaire gebruik wordt gekeken, de andere UB wordt geconfronteerd met gedetailleerde doorberekeningen naar faculteiten en/of instituten, groepen, enz. Over de definitie van de uiteindelijke gebruikersgroep bestaat geen eenheid.

In het papieren tijdperk zijn er veel gebruikersstudies gehouden. Gezien de intrinsieke beperktheid van de mogelijkheden is daar nooit veel beleidsmatig uitgekomen.

In een elektronische omgeving wordt het mogelijk om werkelijk goede methoden te ontwikkelen om het gebruik van een bron (artikel of tijdschrift) te meten in relatie met het gedrag van verschillende (soorten) gebruikers.

In het huidige tijdsgewricht waarbij de technologische mogelijkheden nog ver vóór liggen op de dagelijkse praktijk van de modale onderzoeker zijn de resultaten van enkelvoudige gebruikersonderzoeken bij voorbaat echter achterhaald bij publicatie. Een serieuze Delphi-studie naar de toekomstige ontwikkeling van elektronische tijdschriften laat zien dat het nog alle kanten op kan gaan en er geen eenduidige richting voorspeld kan worden (Kel2001).

Zowel de vorm waarin de elektronische versies van papieren artikelen zich ontwikkelen richting primair product als de computationele infrastructuur als de netwerkbandbreedte zijn geen van alle stabiele parameters voor gebruikersonderzoek.

Het is daarom een onontkoombare conclusie dat gebruikersonderzoek in deze periode intrinsiek een longitudinaal karakter moet hebben.

Desalniettemin zijn er wel degelijk onderzoeken aan te wijzen waarbij beide partijen veel baat kunnen hebben en waarmee de discussie verder gebracht kan worden. In het volgende hoofdstuk wordt hier dieper op ingegaan.

Binnen Nederland is er weinig ervaring met gebruikersonderzoek op het gebied van (elektronische) literatuurvoorziening (Til1999). In het buitenland liepen en lopen er enkele grote projecten, waarbij de vraag is in hoeverre de uitkomsten van die projecten niet ook toepasbaar zullen zijn op de Nederlandse situatie (voor een kort overzicht zie: Bijlage B1 Internationale projecten).

V) Onderzoeksmogelijkheden

V-1) De auteursomgeving

Zoals al in de inleiding is gesteld, ontwikkelen zich op de universiteiten digitale productiecentra. Deze centra zijn vaak bedoeld om de eigen in-huis productie van collegedictaten, proefschriften, voordrachten, maar ook experimenten met (pre-print)-servers, of eigen elektronische tijdschriften vorm te geven. Aan de andere kant investeren de uitgevers veel tijd en geld aan het voortdurend bijwerken van instructies voor auteurs en de permanente confrontatie met de harde werkelijkheid dat de eerste veredelingsstap in het proces, de vertaling van het aangeboden artikel in een gestandaardiseerde vorm is.

In de ontwikkeling naar serieuze auteursomgevingen waarbij de auteur actief geholpen kan worden materiaal zo aan te leveren dat het ook eenvoudig in een gestandaardiseerde database vorm gegoten kan worden is veel werk te doen.

Het feit dat zowel Wolters-Kluwer als Elsevier Science een belangrijk deel van haar digitale productieomgeving vanuit Nederland aanstuurt, biedt een mogelijkheid tot samenwerking.

1.1) Dit kan gaan op fundamentele niveau van standaardisatie discussies (XML-familie (met name DTD's), Open Archive, Z39.50 (en International next generation: ZING), enz. Een belangrijk rapport hier is de zeer solide studie *E-Journal Archive DTD Feasibility Study*, die in opdracht van de Harvard UB is uitgevoerd om de diverse DTD's te vergelijken (Ine2001). Op basis van duidelijke standaardisatie-overeenkomsten kunnen dan productspecifieke (d.w.z. naar vakgebied en/of publicatievorm) tekstverwerkersjablonen worden aangemaakt die een vloeiende overgang van auteur naar publicatieomgeving kunnen garanderen.

1.2) Ook de gezamenlijke ontwikkeling voor (specificaties van) auteurshulpmiddelen zoals multilinguale woordenboeken en vakspecifieke spellingcheckers bieden een mogelijkheid tot gezamenlijke activiteiten.

Voor multilinguale dictionaries zie bijvoorbeeld het EU project voor medische termen in de 9 EU talen (<http://allserv.rug.ac.be/~rvdstich/eugloss/welcome.html>).

Voor specifieke spellingcheckers bijvoorbeeld de firma Spellex (<http://spellex.com/>), die spelcontrole programma's heeft op het gebied van o.m. geneeskunde, tandheelkunde, rechten, biotechnologie, en farmacie.

Een belangrijke uitbreiding op deze al bestaande ontwikkelingen is te geven als een spelcontroleprogramma direct gebruikt zou kunnen worden als indexeringsmachine. Op dit moment controleert zo'n programma alleen of iets klopt en gaat dan verder. Wat zou er niet natuurlijker zijn dan dat zo'n programma ook bij een -nader vast te leggen verzameling woorden- deze woorden direct voorziet van een code (<tagging>) waarmee meteen in de tekst een vakspecifieke indextermcode wordt aangebracht?

Voorbeeld: het programma vindt in een chemische tekst de (al dan niet correct gespelde) term *nucleophilic attack*, geeft nog als keuzemogelijkheid in een pop-up- window de antonieme term *electrophilic attack* en codeert in de tekst deze term meteen met een verwijzing naar een scheikunde theasaurus. Op deze wijze is het mogelijk de "search-hits" van *Bomb-Attacks*, *Heart-Attacks*, of de aanval van een voorhoedespeler bij voetbal, van het gebied chemische binding uit te sluiten.

In de ontwikkeling van dergelijke technologische auteurshulpmiddelen kunnen IWI (Surf), en de universitaire- en uitgeversproductiecentra prima samen werken.

V-2) De lezersomgeving en de ontsluiting van informatie

Over het proces waarom een lezer besluit om een werk ook daadwerkelijk min-of-meer volledig te lezen is het nodig onderzoek gedaan. In de context van dit rapport gaat het erom de vraag te articuleren hoe de diverse vormen van lezen gefaciliteerd worden.

2.1) Biedt een web-site een inhoudsopgave aan met als secundaire mogelijkheid een zoekoptie, dan zal de lezer merendeels bladerend zijn/haar weg zoeken. Het doorlopen van inhoudsopgaven is echter een continu proces geworden in plaats van een stuksgewijs proces (nummer van de plank-bekijken-terugleggen-nieuw nummer van de plank). Bij elektronisch afzoekbare inhoudsopgaven is het tijdschriftnummer echter nog steeds de basis (dit leidt dus tot klik op issue#, scan inhoud, klik op back, klik op volgend issue, scan, enz.; net papier dus!). Voor een elektronisch systeem is het tijdschriftnummer echter oninteressant geworden terwijl de paginering als gestandaardiseerde maat voor de voortschrijding van een betoog een nuttig bibliografisch hulpmiddel kan blijven. Onderzoek naar de wijze waarop inhoudsopgaven aangeboden worden in een elektronische omgeving verdient dan ook de aandacht.

2.2) Met het algemeen worden van SGML/XML gestructureerde documenten is het evident dat informatie-ontsluiting een nieuwe fase in gaat. Er kan nu -in principe- gezocht worden op trefwoorden, zowel als op vrije woorden, in elk specifiek onderdeel van een document. De vraag naar artikelen van de AUteur= Water, van der (Nederlandse achternaam), AND het trefwoord=Zwaar Water in de lopende tekst AND verwijzend naar een publicatie van REF-Au= Water (Engelse achternaam) zal nu iets serieus opleveren. Zoekmachine-interfaces die ook gebruikmaken van de structurering van artikelen zijn hard nodig en zullen een geweldige hoeveelheid inzicht geven in het daadwerkelijk gebruik van informatie. Ook hier ligt het voor de hand dat de ontwikkelingen op Universiteiten op het gebied van zoekmachines, XML en bijvoorbeeld de ontwikkeling van Elsevier Science' Zoekmachine Scirus in elkaar verlengde liggen. Een voorstel is verwoord in VI-2.

