

Bibliometric analyses on repository contents as a library service for the evaluation of research

Marco van Veller
Wageningen UR Library



Wageningen UR

- Hybrid organization
- Wageningen U(niversity) and R(esearch)

Research entities

- Graduate schools (7 coordinated by Wageningen UR)
- Research departments (5 at Wageningen UR)
 - Animal, Plant, Environmental, Social, Agrotechnology & Food Sciences
- Research institutes (9 at Wageningen UR)
- Chair groups (almost 80 with sometimes more than 1 chair)
- Individual staff

Publishing in science

- One of the main activities of doing scientific research
- Various publication types
 - e.g. articles, conference papers, book chapters, reports
- Measurements
 - Scientific output
 - Productiveness of scientists
 - References to scientific output (Citations)
 - Importance of underlying research activities to peers

Bibliometric analyses

- Bibliometric analyses to evaluate research output
 - e.g. peer reviews, job applications
- Analyses performed by the library
- Only articles in (peer reviewed) scholarly journals
 - Citation statistics in Web of Science
- Aggregation
 - Wageningen UR entities
 - graduate school, department, institute, chair group, staff member
 - Range of years

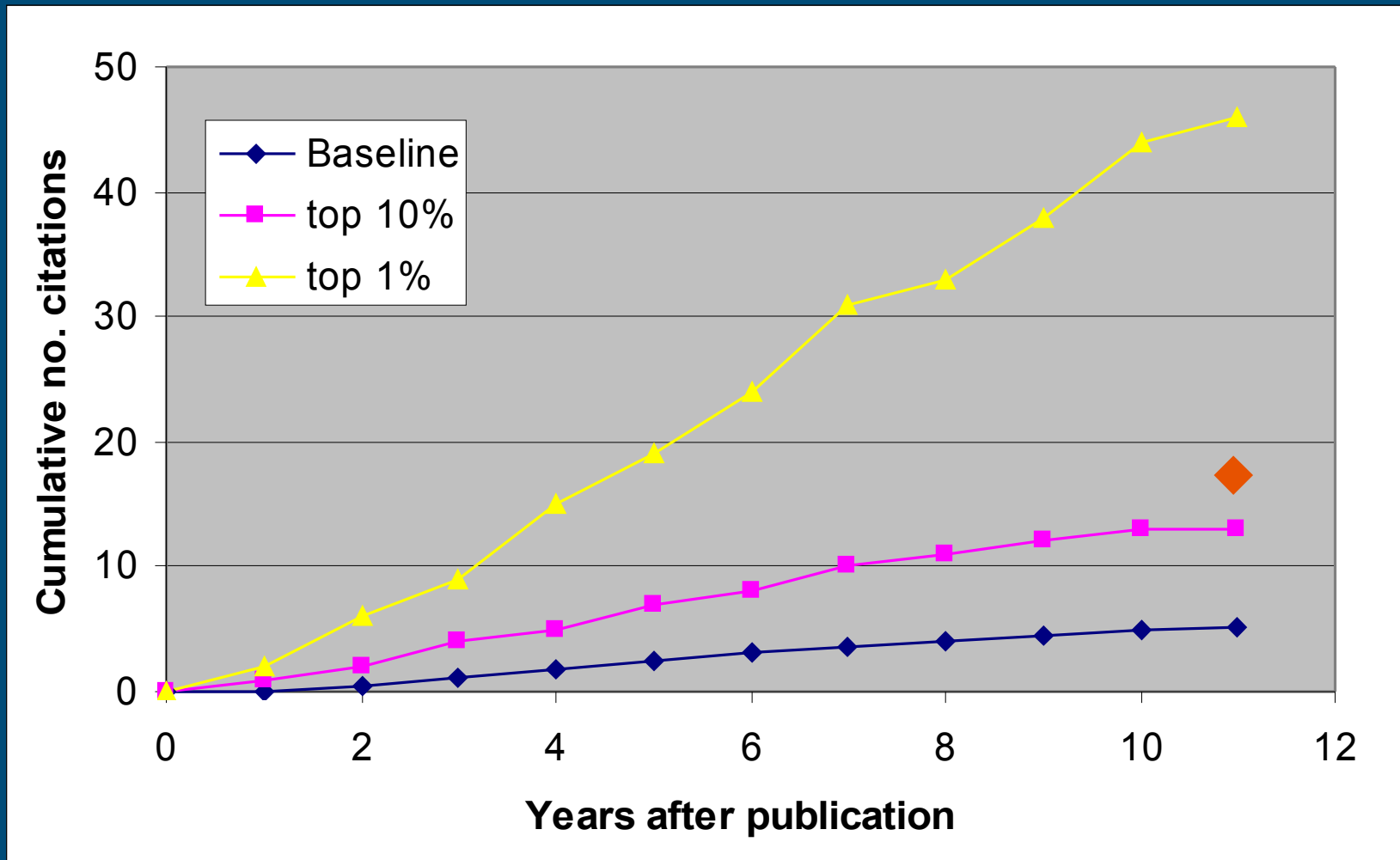
Bibliometric indicators

- Scientific output
 - N = Total number of articles
- Scientific impact
 - C = Total number of citations
- Mean scientific impact
 - CPP = Average number of citations per article = C/N
- However, output and citation behavior can differ considerably per scientific field

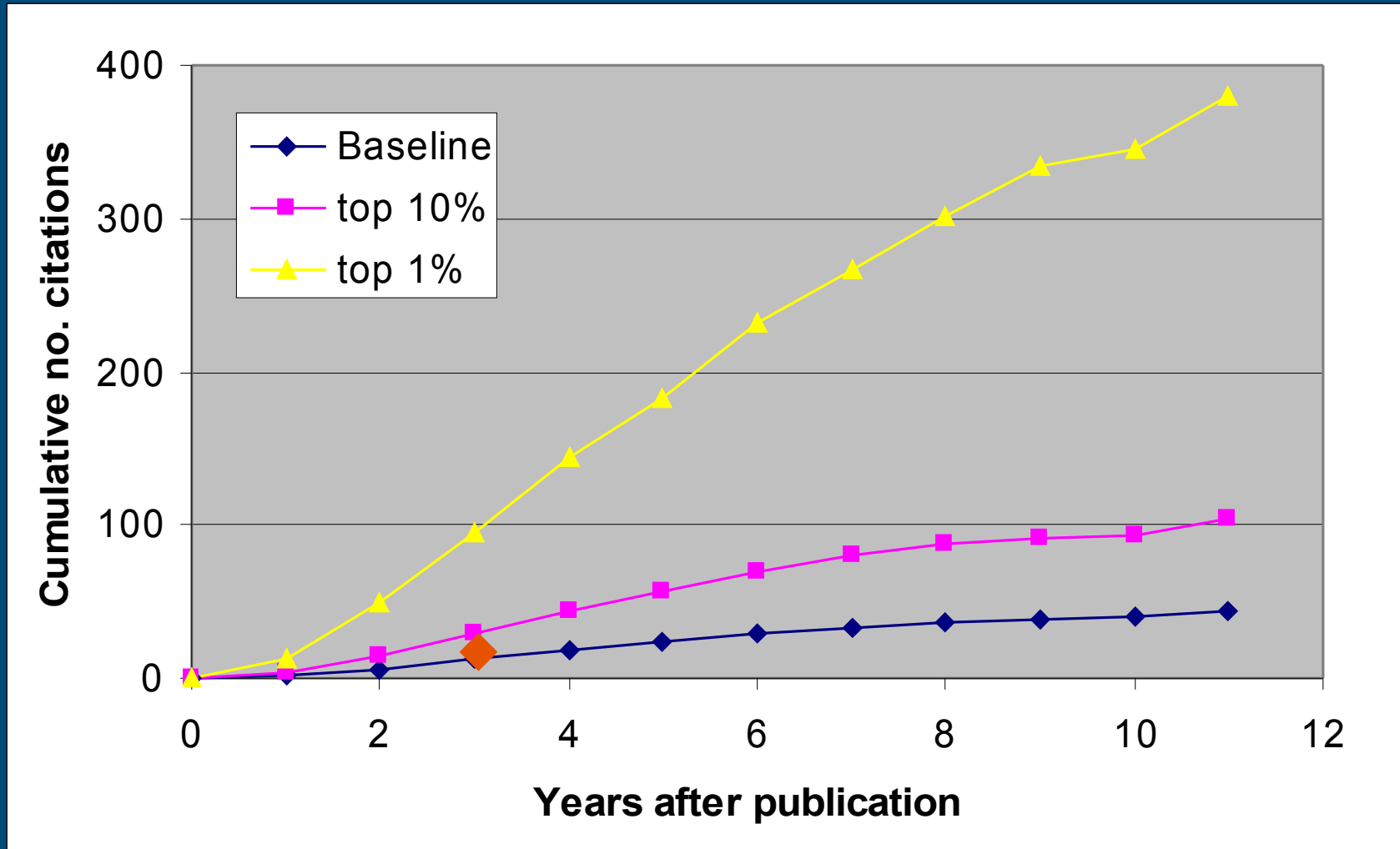
How do we compare numbers?

