The practice of scientific reading: analog is still alive

Joost Kircz

University of Applied Sciences Amsterdam

KRA- Publishing Research

Force 11- *Beyond the PDF2 Conference*. Amsterdam March 19-20, 2013

The essential quest is:

What medium (substrate) is needed for what type of information?

In order to: internalize, or reuse, or look-up, or just quote.

The Scientific Research Paper

Intro
Stating the problem
Positioning
Methods
Results
Interpretation
Findings
Leads to further work

= A standardized argumentational structure to pack a great variety of information & knowledge 'entities' in **essay form**.

= An author centered product

The discussion is always:
Does the publisher caters for the authors and keeps them as friends; hence,
the reader/ library pays?
or
Does the publisher caters for the readers and keep them as friends; hence,
authors pay page charges?

The author proudly presents her knowledge, claims and results
But more often than not the reader only uses parts of the publication for own work

So, what will change beyond PDF?

Writing and Reading are asymmetric

The author wants e.g.:

- Priority & claim
- Proof of expertise and status
- Inform and discuss with colleagues
- Convince sponsors

The reader wants e.g.:

- Information
- Understanding
- Usability
- Anchor points for own work

Different granularities!

Beyond PDF?
Or beyond the essay?

We have to make a clear distinction between two chains

1) Communication chain between author & reader, to inform the reader.

2) Added value chain to certify and validate the author's work, status, job, prestige. Archiving

Beyond PDF?
Or beyond the essay?

Page design is still based on the proven value of paper journals.

PDF is only an electronic transportation vehicle of that very page.

Enhanced PDF is the first step to network knowledge but still based on the page paradigm.

The essay form is as if we address an innocent novice in the field

Analog, Digital, and Binary

Analog is a complete “view” (independent of the substrate: paper, display, retina).

Comprehensive for humans

Digital is the breaking apart in notions (the >1K words of the picture).

Often understandable

Binary is just one numerical calculation scheme.

Unintelligible

Does these various forms require various substrates?

=> Aside 1-6

Aside 1- Language as being digital

Analog **(a complete view)**
The iconograph is a picture
Digital **(a semantic unit)**
The phoneme is a letter from the western alphabet. (German: Buchstabe)

Allows for audio books
Binary is the electronic representation:
Force11=

01000010011001010010000001110011011101010111001001100101001000000111010001101111001000000110010001110010011001011100100000011110010110111101110101011001000100000010011111011101100110000101101100011101000110100101101110011001010010111001101110

Or HEX =66:6f:72:63:65:31:31

By the way: Hex F0rce11= Decimal 240

Aside 2- Except in English

Analog:

Digital/ Alphabetical: Sound = Fish

Phonological: GHOTI (with thanks to Bernard Shaw)

the gh = f as in rouGH

the o = i as in wOmen

the ti = sh as in naTIon

The word - not the syllable - is the unit

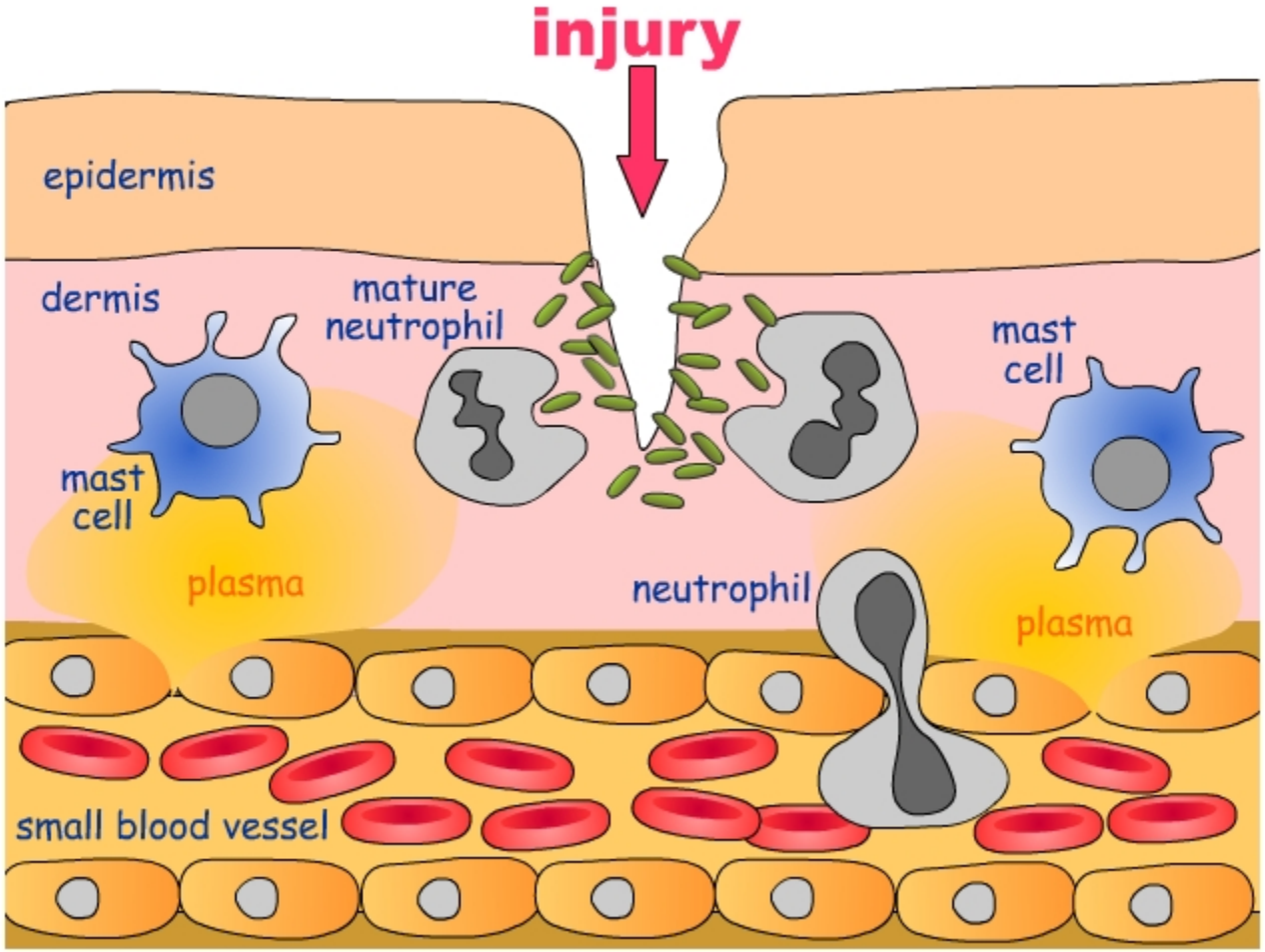
Binary: www.Binaryfish.com. A company for Microsoft apps.

Conclusion: A methods for English does not necessarily fit as model for other languages!