V-3) De meting van gebruik en wat gebruik is

Op basis van een goede logboek (*logfile*) analyse moet het mogelijk zijn de verschillende manieren van gebruik nader te analyseren. Voorwaarde daartoe is dat er een overeenstemming bestaat over wat er nu wel en niet uit een logboekbestand te distilleren is. Op internationaal niveau (zie: Bijlage B1: Internationale initiatieven) wordt hard aan recommandaties voor een minimum standaard gewerkt. Het spreekt voor zich dat deze recommandaties als uitgangspunt worden genomen. Onderzoek leert echter dat het nog niet zo eenduidig is wat een *logfile* weergeeft. Een studie van Borghuis toont bijvoorbeeld aan dat er binnen een relatief korte tijdspanne (tot 1 minuut) onduidelijkheid kan zijn of meerdere "klikken" ook wel meerdere leesacties (*downloads*) vertegenwoordigen (Bor2000). Voor direct bestuurlijk gebruik zijn minimale series gebruikersinformatie nodig die direct waarde hebben voor de sturing van de uitgeef- en bibliotheekprocessen. Zie verder VI-2 & B2.

Op dit moment staan de zogenaamde Digitale Rechten Beheersystemen (*Digital Rights Management systems DRM*) sterk in de belangstelling. Hier wordt informatie zo gecodeerd dat de rechten min-of-meer aan de hardware via licenties zijn vastgebakken (met name Microsoft en de muziekindustrie maken zich hier hard voor). Afgezien nog van de juridische implicaties zou een DRM systeem wel een prima hulpmiddel voor uitgebreide gebruikerstesten kunnen zijn. Concreet wordt een voorstel gedaan dor de KUB, zie VII-1.

V-4) Domein specifieke kenmerken van informatiebehandeling

4.1) Met de ontwikkeling van elektronisch publiceren verandert er veel op het algemene vlak van kennisoverdracht. De wetenschappelijke revolutie ging hand in hand met de boekdrukkunst. Kennis werd in taal eenduidig weergegeven en identiek verveelvoudigd. Iedereen kan nu, onafhankelijk van tijd en plaats dezelfde kennis tot zich nemen en interpreteren. De reformatie stelde dat de waarheid in de schrift ligt en niet in de gesproken persoonlijk interpretatie van de priester.

De laatste eeuwen laat zien dat kennisoverdracht een papieren aangelegenheid is met hier en daar een verhelderend plaatje. In de monomediale opslag van nullen en éenen kunnen we nu eenvoudig tot multimediale presentaties komen, waarbij kennis op verschillende wijze wordt gepresenteerd. Dit is sterk vakafhankelijk. In sterk talige vakken (zoals wiskunde) is er weinig multimediaal vertoon. De website van de ACM (Association for Computing Machinery, www.acm.org) kent geen frivoliteiten, hetzelfde is het geval met de pre-print servers van de Los Alamos cluster ([http](http://)).

Daarin tegen wordt in vakken waar de primaire fenomenologische data essentieel zijn voor het begrijpen van het wetenschappelijk onderzoek enorm vaak gebruik gemaakt van niet tekstuele elementen. Binnen de CCSC-conferentie was hier een hele sessie aan gewijd. Archeologie, scheikunde, astronomie, geneeskunde, enz. zijn allemaal vakken waar grote hoeveelheden getallen, beeldmateriaal, film en binnenkort ook geluid als essentiële onderdelen van het reguliere wetenschappelijk artikel zullen gaan behoren (zie CCS2001 voor voorbeelden en referenties). Dit betekent dat vakafhankelijk onderzocht moet worden hoe de trend zich verder zal gaan ontwikkelen en hoe niet-talige informatie als primaire eenheden van primaire wetenschappelijke kennisoverdracht gebruikt worden. Zowel voor de uitgevers die hun waarde toevoegende werkzaamheden tot deze informatie zullen moeten uitbreiden, als voor de bibliotheken die dit soort informatie transparant en permanent ontsluitbaar moeten hebben, is hier nog een flinke kluit.

4-2) Dichter bij huis is er al het bekende verschil in gebruikersgedrag (Til1999) tussen de wetenschappen waar veeleer naar pertinente informatie wordt gezocht (positief zoeken naar een omschreven vraag) en de wetenschappen waar veel meer gebladerd en “dwalend” gezocht wordt. Vanuit een goed opgezet gebruikersonderzoek waarbij niet de gemiddelden (per bibliotheek, IP-cluster) maar de individuele wetenschapper gevolgd wordt, moet het mogelijk zijn om een betere greep te krijgen over het “nut” of de “waarde” van een artikel. Dit is wat anders dan traditioneel citatie onderzoek (bibliometrie): in dat soort onderzoek wordt immers alleen gekeken naar de referenties van de *schrijvende* wetenschapper. In het volgen van de *zoekende* lezer leren we ook hoeveel informatie nodig is om tot het besluit te komen wel of niet een artikel als nuttige referentie op te nemen in een nieuwe publicatie. Immers een artikel waarin uitgelegd wordt dat een bepaalde weg heilloos is, zal veel gebruikt worden maar nooit geciteerd. De kennis van wel gelezen/bekeken maar niet gerefereerde informatie is van essentieel belang voor een serieuze collectievorming.

Belangrijk is op te merken dat men bij bibliometrische analyse uitgaat van een bekende auteur (waar dus ook extra sociologische informatie over verkregen kan worden), terwijl in het gebruikersonderzoek men meestal te maken heeft met een anonieme lezer. Het is dus nodig dat bij gebruikersonderzoek experimenten worden gedaan waar de gebruikers bekend zijn aan de onderzoekers.

V-5) Kennisdiffusie en pad-analyse

Op dit moment staan de logboekbestanden slecht toe te meten hoe de zoektocht van een wetenschapper is. Dat er een artikel wordt “gedownload” is één zaak, hoe de lezer tot die daad komt en hoe hij/zij verder gaat een geheel andere. Het liefst wil men weten hoe de lezer de verschillende artikelen aan elkaar rijgt. Hoe er van een *Abstract* of Inhoudsopgave naar een artikel wordt gegaan en hoe men van daar (via de referenties?, de auteursnaam?, de tijdschrifttitel?) verder gaat. Dit betekent dat er koppelingen zouden moeten komen tussen de verschillende logboeksystemen (van de uitgevers en de lokale bibliothecaire), op basis waarop volgsystemen kunnen worden ontwikkeld. Voorwaarde tot het ontwikkelen van dat soort systemen is allereerst dat er een standaardisatie van de logboeksystemen komt, of op zijn minst een transparantie en standaardisatie van een minimale verzameling metadata.

V-6) Niet bibliografische links

Via DOI/Crossref of via de Open URL standaarden kunnen we binnenkort naadloos van het ene artikel naar het andere komen. Toch is dit maar een onderdeel van het hele bouwwerk van hyper-links. Steeds meer worden artikelen veredeld met referentiele links naar niet gecertificeerd bibliografisch materiaal. We denken aan: verwijzingen naar websites van organisaties, naar homepages, naar databanken, naar dictionaires en encyclopedieën, naar grijze literatuur, enz., enz. Deze links zijn in de HTML versies te vinden en meestal niet in de PDF versies. Vaak hebben zij geen standaard vorm. Een prachtig, extreem, voorbeeld is het Perseus project (<http://www.perseus.org>) waarbij klassieke teksten on-line worden gebracht. Bijna ieder woord in de Latijnse of Griekse teksten zijn “gehyperlinked” naar aparte pagina’s waar hun vertaling, gebruik, enz., enz. weergegeven zijn. Ook in de scheikunde en de farmacologie zijn dergelijke niet strikt bibliografische links sterk in opkomst. Het in kaart brengen van het gebruik van deze links zal een geheel nieuwe kijk geven op de wijze waarop informatie vergaard wordt.

VI) Aan te vatten projecten

Op basis van bovenstaande is het duidelijk dat er niet 1-2-3 projecten gestart kunnen worden die hapklare parameters voor economische modellen opleveren. De technische omgeving is nog zeer fluïde en de leercurve voor het gebruik van de elektronische bibliotheek nog niet op een stabiel plateau aangekomen.

Het verdient dan ook aanbeveling de beschikbare gelden op een zorgvuldige en getemporeerde wijze aan te wenden om zo longitudinale studies eerst voor te bereiden en dan te laten uitvoeren.

Een aantal aan te vatten projecten zijn in het vorige hoofdstuk benoemd.

Binnen de opdracht is gebleken dat er ondanks de ambitie serieus gebruikersonderzoek aan te vatten niet veel concrete plannen zijn. Een belangrijke oorzaak daarvan is het gebrek aan duidelijk inzicht welke kengetallen in welke context tot relevante bestuurlijke/organisatorische maatregelen kunnen leiden. Het gehele gebied is nog in een staat van experimenteren, een fase die eigenlijk net begonnen is.