- Scientist *Z. Math* has a publication from 1998 with 17 citations
- Scientist *M. Biology* has a publication from 2006 with 24 citations

Baseline Mathematics



Baseline Molecular Biology



In practice: Normalization

- Veller, M.G.P. van; Brooks, D.R.; Zandee, R. (2003)
Cladistic and phylogenetic biogeography: the art and the science of discovery. JOURNAL OF BIOGEOGRAPHY, 30: 319-329
 - Times cited: 20 times
- Journal of Biogeography (look up in journals menu ESI)
 - Environment/Ecology
- Baseline data (from Essential Science Indicators)
 - Article from 2003 in Environment/Ecology: Average: 12.49 citations, 10%: 29 citations, 1%: 87 citations
- $RI = 20 / 12.49 = 1.60$

Advanced bibliometric indicators

- Citation Impact (van Raan, 2004) or Relative Impact (Gerritsma, 2006)

- CI and RI

$$CI = \frac{\sum_i^N C_i}{\sum_i^N W_{avg_i}} \quad RI = \frac{\sum_i^N \frac{C_i}{W_{avg_i}}}{N}$$

- Percentage of 1% or 10% most cited articles
 - %T1 and %T10
- Percentage of non-cited articles
 - %NC

Selection of publications – old situation

- Search for publications in Web of Science
 - Wageningen UR entity
 - range of years

Selection of publications – old situation

- Search for publications in Web of Science
- Search based upon author and affiliation data
 - incomplete
 - errors
 - different spellings/synonyms

Example - RIKILT, Institute for Food Safety

- Rijks- Kwaliteitsinstituut voor Land- en Tuinbouwproducten
- National Institute for Quality Control of Agricultural Products
- State Institute for Quality Control of Agricultural Products
- Research Institute for Quality Control of Agricultural Products
- Institute for Quality Control of Agricultural Products

With or without the combination

- Institute for Food Safety

Example – F.W.T. Penning de Vries

- Penningdevries FWT
- De Vries, FWTP
- deVries FWTP
- PenningD.FW

- de Vries, Frits WTP
- Penning de Vries, F.WT
- Penning de Vries, Frits

Selection of publications – old situation

- Search for publications in Web of Science
- Search based upon author and affiliation data
- Download records from Web of Science
 - Bibliographical meta data
 - Author, Year, Title, Volume, Issue, Pages, Journal, ISSN
 - Number of Citations

Selection of publications – old situation

- Search for publications in Web of Science
- Search based upon author and affiliation data
- Download records from Web of Science
- Results checked with authors
 - Errors corrected
 - Missing records downloaded
 - Corrections/additions are not recorded