Aside 3: Analog THE REAL THING

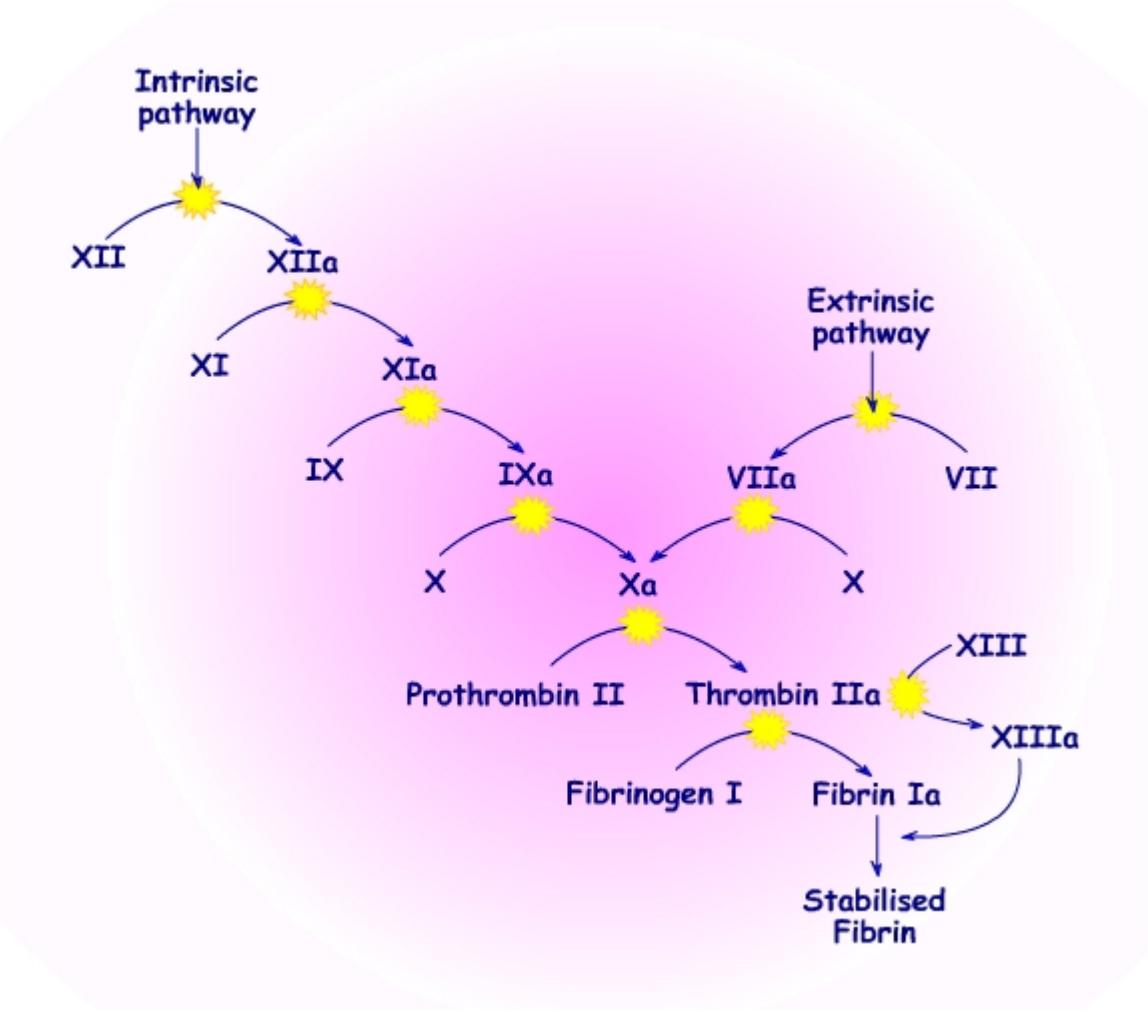
From: http://ewhomecare.com/services/woundcare.html

Aside 4: Digital; Breaking apart in linguistic notions



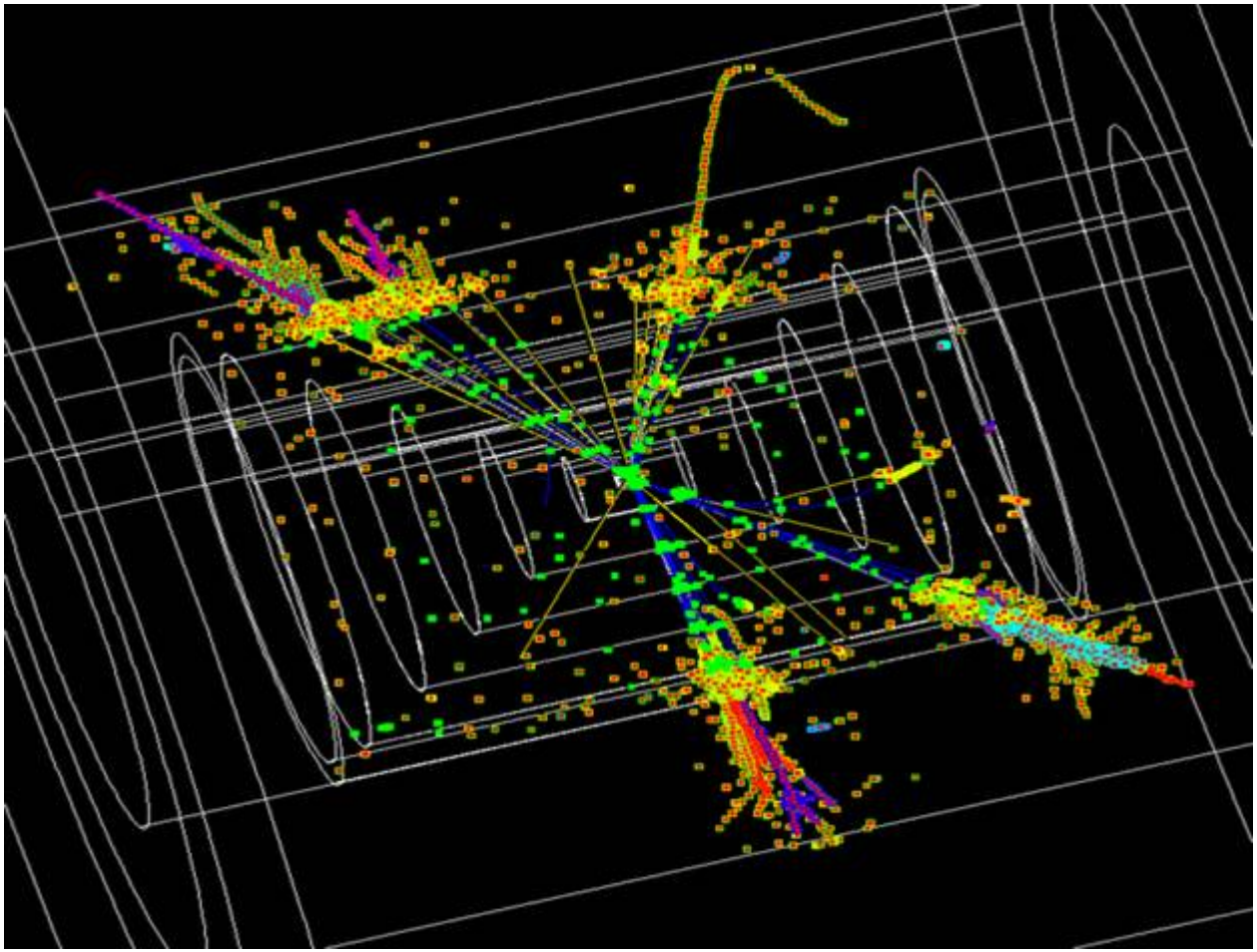
Summary of inflammation showing key inflammatory cells - mast cells and neutrophils
From: http://www.hcc.bcu.ac.uk/physiology/woundhealing.htm

Aside 5: Binary Tree; manipulation & calculation



The clotting cascade showing the intrinsic and extrinsic pathways
From: http://www.hcc.bcu.ac.uk/physiology/woundhealing.htm