Het voorstel is daarom over te gaan tot de organisatie van enkele workshops die op verschillende vlakken een overzicht van de mogelijkheden aan zullen leveren. Op basis daarop kunnen dan realistische voorstellen tot verdergaande, meer intensieve, projecten (testen, studies, enz.) worden uitgewerkt.

a) Primair aan het hele proces is het inzicht in de exacte methoden, mogelijkheden, voorwaarde tot en voorstellen tot uitbreiding van logboekonderzoek. Vandaar dat de eerste workshop er een moet zijn van experts op logboekanalyse en statistiek. Een voorstel tot zo'n workshop is hieronder gegeven (VI-1).

b) Een hieraan nauw gerelateerd onderwerp is de vraag welke parameters die in principe uit logboekanalyse te halen zijn ook relevant zijn voor bestuurlijke sturing en economische planning. Immers, alles wat kan is nog niet nodig en alles wat nodig is, kan niet altijd. Een eerste lijst van zulke vragen is gegeven in B2. Het is een discussiepunt in hoeverre dit binnen de bovengenoemde technische workshop moet worden geschoven. Deze vragen worden immers primair bepaald door de verantwoordelijken voor collectiebeheer, de marketing managers van uitgeverijen en de financieel experts van de partijen. Een mogelijkheid is om een paar experts te vragen om een met redenen omklede lijst van te beantwoorden vragen te maken waarbij de relatie van de vragen duidelijk wordt. Het gaat er immers om rapportages te krijgen die hoofdlijnen en trends laten zien (VI-4).

c) In de voetsporen van het rapport *Benchmarking van wetenschappelijke bibliotheken* (Fro2000) kan nagedacht worden over de termen waarbinnen *Benchmarking* voor de digitale uitgeefproducten en hun banken alsmede de ontsluiting via bibliotheken gestalte kan krijgen. Het spreekt voor zich dat de resultaten van bovengenoemde punten 1 en 2 hierbij randvoorwaarde zijn.

d) Op basis van de structurering van de informatie en in nauwe relatie met de discussies en uitkomsten van bovengenoemde workshop VI-1 wordt een workshop voorgesteld over de nieuwe mogelijkheden van informatieontsluiting en haar relatie met een beter begrip over het daadwerkelijk gebruik van de informatie. Een voorstel daartoe is eveneens hieronder gegeven (VI-2).

e) Het is algemeen bekend dat er grote verschillen zijn tussen de verschillende vakgebieden in gebruikersgedrag. Auteursgedreven formele vakgebieden zoals theoretische hoge-energiefysica kunnen prima leven met preprint banken. Het publiceren in een tijdschrift is meer een finale certificatiestap die onontbeerlijk is in het maatschappelijk verkeer. Het gebruik van de gedrukte versie is vaak niet meer dan een formele verwijzing naar een werk dat in pre-print versie reeds gelezen was. In meer lezersgedreven vakken als de letteren of delen van de geneeskunde zijn pre-print banken nauwelijks van betekenis. De redactionele (Peer-reviewed) kwaliteitscontrole is hier essentieel.

Een werkgroep zou kunnen worden samengesteld waarin de termen worden opgesteld waarbinnen een gericht eindgebruikersonderzoek zou moeten worden aangevat. Het moet dan gaan om een vergelijkende studie tussen verschillende vakgebieden, waarbij de eindgebruiker op zijn/haar eigen werkplek gevolgd wordt in zijn/haar literatuur onderzoek. Dit is onderzoek dat op de universiteit moet gebeuren en gezien de privacy-wetgeving met volledige toestemming en medewerking van de betrokkenen. In dit vooronderzoek werd het nut van een dergelijk onderzoek door alle gesprekspartners onderschreven. Concrete voorstellen zijn er echter nauwelijks gedaan. Een uitzondering is de UB-Nijmegen waar men in een voorbereidende fase is voor een gebruikersonderzoek naar van elektronisch beschikbare tijdschriften in de psychologie.

f) Op een nog algemener niveau is het natuurlijk uiterst noodzakelijk dat er een initiatief komt dat een diepgaande discussie entameert over de toekomst van de bibliotheek als collectievormend en literatuur-ontsluitend instituut voor een goed gedefinieerde gebruikersgroep in het licht van 1) een tendens naar grootschalige licentiecontracten met grote uitgeverijen en de kostenomslag voor de deelnemende bibliotheken onderling, 2) de specifieke collectievorming van lokaal noodzakelijke materiaal en unica en 3) de rol als universitair elektronisch productiecentrum.

Tegelijkertijd is aan de uitgeverskant de discussie over haar veranderende rol t.a.v. 1) de productveredeling in een multi-mediaal tijdperk, 2) de rol van het archief en 3) de rol van aggregatoren.

Binnen de kaders van het onderhavige onderzoek is het van groot belang dat er beter inzicht komt welke vragen kunnen leiden tot antwoorden die de bovengenoemde discussie inhoud kan geven. Een eerste uitwerking is gegeven in VI-4, de fundamenteelere discussie over de toekomstige rol van de bibliotheek ligt natuurlijk in de schoot van de UKB.

g) Om gecontroleerd gebruik te meten kunnen wij *Digital Rights Management* technologie als methode van gebruikersonderzoek gebruiken.

DRM technologie staat een exact volgen van een gebruiker toe. Deze gegevens worden dan weer teruggestuurd naar de eigenaar van de rechten. Op dit moment wordt er veel geëxperimenteerd en is er nog geen industriestandaard. Een testomgeving zou kunnen worden ingericht waarbij een geselecteerde groep gebruikers gevolgd worden.

Zaken die daarbij betrokken kunnen worden zijn:

- De DRM architectuur (o.m. OEM versus plug-in) .
- De weerstand van organisaties en gebruikers om dit soort technologie te gebruiken.
- De privacy aspecten voor de gebruiker.
- De mogelijkheid om een ontwerp te maken dat aansluit op het daadwerkelijk gebruik van een gebruiker of groep gebruikers.

Dit technische experiment zou binnen een samenwerking van Surf (IWI) een van de uitgeverijen en een geselecteerde gebruikersgroep op een van de universiteiten kunnen worden uitgevoerd. In de gesprekken bleek zowel ES als KAP grote interesse te hebben voor

experimenten op dit gebied. Vlak voor het afsluiten van dit rapport ontving ik een zeer to-the-point onderzoeksopzet van de UB-Tilburg over DRM. Dit is toegevoegd als VII-1. Een eerste reactie van de bibliotheek geneeskunde van de Universiteit van Maastricht leert dat men daar graag als object van onderzoek zou willen deelnemen.

VI-1) Workshop logboekanalyse en statistiek

Als grondslag van verdere analyse van gebruik is het noodzakelijk overeenstemming te hebben over welke informatie er nu eigenlijk gesproken wordt.

In de huidige situatie hebben wij aan twee kanten analysemogelijkheden om gebruikers (in termen van IP's of Usernames) te monitoren.

- 1) Aan de uitgeverkant waar logfiles worden bijgehouden; voorbeelden zijn die van JSTOR en ScienceDirect.
- 2) Aan de bibliotheekkant waar het gebruik van de lokale servers gemeten kan worden.

De gebruiker blijft echter niet op de server van de lokale bibliotheek maar gaat vandaar naar uitgevers diensten of gaat via het universitaire netwerk ook direct naar uitgeversdiensten.

Om vragen te kunnen beantwoorden of en op welke wijze en in welke mate tijdschriften en hun artikelen worden gebruikt, is het nodig een protocol te hebben op basis waarvan metingen gedaan kunnen worden. Een deel van de metingen zullen aan de bron plaatsvinden, andere zullen echter op de werkplek zelf gemeten moeten worden.

Voorbeelden o.m.

1- Een artikel wordt alleen gebruikt om als bron voor referenties te dienen. Het wordt dus niet gelezen maar alleen de referenties worden doorgeclickt. Dit gebruik is echter niet vast te stellen: een request (of click) haalt de informatie over naar de computer van de gebruiker (of het nu een abstract, Full Text, HTML versie of een PDF is). Wat er daarna meegedaan wordt is niet te meten door de bron. In zo'n geval moet dus de gebruiker zelf gevolgd worden.

2- Een artikel wordt doorgebladerd op het scherm maar niet ook lokaal opgeslagen (één request). Sommige rapporten die in boekvorm te koop zijn hebben de mogelijkheid om slechts per pagina over te halen (meestal HTML of PDF).

3- Een artikel wordt onmiddellijk gedownload zonder eerst bekeken te worden. Ook dit is niet te meten door de bron want een click op de link is een request.

Aan de gebruikerskant is het wel van belang om te weten of een document alleen bekeken wordt en daarna weggegooid of bewaard wordt en mogelijk zelfs daarna lokaal verspreid.