Aside 6: Making an Analog Explanation (from the unintelligible)



A simulation of the decay of a Higgs boson in a linear collider detector. (Image courtesy of Norman Graf.)
From: http://public.web.cern.ch/public/

Fireworks as metaphor

A typical scientific article

Development of a two-leaf light use efficiency model for improving the calculation of terrestrial gross primary productivity
By: Mingzhu Hea, + 11 co-authors
Agricultural and Forest Meteorology
Volume 173, 15 May 2013, Pages 28–39

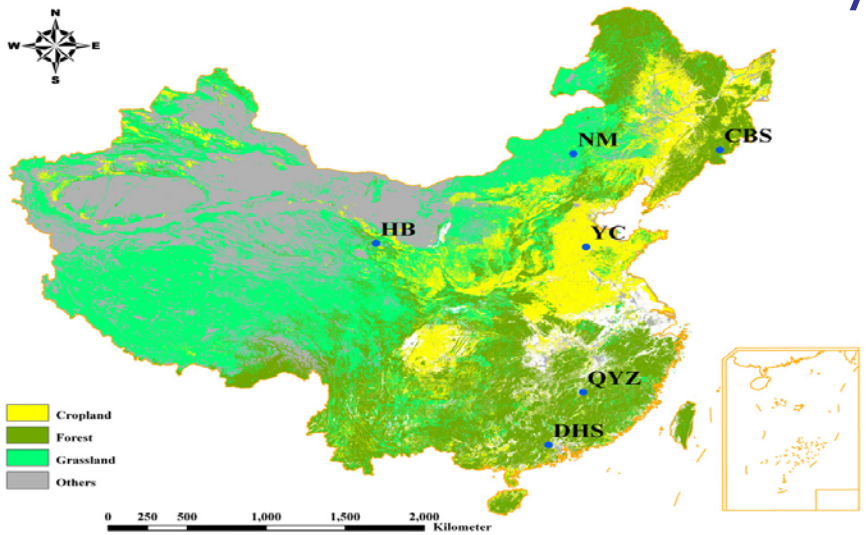


Fig. 1. Distribution of the 6 sites in China at which measured GPP was used for calibrating and validating the TL-LUE model developed in this study. The background is the GLC2000 land cover map.

used for the XLHT site. Both the MODIS GPP and LAI products have a spatial resolution of 1 km. The projection of these data is Sinusoidal, and MRT (MODIS Reprojection Tools) was used to reproject them into an UTM/WGS 84 projection. Because of residual cloud contamination, the MODIS LAI product has some unrealistically abrupt short-term fluctuations, and the locally adjusted cubic-spline capping (LACC) method (Chen et al., 2006) was used to smooth MODIS LAI. The smoothed LAI series were then input into the MOD17 algorithm and the TL-LUE model for calculating *f*PAR.

2.2. Method

2.2.1. The MOD17 algorithm

The MOD17 algorithm is based on the radiation conversion efficiency concept of Monteith (1972). GPP is calculated as (Running et al., 2000):

$$\mathrm{GPP} = \varepsilon_{\max} \times f(\mathrm{VPD}) \times g(T_a) \times \mathrm{PAR} \times f\mathrm{PAR} \quad (3)$$

where *f*PAR is the fraction of PAR absorbed by the canopy and calculated as:

$$f\mathrm{PAR} = 1 - e^{-k \times \mathrm{LAI}} \quad (4)$$

where k is the light extinction coefficient and set as 0.5; LAI is the green leaf area index of the whole canopy.

In Eq. (3), $\varepsilon_{\max}$ is the maximum LUE and changes with vegetation types (Table 2). f(VPD) and $g(T_a)$ are the scalars of VPD and the minimum air temperature (T_a) used to downscale $\varepsilon_{\max}$ to the actual. They are calculated as:

$$f(\mathrm{VPD}) = \begin{cases} 0 & \mathrm{VPD} \geq \mathrm{VPD}_{\max} \\ \dfrac{\mathrm{VPD}_{\max} - \mathrm{VPD}}{\mathrm{VPD}_{\max} - \mathrm{VPD}_{\min}} & \mathrm{VPD}_{\min} < \mathrm{VPD} < \mathrm{VPD}_{\max} \\ 1 & \mathrm{VPD} \leq \mathrm{VPD}_{\min} \end{cases} \quad (5)$$

$$g(T_a) = \begin{cases} 0 & T_a \leq T_{\min} \\ \dfrac{T_a - T_{\min}}{T_{\max} - T_{\min}} & T_{\min} < T_a < T_{\max} \\ 1 & T_a \geq T_{\max} \end{cases} \quad (6)$$

Table 1
Summary of climate and vegetation characteristics of the 6 tower sites.

Sites	Changbaishan	Qianyanzhou	Dinghushan	Yucheng	Haibei	Xinlinhot
Lat/Lon	42°24′N 128°06′E	26°45′N 115°04′E	23°10′N 112°32′E	36°57′N 116°36′E	37°40′N 101°20′E	43°33′N 116°40′E
Climate type	Temperate continental climate influenced by monsoon	Sub-tropical monsoon climate	The monsoon humid climate of torrid zone of south Asia	Semi-humid and monsoon climate	Plateau continental climate	Temperate semiarid continental climate
Annual mean precipitation (mm)	600–900	1489	1956	582	580	350–450
Annual mean temperature (°C)	3.6	18.6	21	13.1	−1.7	−0.4
Vegetation type	Mixed forest	Evergreen	Evergreen	Winter	Alpine meadow	Grassland

Is a Scientific Article a hors d'oevre of consumable knowledge and information entities?

Randomized or structured consumption?

A scholarly paper has a traditional purpose

Show the reader that you know about the issue

Define the quest you are dealing with

Explain the context of the research

Explain the working methods

Present and defend the results

Draw conclusions and state further questions

A scholarly paper has a reasoning, a rhetoric and an argumentational structure

Breaking up the essay

Researchers/students leaf, browse, read haphazardly, go to and fro. (work of e.g. Kircz et al., Tenopir et al.)

Two ways of attacking the traditional essay:

1) Breaking apart in comprehensive modules, each representing a stage in the reasoning and presentations. (Kircz/Harmsze)

2) Leaving the flow of the essay (or module) and look at the grammar.

Anita de Waard's dictum: *"In defense of the clause as the unit of thought"*

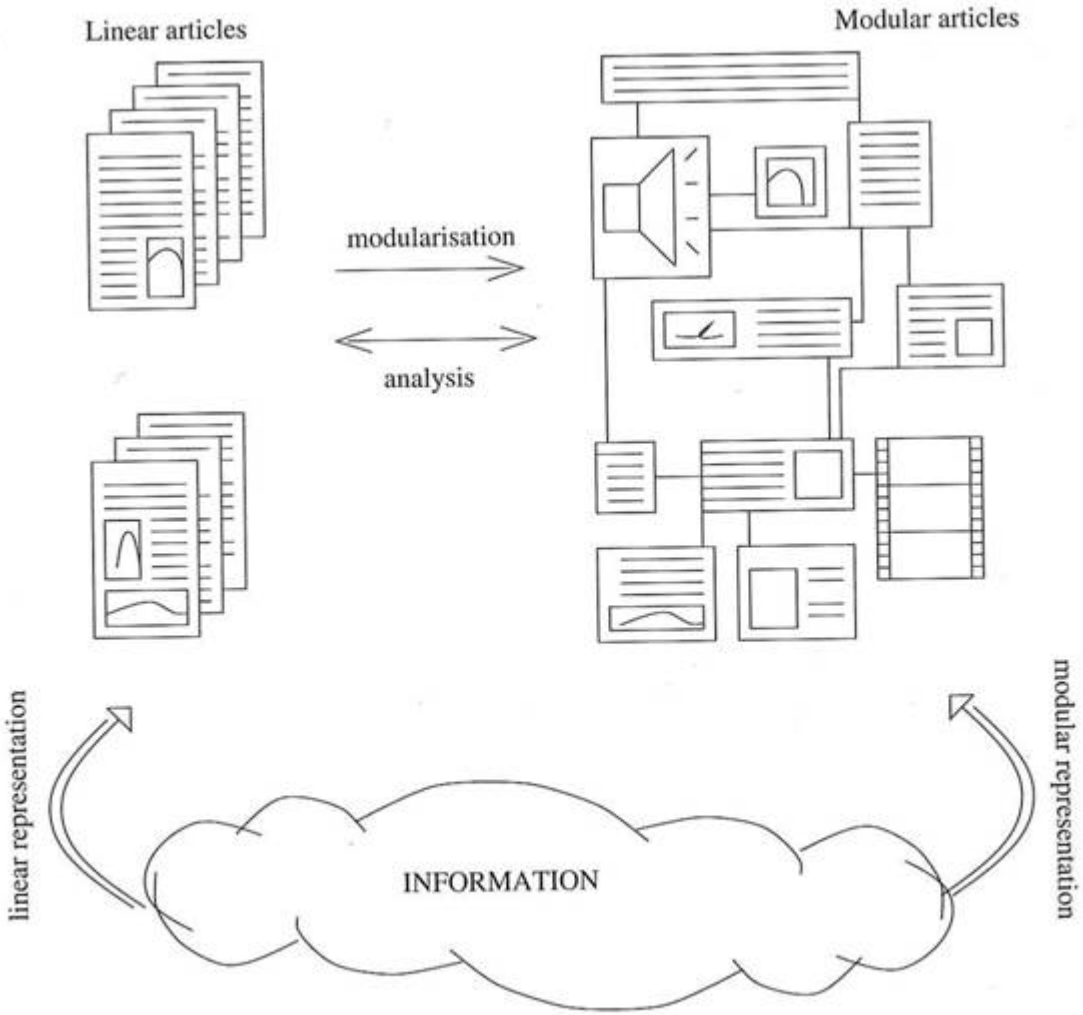


Figure 4.1. Linear and modular representations of information: mappings from the information space.

From: Fréderique Harmsze: A modular Structure for scientific articles in an electronic environment PhD Univ. of Amsterdam 2000

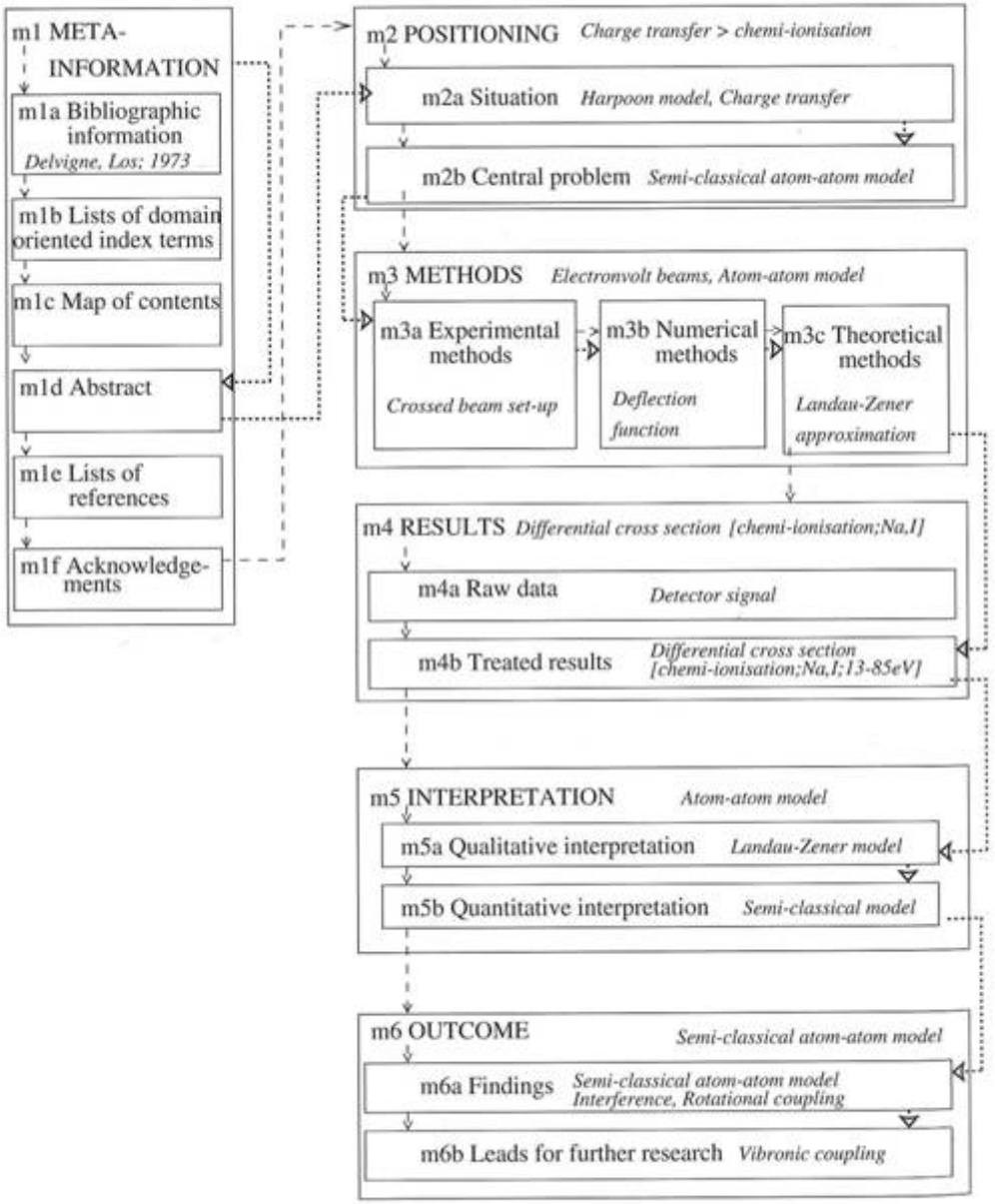


Figure A.1. The module meta-information and the modules distinguished following the conceptual function of the information, and the sequential paths leading through the article. The dashed line indicates the complete sequential path, and the dotted line the essay-type sequential path.

Demands typed hyperlinks!