Om gebruikersbewegingen te kunnen analyseren (pad-analyse) is het nodig te weten welke activiteiten te meten zijn, m.a.w. welke kenmerken, patronen en volgordelijke activiteiten zou je uit een logfile willen halen. Dit zal vooral lokaal moeten gebeuren.

Internationale activiteiten: standaardisering en definitie van data-elementen

Op het gebied van logfile analyse wordt er internationaal al veel gedaan.

We noemen:

A recommendation for Content Information: A report by the OCLC/RLG working group on preservation metadata, October 2001. www.oclc.org/research/pmwg

International coalition of library consortia. ICOLC. Guidelines for statistical measures of usage of web-based information resources (December 2001):

www.library.yale.edu/consortia/2001webstats.htm

PALS Usage statistics working group. (The publishers and Libraries Solution group of JISC, the PA, and ALPSP): <http://www3.oup.co.uk/central/pals/index.htm>

Ook hebben wij nu ervaring met de rapportages van bijvoorbeeld JSTOR en ScienceDirect.

De centrale vraag is natuurlijk: welke antwoorden willen we hebben en welke vragen moeten we daarom stellen?

Voorstel

Het voorstel is om een eendaagse workshop te houden met de specialisten van uitgevers en bibliotheken die zich met de statistische analyses van logfiles en verder gebruik bezig houden.

De workshop is voorwaardenscheppend voor de beide andere workshops.

De workshop leidt tot:

- een geaccepteerd kennisniveau over wat usage van elektronische tijdschriften nu eigenlijk is en hoe het zich verhoudt tot gebruiksmetingen in de print-omgeving en tot het citatiegedrag van auteurs.
- heldere definities voor alle data-elementen die in usage-analyse een rol spelen en tevens hoe deze samenhangen met definities van de data-elementen in de print-gebruik records en citatie records.
- een initiële beschrijving van de specifieke informatie-behoefte (zowel bij bibliotheken als bij uitgevers) die door analyse van usage data en daaraan gerelateerde referentie-informatie kunnen worden vervuld.

Deelnemers:

Primair de specialisten van uitgevers en bibliotheken (15 mensen).

Secondair de direct betrokkenen bij collectievorming, marketing, planning, productie (15 mensen).

Duur: 1 dag.

Organisatie:

Drs. Marthijn Borghuis (ES)

Drs. Boy de Haas (UB-UvA)

Drs. Peter van Laarhoven (UB-RUG)

Joost Kircz (KRA) Coördinatie

Deze workshop zou goed met medewerking van de Werkgemeenschap Informatiewetenschap kunnen plaatsvinden.

VI-2) Workshop ontsluiting

De beschikbaarheid van *full-text* digitale tijdschriften houdt een belangrijke verandering in voor zowel gebruikers, bibliotheken, en uitgevers. Omdat de software-omgeving waarbinnen deze informatie wordt aangeboden volstrekt bepalend is voor de gebruiksmogelijkheden die de gebruiker heeft, dienen de aanbieders van deze informatie zich veel sterker dan vroeger te verdiepen in de manieren waarop gebruikers met deze informatie willen omgaan.

Globaal kunnen we op dit moment drie manieren van gebruik onderscheiden:

- 1- Browsen: het bladeren door lijsten relevante tijdschrifttitels, het opvragen van recente nummers, het bladeren door inhoudsopgaven van de nummers op zoek naar interessante nieuwtjes, het opvragen van voldoende uitgebreide informatie om mogelijke interesse bevestigd te zien, enzovoort. Deze manier van gebruik komt overeen met het vroegere bezoek aan de bibliotheek, waar meestal de recente nummers van de tijdschriften waarop de bibliotheek was geabonneerd, op een paar planken bij elkaar stonden en eenvoudig doorgebladerd konden worden.
- 2- Zoeken: het met behulp van zoektermen doorzoeken van de gehele collectie, een vóórgeselecteerd deel van de collectie, een enkel tijdschrift, enzovoort. In de papieren collectie bestond geen echt gelijkwaardig equivalent voor deze retrieval-toepassing. Slechts enkele materiaalkundige tijdschriften hebben een materialenindex, nog minder tijdschriften een serieuze onderwerpindex.
- 3- Persoonlijke attendering: het automatisch (bijvoorbeeld per e-mail) geattendeerd worden op nieuw ontvangen informatie die aan een vastgelegd persoonlijk interesseprofiel voldoet (z.g. Selective Dissemination of Information). Daarbinnen kunnen opnieuw de twee voorgaande gebruiksmogelijkheden worden onderscheiden. Een interesseprofiel kan bestaan uit een aantal tijdschrifttitels, waarvan de inhoudsopgaven worden ontvangen, zodat daarin gebladerd kan worden, op zoek naar interessante nieuwtjes. Een interesseprofiel kan ook bestaan uit een combinatie van zoekwoorden waarop de gehele aanwas van de collectie (of een voor geselecteerd deelgebied daarvan) gefilterd wordt. In de papieren collectie kon door tijdsdruk ingegeven – het was daar een handmatige activiteit – meestal hooguit sprake zijn van eerstgenoemde vorm van attendering, waarbij het “persoonlijke” zelfs vaak nog ontbrak.

Voor elk van deze drie manieren van gebruik – het lijkt aannemelijk dat elke directe wijze van gebruik van digitale tijdschriften wel onder één van deze drie kopjes te vangen zal zijn – is nader onderzoek geboden.

Voor een analyse van het gebruikersgedrag, zowel in de afweging tussen browsen en zoeken, als in de details van het browse-gedrag, komt logboek-analyse als onderzoeksmethode zeker in aanmerking (zie voorstel VI-1). Daarnaast komen ook andere manieren van gebruikersonderzoek, zowel gebaseerd op observatie als op enquêtes, in aanmerking. Dit geldt zeker voor de te bieden gebruiksmogelijkheden voor persoonlijke attendering. Maar ook wat betreft de te bieden retrieval-functionaliteit is nader onderzoek geboden. Het gebruik verandert immers flink met de elektronische beschikbaarstelling van vele duizenden tijdschriften in de UB's. Onderwerpen zijn onder meer:

1. Welke geavanceerde, veelal op taaltechnologische methoden gebaseerde functionaliteit komt in aanmerking om bij deze specifieke informatiesoort te worden toegepast?

2. Welke terugkoppelmechanismen, inclusief het op persoonlijk niveau opslaan van preferenties en eerder zoekgedrag, kunnen bijdragen aan verhoging van precisie en recall van zoekacties door gebruikers?
3. Hoe kunnen precisie en recall van zoekacties worden verbeterd door het toevoegen van gestandaardiseerde metadata?
4. Direct full-tekst zoeken versus reference linking vanuit bibliografische databases.
5. Synchron zoeken in meerdere bestanden tegelijk (cross-file searching).

Enigszins gechargeerd kunnen we stellen dat er een fase is geweest waarbij collectievorming (de bibliotheek) een doel in zich zelf was. Collectievorming was l'art pour l'art, nu zitten we een fase waarbij de bibliotheek zich ontwikkelt als dienst (toegepaste kunst fase), in de toekomst zal de bibliotheek echter ook een actieve/creatieve rol als onderzoeksmotor/atelierfunctie vervullen.

Een collectie zal uiteindelijk dusdanig gestructureerd moeten worden dat op basis van het onderzoeksprofiel van een onderzoeker/onderzoeksgroep actief materiaal gesuggereerd wordt. Dit is een fundamentele stap verder dan de passieve situatie van nu waarbij de bibliotheek wacht op een al dan (meestal) niet wel-geformuleerde zoekvraag. De ontwikkeling van goed toegesneden *portal-sites* is hierbij een kernactiviteit.

In deze workshop gaat het erom de effectiviteit en efficiëntie van ontsluiting beter te benoemen om zo tot strategische richtlijnen voor een vollediger en intensiever gebruik van materiaal te komen.

In de workshop willen we de volgende vraagstellingen aan de orde stellen:

- 1) Wat is de relatie tussen de structurering van het aanbod en de vakgebied specifieke zoekstrategieën en methoden?
- 2) Hoe kan een interface (dwingend) helpen om complexere vragen te stellen en zo betere resultaten uit zoekacties te bewerkstelligen?
- 3) Hoe wordt een vakafhankelijke collectie samengesteld uit de zeer gevarieerde bulkaanbiedingen (segmentering) en hoe kan daar een portal site van worden gemaakt?
Wat is de beschikbaarheid aan informatie binnen een organisatie (-eenheid)?
- 4) Het probleem van zoeken in verschillende bestanden. Vaak zijn de meta-data velden zelf min of meer gelijkaardig maar zeker niet de vulling daarvan. Meer thesauri moeten aan een document gekoppeld kunnen worden, indien trefwoord concordantie onmogelijk blijkt.
- 5) De technische ontwikkelingen op het gebied van Z39.50 (ZING), waarbij ook XML standaarden worden geïntegreerd (Search/Retrieve Web)
- 6) De technische ontwikkelingen op het gebied van Open URL's, SFX linking ed.