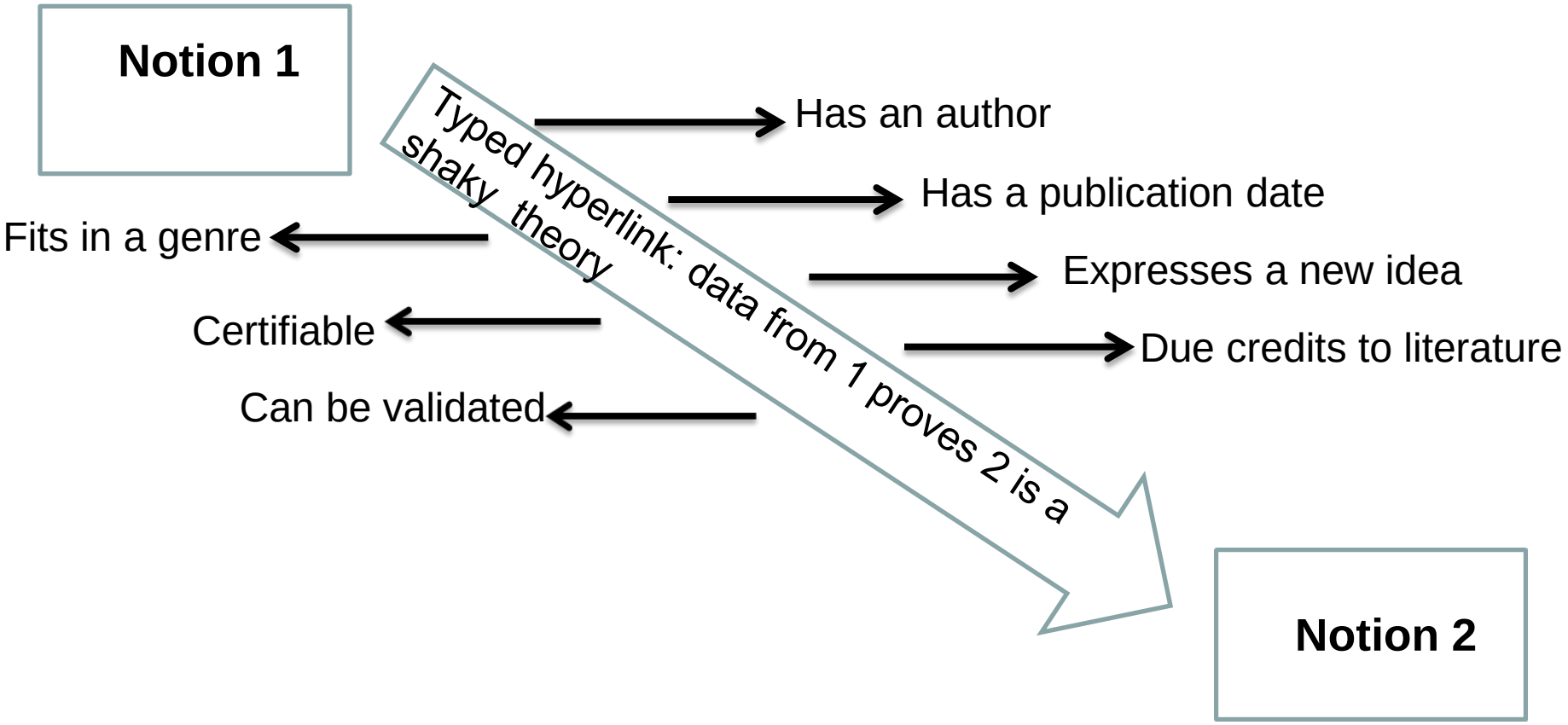
Hyperlinks have a meaning!

Organizational hyperlinks

Argumentative hyperlinks

From: Fréderique Harmsze:
A modular Structure for scientific articles in an electronic environment
PhD Univ. of Amsterdam 2000

In this model the shortest Scientific paper will be a typed Hyperlink



Old & new added value
Publisher's tasks/rôles

Primary key values are still:

validation & certification (V&C)

Secondary key values are still:

presentation & dissemination

New key values are related also to readers' needs:

V&C of components of scientific reporting,

Aggregation per genre as well as per kind of information/knowledge (text, animation, data,…)

Multiple presentations on various substrates

V&C of components of scientific reporting 1

Text:

1)Language brushing-up in order to enable English to be a lingua franca. Readers must understand

2)Text structuring: themes & tools. Rs must follow

3)Integration of specific keyword ontologies in the text processors (pre-coordination in IR terms). Rs must find

4)Plagiarism checks & balances (due citation). Rs must trust

5)Aggregation in repositories. Rs must roam around

V&C of components of scientific reporting 2

Sound & Visuals:

1)Standardization of illustrations, sound & (moving) pictures to enable overlays & comparison. Rs compare

2)Metadata grammars for sound & vision. (Here also pre-coordination by the author where possible.) Rs must find and understand in context

3)Standards for including animations (games) Rs must understand the underlying constraints

V&C of components of scientific reporting 3

Data:

1)Standardization units (per field & genre).
Readers reuse and compare.

2)Metadata grammars for storage

3)Plagiarism checks & balances (due citation)

4)Shared repositories of original (raw) data

5)Etc.

V&C of components of scientific reporting 4a

Aggregation & Dissemination :

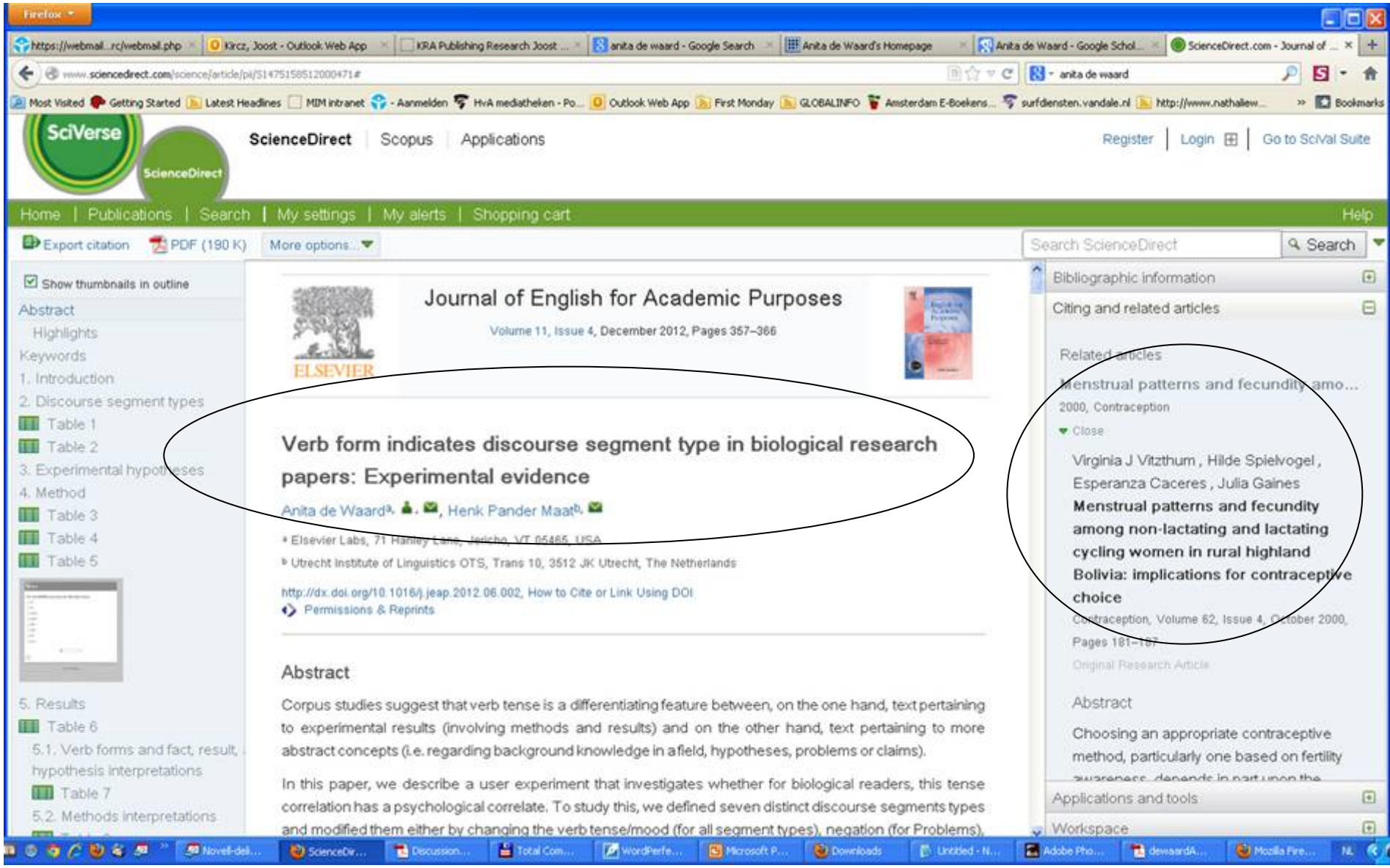
1) Comprehensive coupled databases. DOI per “module” (textual, dataset, visuals…)

2)Relevant search methodologies based on content (genre dependent ontologies) as well as probabilistic methods.

E.g. Elsevier’s Clinical Key or Geofacits experiment.

= > see aside 7

Aside 7: To what extent can technology help us



V&C of components of scientific reporting 4b

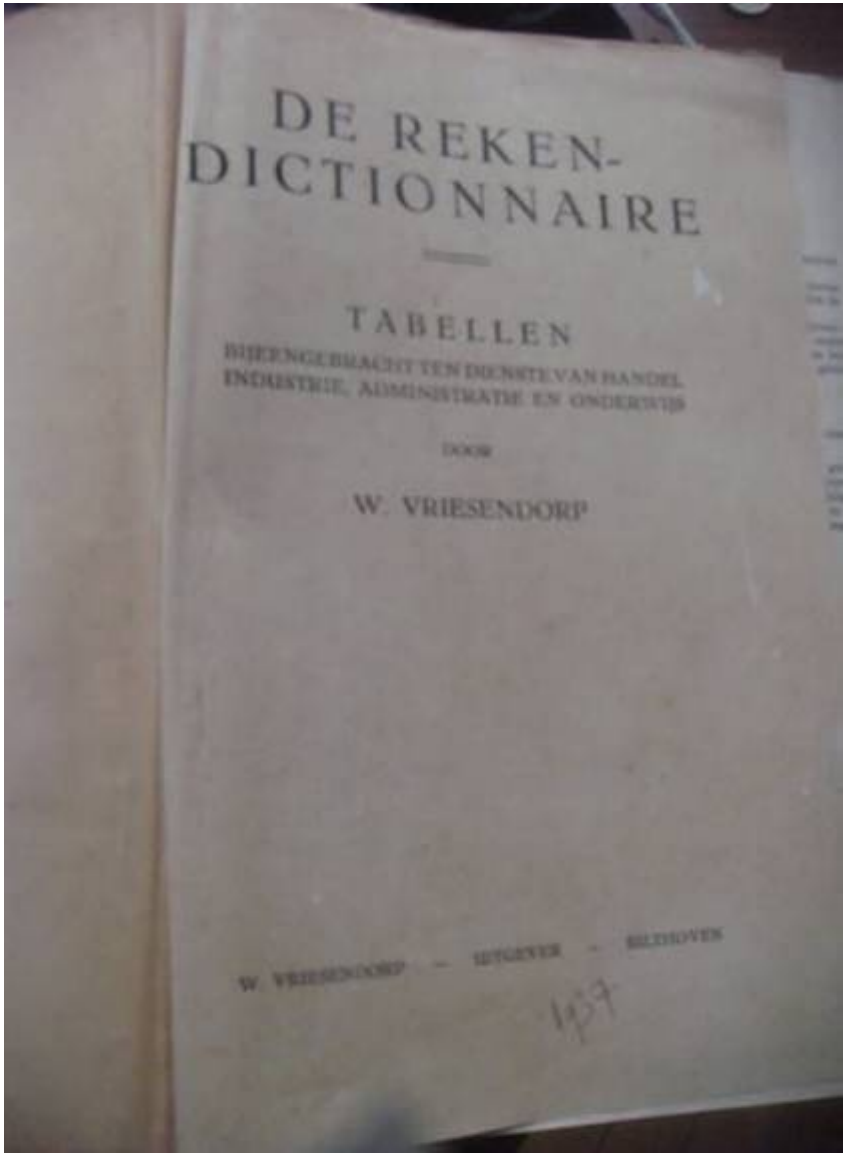
Aggregation & Dissemination:

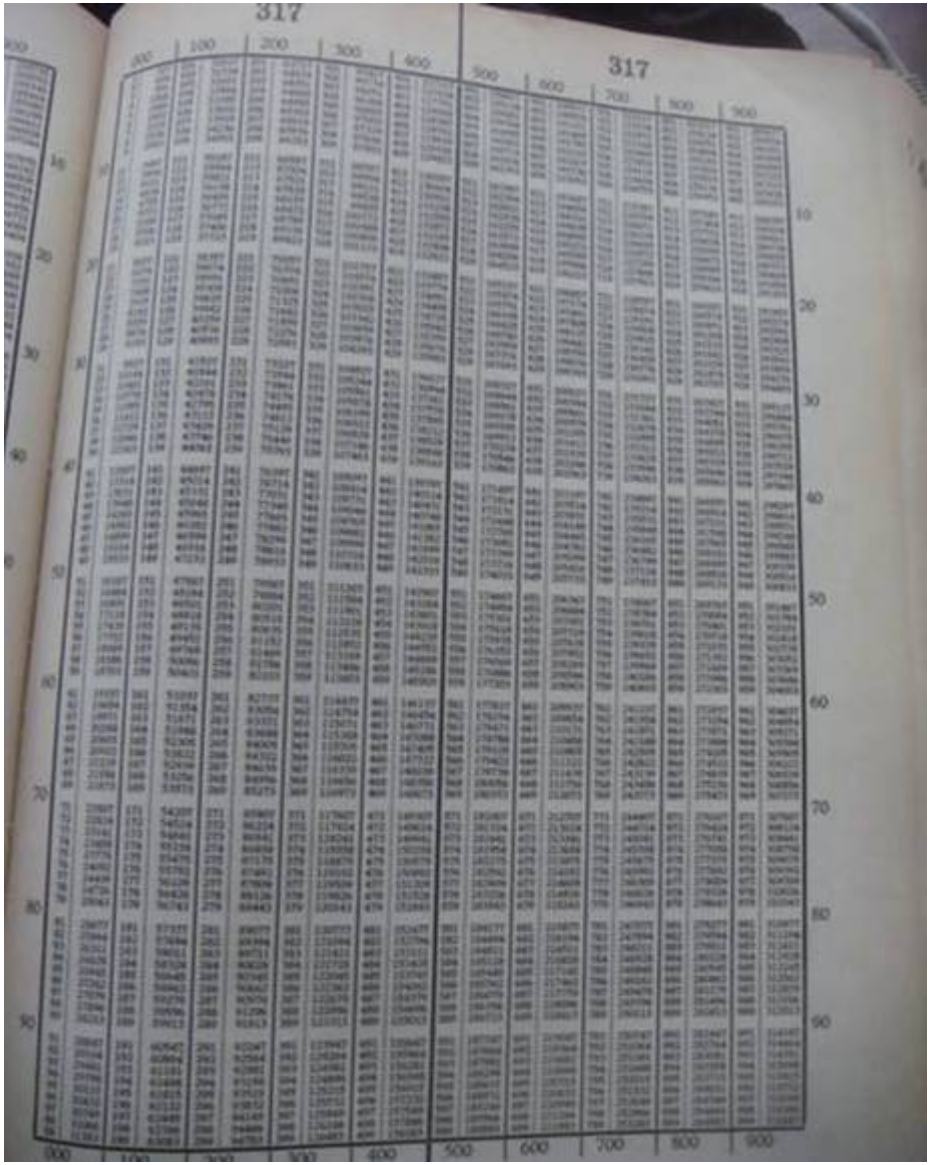
1. What message demands what substrate???
2. Colour dynamic range is still better on film than digital, pace Kodak's bankruptcy
3. Data lists are out of print (telephone directory)
4. Look-up tables