Resultaat:

Als resultaat willen we naast een dieper inzicht vooral komen tot taakstellende onderzoeksvragen en richtlijnen voor aanlevering en gebruikersomgevingen, eveneens om voorstellen t.a.v. het faciliteren en meten van zoekgedrag.

Dit zowel op het vlak van technische voorwaarden inclusief de structurering van de aanbod kant, als op het niveau van de vakgebied specifieke eisen t.a.v. indexen, interfaces en materiaal clustering.

Deelnemers:

Primair de specialisten van uitgevers en bibliotheken (15 mensen).

Secondair de direct betrokkenen bij informatieontsluiting, marketing, publieksdiensten, enz. (15 mensen).

Duur: 1 dag.

Organisatie:

Dr. Eric Sieverts (UB-UU & HvA)

Dr. Kurt de Belder (UB-UvA)

Drs. Thomas Place (UB-KUB)

Drs. Nicolette van Dijk (ES-Scirus)

Joost Kircz (KRA) Coördinatie

Deze workshop zou goed mede georganiseerd kunnen worden met de Werkgemeenschap Informatiewetenschap.

VI-3) Redacteurenconferentie

De Nederlandse redacteurenconferentie moet gezien worden als een lokale versie van de succesvolle CCSC (www.niwi.knaw.nl/ccsc) conferentie.

Wij willen Nederlandse redacteuren bijeenbrengen die experimenteren met elektronische tijdschriften, of daar grote belangstelling voor hebben. Dit kunnen redacteuren zijn van tijdschriften van de grote uitgeverijen (ES, KAP) en van de experimenten van het Roquade project (www.roquade.nl).

Het doel van de conferentie is het verwerven van inzicht in het veranderende schrijf-publicatie en leesproces in een elektronische omgeving.

Met name willen wij de volgende aspecten belichten:

- 1) de eisen die verschillende vakgebieden stellen aan het gebruik en de leesbaarheid van hun artikelen wat betreft tekst, niet-tekstuele elementen, hyperlinking en kruisreferentiering, enz., alsmede de daaruit volgende eisen t.a.v. data structuur (SGML/XML etc.), opslag en presentatie,
- 2) veranderingen in redactionele processen wat betreft het peer-review systeem, in het bijzonder wat betreft niet-tekst elementen (data sets, multimedia etc) en open-peer review experimenten,
- 3) de veranderingen in de logistiek van de document administratie en het uiteindelijke publicatietraject,
- 4) de veranderingen in interactie tussen auteur en redactie, alsmede tussen lezer en tijdschrift, en lezer en auteur/lezer.

Doel:

Uiteindelijk proberen we te komen tot voorstellen of richtlijnen om het gebruik en de ontwikkeling van elektronische tijdschriften te stimuleren, te bestuderen en te meten.

Deelnemers:

Het aantal deelnemers kan bij deze workshop wat groter zijn dan bij de andere twee omdat we hier met een duidelijk leertraject te maken hebben. Een maximum van 50 aanwezigen maakt het nog goed mogelijk een goede discussie te hebben.

Duur: 1 dag.

Organisatie:

Prof. dr. Marian C. Horzinek

Redacteur Veterinary Science Tomorrow, Faculteit voor diergeneeskunde U-Utrecht

Prof. dr. Sjef van Erp,

Hoofdredacteur Electronic Journal of Comparative Law (EJCL), Rechtenfaculteit U-Maastricht

Drs. Natalia Grygierczyk

Projectleider Innovatieve projecten UB-UU

Drs. Anita de Waard

Manager Information Science ES

Joost Kircz (KRA) Coördinator

VI-4 Werkgroep onderzoeksvragen

In de workshop "logboekanalyse" VI-1, wordt het grondwerk verricht dat nodig is om methoden te ontwikkelen tot gebruikersonderzoek. In de workshop "ontsluiting", VI-2, worden ideeën ontwikkeld hoe goed gestructureerde informatie beter ontsloten kan worden en welke nieuwe mogelijkheden er zijn aan te geven. Vanuit de daar ontwikkelde gedachten over meta-data komen ook weer mogelijkheden tot gebruikersonderzoek naar voren. In de redacteursconferentie, VI-3, wordt uitgegaan van de auteur en redacteur en geconcludeerd wat als nieuwe mogelijkheden en wensen naar voren komt. Ook dit leidt natuurlijk naar nieuwe mogelijkheden tot gebruikersonderzoek.

Het probleem blijft echter dat, zoals zo vaak in de automatisering gebluft wordt, "wij alles kunnen, als u maar zegt wat u wilt".

De meeste resultaten van eerder genoemde workshops zijn voorwaardelijk voor beleidsbepaling. De wat en waarom vraag zweeft in alle drie workshops "boven de markt" maar is ondergearticuleerd. Dooddoeners als beter, mooier, goedkoper, effectiever en efficiënter zijn door iedereen geaccepteerde doelen, maar zijn geen operationele onderzoeksvragen. Concreet gaat het erom welke gebruiks- en gebruikersinformatie heeft de bibliotheek en uitgever nodig om beleid te onderbouwen. Voor de bibliotheek gaat dat om de inrichting van haar gebruikersomgeving en collectievorming. De informatie moet dus toe te snijden zijn op verschillende niveaus: gebruiksgroep/instituut - faculteit - UB - Bestuurslagen - UKB. Voor de uitgever is met name de eindgebruikersinformatie van belang voor productontwikkeling en de informatie-eisen van de bibliotheekorganisaties om aanlevering, onderhoud en rapportages te kunnen leveren.

Uit het vooronderzoek concludeer ik dat het van groot belang dat er een breed overleggremium komt dat, uitgaande van de huidige en in de workshops verder uitgewerkte ideeën, niet alleen de mogelijkheden maar ook de doelmatigheden bespreekt. Er moet op basis van de resultaten van deze samenvattende discussie inzicht komen t.a.v. de vraag welke vraagstellingen en resultaten van logboekanalyse, zoek- en ontsluitmethodieken en auteurswensen -en op welk aggregatieniveau- leiden tot beleidsbepaling bij collectievorming en productontwikkeling.

De aanbevelingen van de workshops moeten omgezet worden in onderzoeksprojecten die binnen de taakstelling van de stuurgroep inzicht geven in de nieuwe ontwikkelingen die verder reiken dan momentopnames en fenomenologie.

Er moeten duidelijker handvatten komen om gebruik -in al zijn vormen- te kunnen meten en wel op dusdanige wijze dat de resultaten gebruikt kunnen worden voor het ontwikkelen van economische modellen voor een nieuwe ronde licentie/ abonnement onderhandelingen.

Het ligt in de schoot van de stuurgroep om zo'n werkgroep al dan niet in te stellen en voorstellen tot samenstelling van dit gremium te doen.

VII) Ingediende onderzoeksvoorstellen

VII-1) Project voorstel Digitale rechten beheerssystemen UB-KUB

DREAM:
Digital Rights for Evolving Academic Models

Aldo de Moor, Manfred Jeusfeld
Hans Roes, Thomas Place
University of Tilburg

This text describes our initial ideas about an R&D project on rights management for digital libraries. As there is no such system in place today, we see an opportunity to design such a system according to the user needs, i.e. the library users, the authors, the libraries, and the publishers.

The increase of digital formats in libraries and the trend towards self-publishing in academia have led to the problem of properly maintaining the rights for digital publications. The potential benefit of digital rights management is motivated by the following questions:

From the author's perspective: *Which rights to my publications did I grant to which persons or organizations?*

From the publisher's perspective: *Which rights to digital publications did we acquire and to whom did we grant derived rights from them?*

From the reader perspective: *To which document collections do I have access?*

From the libraries perspective: *Do we have all necessary rights to make our digital archives available to our members?*

Beyond that we argue that every publishing community or company shall be able to design/configure a digital rights management system according to their own norms. The norms describe what behaviour in the system is regarded as acceptable.

Following, we sketch the vision of what we aim to achieve on the long term, second we describe the framework of a realistic first project to realize this vision.