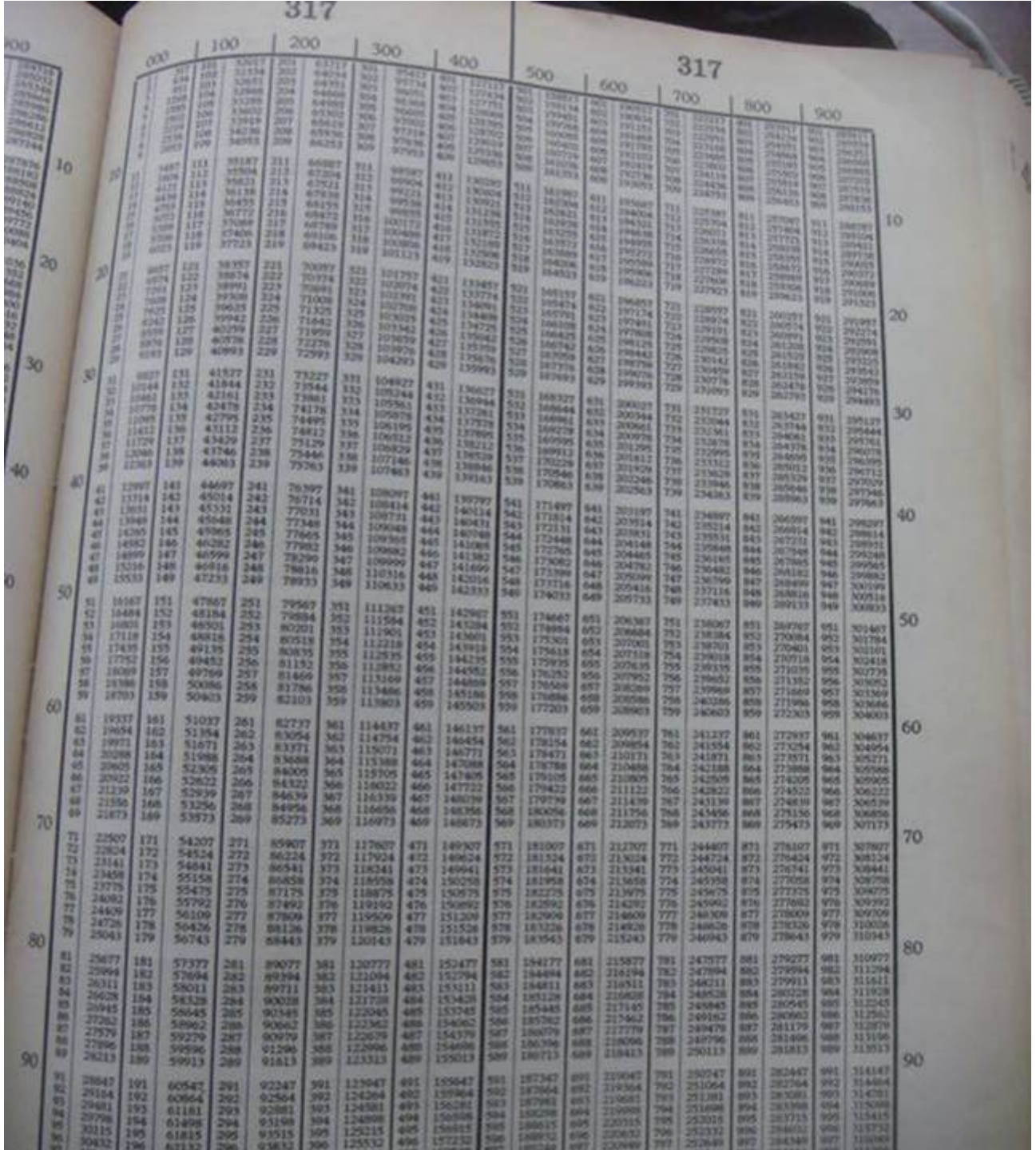
Though! In a paper encyclopedia or dictionary we often also read the adjacent lemmas

=> Aside 8

Aside 8. Paper books







V&C of components of scientific reporting 4c

Aggregation & Dissemination:

Finally: Paper vs. Displays!!

Digitaal Vrouwenlexicon van Nederland

Menu
Home
Nieuw !
Biografieen A-Z
Zoeken & selecteren
Auteurs
Afkortingen
Groslijst
Informatie voor auteurs
Nieuwsbrief (archief)

Titel
Digitaal Vrouwenlexicon van Nederland
Periode
oudheid - heden
Status
Lopend
Projectleider
E.M. Kloek
Betrokkenen
J. Zijlstra
Bijdragen
Een stoet van Vrouwen. Het Digitaal Vrouwenlexicon van Nederland (DVN), van de vroegste tijden tot circa 1850 (PDF)
Rubrieken
Naslagwerken

reactie

Het *Digitaal Vrouwenlexicon van Nederland* (*DVN*) biedt informatie over de opmerkelijkste vrouwen uit de geschiedenis van Nederland en zijn overzeese gebiedsdelen van de vroegste tijden tot nu. In korte biografische schetsen vindt u gegevens over vrouwen die ooit invloedrijk, beroemd of berucht waren, maar nu meestal zo goed als vergeten zijn.

- Nieuw !
- Biografieën A-Z
- Zoeken en selecteren

NB. Niet alle bekende vrouwen staan in het DVN. Zie ook Biografisch Portaal

Inst. for Netherlands History
Digital Women Lexicon
~1450 C&V entries
Web based!

Wordt het DVN een boek?

Ja! Het boek verschijn op 14 februari 2013 (bij Vantilt). Het boek bevat een selectie van ruim duizend lemmata en de titel luidt: *1001 Vrouwen uit de Nederlandse geschiedenis*. Ontwerpster: Irma Boom. Omvang: ca. 1550 blz - zie www.1001-vrouwen.nl

Voor meer informatie:

- Werkwijze
- Organisatie van het project
- Auteurs

HERESE SCHWARTZE, ALEXANDRINE TINNE EN ANNIE M.G. SCHMIDT

001
ROUWEN

GEPRESENTEERD: 14 februari 2013

BEROEMDE, BERUCHTE, OPMERKELIJKE, GELIEFDE, SLECHTE, SPRAAKMAKENDE EN INVLOEDRIJKE VROUWEN.

Gebundeld in een vuistdik naslagwerk

1001 VROUWEN UIT DE NEDERLANDSE GESCHIEDENIS

INT. VROUWENDAG

Drie activiteiten op 8 maart, Internationale Vrouwendag:

1. Lezing van Els Kloek over 'the making of 1001 Vrouwen'.

2. Rondleiding over de expositie.

3. Stadswandeling '1001 Vrouwen'.

Paper vs. Displays 1

What do in-depth tests teach us.

Electronic media

Difficult to search by visual recognition / memory (data driven)

E-Ink reads like paper and therefore, many novels can be read on e-readers, because the flow of reading is defined by the author. Page design is less important.

LCD, etc. screens flicker and back-lit screen weary the eye after intense reading for hours. (Yes we can, we also read form CRTs and accepted Camera ready books)

Paper vs. Displays 2

Paper media

Ideal for long reading, learning, internalizing the content

Leafing and browsing

Note making, scribbles, annotation

Coffee spillage, etc.

Visual inspection when in piles, dog ears

Writing is still faster than typing

Writing is emotionally/mentally

more than punching letters!

Unity of text and visuals

PDF/DjVu/Postscript are Polaroids of Oil paintings.

They act as intermediate transmission medium and are beautiful for printing!

For a report of an in-depth study: Judith Stoop, Paulien Kreutzer, and Joost Kircz: . Reading and learning from screens versus print a study in changing habits. Draft 2012. To be published. And references therein!

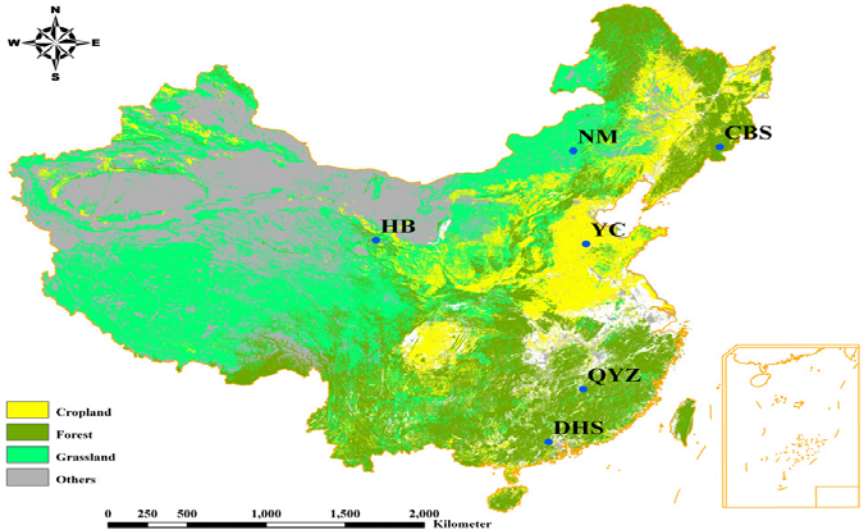


Fig. 1. Distribution of the 6 sites in China at which measured GPP was used for calibrating and validating the TL-LUE model developed in this study. The background is the GLC2000 land cover map.

used for the XLHT site. Both the MODIS GPP and LAI products have a spatial resolution of 1 km. The projection of these data is Sinusoidal, and MRT (MODIS Reprojection Tools) was used to reproject them into an UTM/WGS 84 projection. Because of residual cloud contamination, the MODIS LAI product has some unrealistically abrupt short-term fluctuations, and the locally adjusted cubic-spline capping (LACC) method (Chen et al., 2006) was used to smooth MODIS LAI. The smoothed LAI series were then input into the MOD17 algorithm and the TL-LUE model for calculating *f*PAR.

2.2. Method

2.2.1. The MOD17 algorithm

The MOD17 algorithm is based on the radiation conversion efficiency concept of Monteith (1972). GPP is calculated as (Running et al., 2000):

$$\mathrm{GPP} = \varepsilon_{\max} \times f(\mathrm{VPD}) \times g(T_a) \times \mathrm{PAR} \times f\mathrm{PAR} \quad (3)$$

where *f*PAR is the fraction of PAR absorbed by the canopy and calculated as:

$$f\mathrm{PAR} = 1 - e^{-k \times \mathrm{LAI}} \quad (4)$$

where k is the light extinction coefficient and set as 0.5; LAI is the green leaf area index of the whole canopy.

In Eq. (3), $\varepsilon_{\max}$ is the maximum LUE and changes with vegetation types (Table 2). $f(\mathrm{VPD})$ and $g(T_a)$ are the scalars of VPD and the minimum air temperature (T_a) used to downscale $\varepsilon_{\max}$ to the actual. They are calculated as:

$$f(\mathrm{VPD}) = \begin{cases} 0 & \mathrm{VPD} \geq \mathrm{VPD}_{\max} \\ \dfrac{\mathrm{VPD}_{\max} - \mathrm{VPD}}{\mathrm{VPD}_{\max} - \mathrm{VPD}_{\min}} & \mathrm{VPD}_{\min} < \mathrm{VPD} < \mathrm{VPD}_{\max} \\ 1 & \mathrm{VPD} \leq \mathrm{VPD}_{\min} \end{cases} \quad (5)$$

$$g(T_a) = \begin{cases} 0 & T_a \leq T_{\min} \\ \dfrac{T_a - T_{\min}}{T_{\max} - T_{\min}} & T_{\min} < T_a < T_{\max} \\ 1 & T_a \geq T_{\max} \end{cases} \quad (6)$$

Table 1
Summary of climate and vegetation characteristics of the 6 tower sites.

Sites	Changbaishan	Qianyanzhou	Dinghushan	Yucheng	Haibei	Xinlinhot
Lat/Lon	42°24′N 128°06′E	26°45′N 115°04′E	23°10′N 112°32′E	36°57′N 116°36′E	37°40′N 101°20′E	43°33′N 116°40′E
Climate type	Temperate continental climate influenced by monsoon	Sub-tropical monsoon climate	The monsoon humid climate of torrid zone of south Asia	Semi-humid and monsoon climate	Plateau continental climate	Temperate semiarid continental climate
Annual mean precipitation (mm)	600–900	1489	1956	582	580	350–450
Annual mean temperature (°C)	3.6	18.6	21	13.1	−1.7	−0.4
Vegetation type	Mixed forest	Evergreen	Evergreen	Winter	Alpine meadow	Grassland

Development of a two-leaf light use efficiency model for improving the calculation of terrestrial gross primary productivity
By: Mingzhu Hea, + 11 co-authors
Agricultural and Forest Meteorology
Volume 173, 15 May 2013, Pages 28–39

Conclusions for Publishing 1
Added value for author & reader

A) Take article components as serious as the whole article.
B) Develop standards for non-text elements (XML family)
C) Organize dedicated reviewing for non-text elements (standards & procedures)
D) Develop genre specific metadata (semantics)
E) Develop genre specific hyper-link grammars (see my previous work on typed hyperlinking).
F) Develop understanding of the design of electronic works where text, visuals and sound are integrated.
G) Ensure validation and certification of the above.

Conclusions for Publishing 2
Added value for reading

Make sure that the various content components are collected in a DB, in such a way that reuse is possible and presented in all necessary dedicated substrates.

Data storage: 1) reuse (whole & partly), 2) merging with ditto data. 3) feeding into visualization software,….

Visuals: 1) colour integrity, 2) resolution integrity, 3) overlays, ….

Text: Overview of the flow of reasoning, typed hyper-links (intra & inter), ergonomics of presentation,……

Allow digital for fast browsing, allow analog for internalization by humans.

Experiment as much as you can. Publishing is a craft!

Please follow or disagree, but argue!

CREATE-IT APPLIED RESEARCH
MEDIA, FASHION, IT

University of Applied Sciences Amsterdam

KRA publishing research

www.kra.nl (for all publications)