Hypotheses & Goals

- Following the trajectory of an object endowed with Intellectual Property rights, is critical for the understanding and success of all publishing environments in particular also for self-publishing communities.
- Only few DRM-frameworks exist, many of them focused in IPR-issues in the entertainment industry.
- DRM in self-publishing communities is more complex, however. First, there are many more stakeholders involved playing complex *roles*. Second, there are numerous types of *content* elements, related in intricate ways. Third, self-publishing communities have a strong tendency to *evolve*.
- Current frameworks do not sufficiently support the management of this complexity. The goal of the DREAM project is to design and implement a framework for the digital rights

management in such evolving academic community models. These communities should be seen in the broad sense, ranging from doing information searching, co-authoring, to publishing activities.

Framework

The DREAM framework consists of the following components (Fig.1):

- *Workflows*: descriptions of the various workflows comprising the academic self-publication process.
- *Services*: various information services needed to enable the workflows.
- *Document-structures*: the elements needed to generate the documents needed in the publication process, ranging from information search results to review forms, to electronic journal issues.
- *DRM-norms*: normative structures that describe acceptable workflow, service, and document structure processes.
- *DR repository*: database which maintains digital rights for a user/organization community.

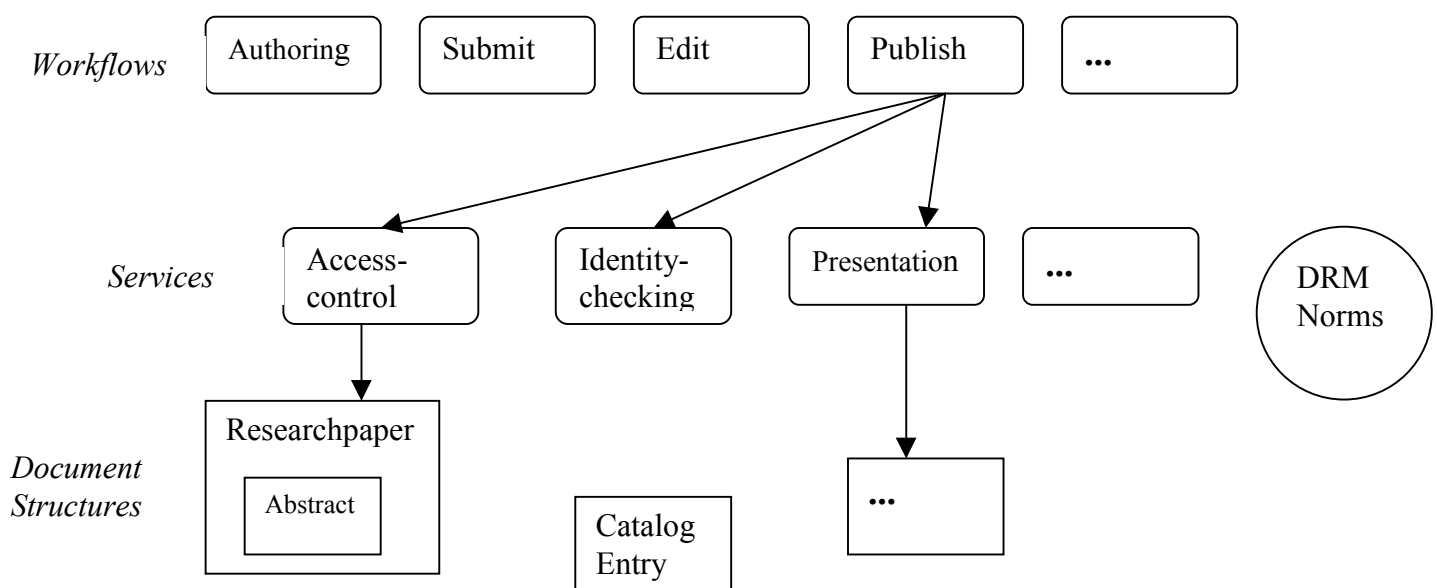


Fig.1 The DREAM Framework

The Project

In the project, we adopt a phased approach. We begin focusing on publishing activities, since much experience has been gained in this domain already via, for example, the EJCL and ARNO projects. Later, we extend our focus with, for instance, support for the co-authoring process.

- As a case, we will study the ARNO (Academic Research in the Netherlands Online) project. This is a good example of the university community building its own academic publishing infrastructure. Building on the results of this project, a sophisticated academic materials rights management system can be developed.
Both the DRM of teaching materials and academic publications can be studied.
- Important research questions to be answered in the project are:
 - What are DRM systems?
 - How can we integrate them with the social norms of the community?
 - How are they actually used?
 - How can the whole community agree on DRM issues?
- Research plan:
 - A: State-of-the-art in DRM literature survey.
 - B: DRM requirements in academic communities
 - B1: preliminary research (defining the questions):
 - "what *additional* services can DRM offer the various user *roles*, e.g. "as an *author*, a DRM system allows me to *more efficiently write papers*", etc.
 - B2: empirical research (asking the questions)
 - more extensive literature survey (focused on concrete DRM systems design)
 - interviews
 - user surveys
 - C: implementation (in the form of prototypes, working systems with selected functionality)

The parts A,B1 of the research plan are completed before the proper project proposal.

VII-2) Project voorstel Universiteit Twente

Zoeken van wetenschappelijke informatie door wetenschappers en studenten

Achtergrond:

Met de komst van ICT is wetenschappelijke informatie op een meer diverse manier te vinden dan voorheen het geval was. De verwachting is dat de zoekprocedures die wetenschappers en studenten volgen bijgevolg ook meer divers zal zijn.

Doel:

Achterhalen van het huidige en toekomstige zoekgedrag in een gemengde papieren en ICT zoekomgeving van gebruikers van wetenschappelijke informatie.

Vraag:

Welke patronen van zoeken laten wetenschappers en studenten zien in een gemengde informatieomgeving en in hoeverre variëren hun zoekgewoontes met de graad van digitalisering van de informatieomgeving?

Belang:

Informatie over persoonlijk zoekgedrag maakt het mogelijk de informatievoorziening dusdanig in te richten dat optimaal gebruik gemaakt wordt van ICT.

Werkwijze:

Selectie van een representatieve groep wetenschappers en studenten, welke een protocol bijhouden van hun zoekgedrag gedurende een of meer nader te bepalen periodes. De periode is variabel, omdat zoekgedrag samenhangt met het soort onderzoek waarmee men bezig is. Bovendien zal zo mogelijk tenminste twee keer worden gemeten om de stabiliteit van het gedrag te kunnen bepalen. Vooralsnog zullen wetenschappers en studenten uit twee verschillende disciplines ondervraagd. Bovendien zal het digitaal zoekgedrag digitaal worden vastgelegd.

Looptijd:

1,5 Jaar. De dataverzameling zal over een ruim jaar worden uitgespreid om zodoende de stabiliteit door de tijd heen te kunnen beoordelen.

Leverantie:

Zoekgedragmodellen gedifferentieerd naar discipline en eventueel naar wetenschappelijke senioriteit en de mate van digitalisering van de informatieomgeving.

Kosten:

- uitvoering door medewerker onderzoek 0,75 fte
- begeleiding door senioronderzoeker 0,1 fte
- begeleiding hoogleraar 0,1 fte
- reiskosten en verblijfkosten op declaratiebasis

- kosten voor presentatie van de resultaten aan de doelgroep op declaratiebasis en na overleg.

Onderzoekers:

Prof. dr. H.E. Roosendaal, hoogleraar wetenschappelijke informatie

Dr. P.A.Th.M. Geurts, methodoloog

Dr. P.E. v.d. Vet, informaticus

Medewerker onderzoek

Enschede, 27 feb. 2002

Bijlagen

B1) Internationale initiatieven

Op het gebied van gebruikersonderzoek en logboekanalyse zijn verschillende initiatieven die voor ons doel de moeite waard zijn.

Ongelukkigterwijs zijn de meeste gebruikersonderzoeken van een tamelijk plaats en tijd gebonden behavioristisch gehalte met weinig uitzicht op langere termijn trends.

Een eerste belangrijk experiment was het vaak geciteerde Tulip (The University Licensing Program) project van Elsevier Science samen met enkele grote Amerikaanse bibliotheken (Tul1996). Het project dat liep begin jaren negentig en ondanks de lessen die eruit getrokken konden worden een perfect voorbeeld van de stormachtige ontwikkeling van de technologie gebruikersonderzoek moeilijk maakt. Tijdens het project kwam de browser technologie op en werd het scannen van documenten definitief achterhaald door de noodzaak van SGML gecodeerde documenten.

In de afgelopen jaren zijn er veel projecten gestart voor digitale bibliotheken. In al deze projecten is het vraagstuk van *Usage* wel aan de orde.

In het recent afgesloten PEAK project, eveneens door Elsevier Science geëntameerd, werd geëxperimenteerd met verschillende soorten prijsmodellen (abonnement versus losse verkoop in verschillende mengingen). Diepe inzicht waarom en onderzoeker nu wel of niet een document tot zich neemt komt helaas hier op niet goed naar voren (PEA2000). Daarbij was het project gericht op het kopen van artikelen, waardoor het aspect van vrijbladeren onderdruk komt te staan.

Op dit moment zijn de volgende projecten van belang. Deze projecten zijn allemaal breed opgezet en omvatten zeer veel gebruikers, bibliotheken en uitgevers. Het is dus van belang om vooral bij deze activiteiten aan te sluiten en van te leren, alvorens grootschalig eigen Nederlandse initiatieven op te zetten. Een overzicht van lopende initiatieven op het gebied van gebruikersstatistiek is te vinden bij de PALS werkgroep op:

<http://www3.oup.co.uk/central/pals/initiatives.htm>

Het Equinox programma van de universiteiten van Manchester en Dublin en de bibliotheken van Münster is door de EU gesponsord in het Telmatics for Libraries programme.

<http://equinox.dcu.ie./reports/pilist.html>

Dit *Library Performance Measurement and Quality Management System :Performance Indicators for Electronic Library Services* project had tot doel om software te ontwikkelen voor de hybride bibliotheken (E&P) en om daar prestatie-indicatoren voor te ontwikkelen.

Het Collection Management Initiatief (www.ucop.edu/cmi/) van de universiteit van Californië gaat verder op in de Amerikaanse behavioristische traditie door een zeer uitgebreide test te doen waarbij enkele campussen alleen papier en anderen alleen elektronisch materiaal krijgen. In de loop van 2002 moeten de eerste resultaten binnen zijn. Net als bij andere experimenten zijn ook Elsevier Science en Kluwer van de partij.

Een fundamenteeler project is het PALS (de Publishers and Libraries Solutions group) project van de JISC (Joint Information Systems Committee of the UK further and higher education

funding bodies), ALPSP (Association of Learned and Professional Society Publishers) en de PA (The Publishers Association). <http://www.usagestats.org/>

Dit uiterst belangwekkende project probeert een platform te vinden waarbij richting wordt gegeven aan de volgende zaken:

- Welke data-elementen moeten gemeten worden?
- De definitie van deze data-elementen
- Presentatie van de formaten/ frequentie en granulariteit van de rapporten
- De aanleveringsmethode
- Methoden om te komen tot een combinatie van gebruikersrapportages van direct gebruik en gebruik via intermediären -inclusief *gateways*, aggregatoren, en elektronische document leveranciers.
- Methoden om *remote* gebruik van institutionele licenties te meten
- Methodes voor het meten van *caches* en *mirrors*

Dit zwaar aangezette project zal naar alle waarschijnlijkheid in ultimo 2002 een rapport uitbrengen. In voorstel VI-1 wordt dan ook gesteld dat dit uitgangspunt moet zijn van eigen activiteiten.

Binnen de International Coalition of Library Consortia (ICOLC) zijn er in november 1998 richtlijnen gepubliceerd voor statistische metingen van gebruik van web-gebaseerde informatie bronnen. Deze zijn recentelijk in december 2001 bijgesteld.

www.library.yale.edu/consortia/2001webstat.htm

Ook deze richtlijnen zijn basis voor de voorgetelde workshop (VI-1).

Recent is er een samenwerking afgesproken tussen de Council on Library and Information Resources (CLIR) en de Association of American Publishers (AAP) om te komen tot overleg over: digitale archivering, gebruikersstatistiek en organisatorische barrières binnen uitgevers en bibliotheken (Cli2001). Dit initiatief kan duidelijk gevoed worden door de initiatieven die in dit rapport worden voorgesteld

B2) Initiële lijst van te stellen vragen aan een logboekanalyse

Onderstaande lijst vragen is een compilatie van vragen die in de loop van dit vooronderzoek genoemd zijn. Deze lijst kan gelden als leidraad voor verdere discussie over de waarde en het nut van sommige vragen. In logboekanalyse gaat het immers om 1^e) wat men idealiter zou willen weten en 2^e) wat technisch redelijk haalbaar is.

Het gaat daarbij om zowel het aantal keren dat een artikel wordt gebruikt als de dynamiek van het gebruik.

Het antwoord op deze vragen hangt af van een duidelijk inzicht wat de termen betekenen. Daartoe moeten er duidelijke operationele definities worden vastgesteld. Een typisch voorbeeld is het begrip "sessie", dat vanuit de gebruiker gedefinieerd kan worden (de tijd dat de gebruiker on-line iets zoekt) of vanuit de bron (de tijd dat de gebruiker actief zocht op Science Direct). Binnen één gebruikerssessie kunnen meerdere bronsessies bestaan, als de gebruiker uitlogt, zich bedenkt en een half uur later weer inlogt. Voor de bron is het niet te zien of dit dezelfde gebruiker is of twee personen die van de zelfde plek inloggen (bijv. vanuit een studiezaal).

Onderstaande lijst is dus zuiver indicatief om de discussie verder te helpen.

1. Per geïdentificeerde gebruiker

- 1.1. Aantal sessies.
- 1.2. Aantal zoekvragen per sessie naar categorie (browse, search, basic, advanced, enz.).
 - 1.2.1. Binnen de zoekvraag aantal bewegingen van en naar resultaat pagina's, resp. inhoudsopgaven.
- 1.3. Aantal bekeken pagina's HTML (tijdschrift titel, artikel).
 - 1.3.1. Onderscheid tussen abstracts en full text
 - 1.3.2. Aantal malen dat binnen het HTML document gebruik wordt gemaakt van hyperlinks.
 - 1.3.3. Aantal malen dat er via een hyper-link "uit" het document wordt gesprongen en weer terug.
- 1.4. Aantal bekeken (downloaded) pagina's PDF, PS, RealPage, enz. (tijdschrift titel en artikel).
- 1.5. Aantal verschillende bronnen die gebruikt worden.
 - 1.5.1. On-site.
 - 1.5.2. Direct bij de leverancier.
- 1.6. Aantal bewegingen (doorklikken) tussen bronnen.
- 1.7. Aantal bewegingen naar niet artikel informatie (databases, multimedia component, web-pagina's buiten de collectie, zoals home-pages).
- 1.8. Type en aantal van gebruikte multimedia plug-ins, besturingssysteem, type browser.
- 1.9. Hoe lang duurt een sessie.
 - 1.9.1. Op welke tijd van de dag.
- 1.10. Hoe vaak wordt, waar, teruggekomen.
- 1.11. Hoeveel missers zijn er (404,403, enz).
- 1.12. Hoe vaak wordt toegang geweigerd doordat het maximaal aantal gebruikers is bereikt.

2. Per IP cluster (in principe een aggregatie van het eerste punt)

- 2.1. Aantal sessies
- 2.2. Aantal geraadpleegde bronnen.
 - 2.2.1. Aantal bezochte HTML pagina's per bron naar tijdschrifttitel en artikel.
 - 2.2.2. idem voor PDF, PS, RealPage, enz.
- 2.3. Metingen van verkeer tussen bronnen.
- 2.4. Aantal gebruikte multimedia elementen.
- 2.5. Onderscheid student, AIO, medewerker.

3. Per tijdschrift(en)

- 3.1. Aantal tijdschriften licentie/abbonement- situatie.
- 3.2. Aantal artikelen naar abbonement/licentie- situatie
- 3.3. Totaal aantal HTML pagina's per titel per tijdseenheid.
 - 3.3.1. Uitsplitsing naar IP-cluster.
- 3.4. Totaal aantal PDF, PS, RealPage, enz. pagina's per titel per tijdseenheid.
 - 3.4.1. Uitsplitsing naar IP-cluster.
- 3.5. Bewegingen binnen een tijdschrift van artikel naar artikel.
- 3.6. Bewegingen tussen tijdschriften (al dan niet via Crossref).
- 3.7. Aantal artikelen gevonden via Zoeken of via bladeren.
- 3.8. Meest/minst gevraagde titel/ artikel/ pagina.

4. Per cluster van bronnen

- 4.1. Linking van een document naar een ander (padanalyse)

B3) Lijst van gesprekspartners

Uitgevers:

Kluwer Academic Publishers:

Peter Coebergh- Managing director global sales, Ad Plaizier- Publishing manager Electronic Publishing Unit, Rob Doornebal- General Manager Manufacturing Department

Elsevier Science:

Leo Voogt- Director Global library relations, Frank Vrancken Peeters,- Managing Director ScienceDirect, Marthijn Borghuis-General Manager End-user and Library Reserach, Jonathan Clark- Technology director S&T division.

Bibliotheken:

Amsterdam:

Nol Verhagen-Bibliothecaris, Kurt de Belder- Chief, Division of Electronic Services, Rense Brandsman-Hoofd digitaal productiecentrum, Boy de Haas-Division of Electronic Services.

Den Haag:

Koninklijke Bibliotheek: Gerard van Trier- Directiesecretaris, Repke de Vries- Coördinator digitale bibliotheek, Henk Voorbij-Onderzoeker.

Delft:

Cees Konings

Groningen:

Alex Klugkist-Bibliothecaris, Peter van Laarhoven- Hoofd elektronische bibliotheek

Maastricht:

John Gilbert-Bibliothecaris, Fons van den Eeckhout-Hoofd Medische bibliotheek, Henk van den Hoogen- Hoofd. afd. wet.inf.&doc., Dick Schreurs-Benchmarking, Roel Tilly- acquisitie en Licentie bureau UKB.

Nijmegen:

Anja Smit- Adjunct-bibliothecaris

Tilburg:

Hans Roes- Adjunct bibliothecaris, Thomas Place Adjunct -bibliothecaris

Utrecht:

Bas Savenije- Bibliothecaris, Natalia Grygierczyk-projectleider innovatieve projecten, Eric Sieverts- Afd. informatie technologie

Wageningen:

Leo Waaijers-Bibliothecaris

Overigen:

John Mackenzie Owen- Hoogleraar Informatiewetenschappen UvA

Jan Peters- CVB- KUN

Peter Geurts- wetenschappelijk medewerker UTwente

Paul van der Vet- wetenschappelijk medewerker UTwente

B4) Literatuurreferenties

Selectie van gebruikte literatuur, niet noodzakelijk in de tekst genoemd

Voor een excellent overzicht van de literatuur zie:

Charles W. Bailey, Jr. Scholarly Electronic Publishing Bibliography (SEPB).

“This selective bibliography presents over 1,500 articles, books, and other printed and electronic sources that are useful in understanding scholarly electronic publishing efforts on the Internet” <http://info.lib.uh.edu/sepb/sepb.pdf> (resp.html).

Deze bibliografie wordt geregeld bijgewerkt.

Een artikel dat deze bibliografie beschrijft is:

Charles W. Bailey, Jr: The Scholarly Electronic Publishing Bibliography. Jnl of Electr.Publ. vol 7 issue 2. December 2001. <http://www.press.umich.edu/jep/07-02/bailey.html>

Een ander belangrijk overzichtswerk is het recente boek:

Carol Tenopir and Donald W. King. Towards Electronic Journals. Realities for scientists, librarians and publishers. Special Libraries Association (SLA) publishing, 2000, 488 pp.

Verder:

- Arn2002 ARNO project (Academic Research in the Netherlands Online).
<http://www.uba.uva.nl/en/projects/arno/>
- Bor2000 Marthyn G.M. Borghuis. What to count & what not. A white paper on the filters to be applied to a web-server log file, before usage-analysis and reporting can start. September 2000, 7pp. Verkrijgbaar via: <http://reports.sciencedirect.com>
- CCS2001 Conferentie: Change and Continuity in Scholarly Communication. 2001.
Bijdragen en verslagen via: www.niwi.knaw.nl/ccsc
- Cli2001 CLIR issues Number 25 Ja/Febr 2002. CLIR and AAP form joint working group.
- Har1999 Stevan Harnad Free at last: The future of peer-reviewed journal. D-Lib magazine, 5(12). December 1999. www.dlib.org/dlib/december99/12harnad.html
- ISC1996 Conference Recommendations of the Joint ISCU-Press/UNESCO Conference on Electronic Publishing in Science. Paris February 1996.
<http://associnst.ox.ac.uk/~icsuinfo/recommen.htm>
- ISC2001 Conference Recommendations of the Second ICSU-UNESCO Conference on Electronic Publishing in Science, Paris. February 2001.
<http://associnst.ox.ac.uk/~icsuinfo/confer01.htm>
- Ine2001 Inera(TM) Incorporated: E-Journal Archive DTD Feasibility Study. Prepared for the Harvard University Library Office for Information Systems E-Journal Archiving project. December 5, 2001. Verkrijgbaar via:
www.diglib.org/preserve/hadtdfs.pdf (65 pp)
- Int2000 International Working Group. Defining and Certifying Electronic Publication in Science. A proposal to the International Association of STM Publishers, Published as a report. Learned Publishing 2000:13(4). Oct., 251-8. ook via: www.learned-publishing.org.
- Fro2000 J.C. Frowijn, A.H. Laeven en H.J. Voorbij. Benchmarking van wetenschappelijke bibliotheken. Surf/IWI, 2000, 149 pp
- Kel2001 Alice Keller: Future developments of electronic journals: a Delphi survey. The Electronic Library, 19(6), 2001. pp. 383-396
- Kir2001 Joost G. Kircz. New practices for electronic publishing 1: Will the scientific paper keep its form. Learned Publishing. Volume 14. Number 4, October 2001. pp. 265-

272. Ook via www.learned-publishing.org
- Kir2002 Joost G. Kircz. New practices for electronic publishing 2: New forms of the scientific paper. Learned Publishing. Volume 15. Number 1, January 2002. pp. 27-32. Ook via www.learned-publishing.org
- Kli2002 Rob Kling, Lisa Spector & Geoff McKi., Locally controlled scholarly publishing via the Internet: The Guild Model. CSI Working paper, SLIS Indiana Univ. www.slis.indiana.edu/csi/wo/wp02-01B.html
- Lut2001 Judy Luther. White paper on electronic journal usage statistics. Council on Library and Information Resource. Wash. DC. 2001, 26 pp. Herdrukt in JEP 6(3) . www.press.umich.edu/jep/06-03/luther.html
- Man2000 Managementrapportage Digitale Bibliotheek 2000. Intern rapport UB-UvA.
- Mea1998 A. J. Meadows, Communicating research, Academic Press, 1999, 266 pp.
- Mos1998 P. Mostert e.a. Digitaal academisch erfgoed. IWreeks/erfgoed1, Surf/IWI 1998.
- PEA2000 PEAK-Pricing Electronic Access to Knowledge. Conference programme: www.si.umich.edu/PEAK-2000?program.htm en bijdragen daarin, met name: Hunter, King&Tenopir, Gazzale&MackieMason.
- Pee1996 Robin P. Peek and Gregory B. Newby. Scholarly Publishing: the electronic frontier. The MIT Press. 1996, 363pp.
- Raa200 Raad voor Cultuur: Advies De rol van de bibliotheek bij het beheersen van de informatievloed. mei 2000.
- Til1999 UB-Tilburg: The use of paper and electronic journals by researchers. Prepared for Steering Groep IWI, Janaury 1999 : <http://cwis.kub.nl/~dbi/project/journal/eindrep.pdf>
- Tja1998 H.D. Tjalsma, e.a. Digitaal academisch erfgoed *Getoetst*. IWreeks/erfgoed2, Surf/IWI 1998.
- Tol1999 M.. C. van der Tol. The abstract as an orientation tool in modular electronic articles.' In: Alfons Maes, Hans Hoeken, Leo Noordman & Wilbert Spooren (eds.), Document Design. Linking writers' goals to readers' needs. Tilburg: Tilburg University, 1999. pp. 175-186.
- Tol2000 M. C. van der Tol. 'Naar richtlijnen voor elektronische abstracts in een modulaire publicatieomgeving'. In: *Tekst[blad]*, jrg. 6, no. 1. 2000. pp. 9-13.
- Tol2001 Maarten van der Tol, 'Abstracts as orientation tools in a modular electronic environment.' In: *Document Design*, vol. 2, no. 1. 2001. pp. 76-88.
- Tul1996 Tulip Final report, Elsevier Science 1996. 368 pp.
- Voo2001 H. J. Voorbij. Elektronische tijdschriften. Handboek Informatiewetenschap bibliotheek en archief. Red. G. van Trier, D.W.K. Jansen en H. Prins.Bohn Stafleu van Loghum, nummer 19, september 2001, pp I 520-1- I 520-34.
- Wel2001 Ann. C. Weller. Editorial Peer Review. Its strengths and weaknesses. ASIST Monograph series. Information Today Inc. 2001. (342 pp)