Selection of publications – old situation

- Search for publications in Web of Science
- Search based upon author and affiliation data
- Download records from Web of Science
- Results checked with authors
- Bibliometric analysis on downloaded records
 - Number of Citations from Web of Science
 - Baselines from Essential Science Indicators
 - Field determined by journal
 - Connection between record and baseline
 - ISSN

Selection of publications – new situation

- Research registration system - Metis
 - Wageningen UR staff
 - Affiliations
 - Work relations
 - Research projects
 - Meta data of publications
 - Data entry by local contacts
 - Local users
 - Secretaries/management assistants
 - Validation by researchers and library staff

Selection of publications – new situation

- Research registration system - Metis
- Repository to scientific output – Wageningen Yield
 - All records from Metis exported to Wageningen Yield
 - Collection of publications by Wageningen UR staff
 - Different levels of aggregation
 - Publication lists
 - Source to select publications for bibliometric analyses

Selection of publications – new situation

- Research registration system - Metis
- Repository to scientific output – Wageningen Yield
- Bibliographic database – Web of Science
 - Contains statistics on citations for articles
 - Search address Wageningen UR
 - Search also for synonyms
 - Download bibliographical records
 - ISI-number
 - DOI
 - Title
 - ISSN, volume, pages

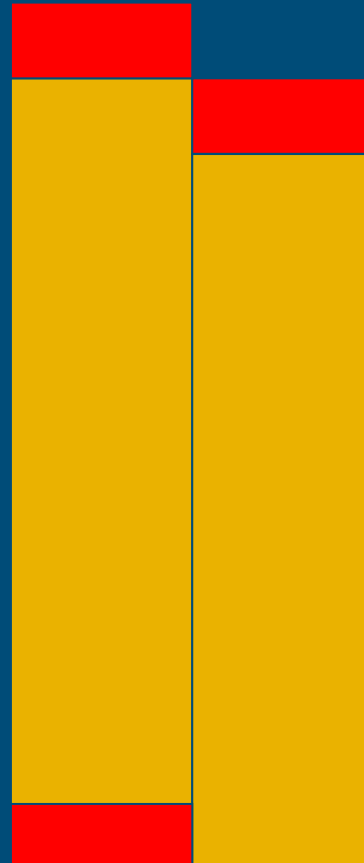
Selection of publications – new situation

- Research registration system - Metis
- Repository to scientific output – Wageningen Yield
- Bibliographic database – Web of Science
- Connection between Metis, Wageningen Yield and Web of Science
 - Match records from Web of Science with records in Metis
 - DOI → ISSN, Volume, Pages → Title
 - Onetime action
 - Import ISI-number in Wageningen Yield

Matching Metis and Web of Science

Metis, 9929 articles

Missing in Metis,
807 articles



Missing from WoS,
1159 articles

WoS, 9577 articles

Selection of publications – new situation

- Research registration system - Metis
- Repository to scientific output – Wageningen Yield
- Bibliographic database – Web of Science
- Connection between Metis, Wageningen Yield and Web of Science
- Retrieve citations from Web of Science
 - To be done for every bibliometric analysis
 - Identification of records by ISI-number
 - Wageningen Yield
 - Tool to do the job

Selection of publications – new situation

- Research registration system - Metis
- Repository to scientific output – Wageningen Yield
- Bibliographic database – Web of Science
- Connection between Metis, Wageningen Yield and Web of Science
- Retrieve citations from Web of Science
- Bibliometric analysis
 - Number of Citations from Web of Science
 - Baselines from Essential Science Indicators

Example – by years

Year	N	C	CPP	CI	RI	%T10	%T1	%NC
2002	10	527	52.7	3.46	4.92	50%	30%	0%
2003	21	545	26.0	1.54	1.89	33%	5%	0%
2004	25	509	20.4	1.71	2.53	28%	8%	0%
2005	25	323	12.9	1.35	1.83	28%	4%	0%
2006	26	319	12.3	2.52	3.22	38%	15%	8%
2007	32	285	8.9	3.73	4.69	38%	19%	13%
All years	139	2508	18.0	2.01	3.10	35%	12%	4%

Example – by research fields

Year	N	C	CPP	CI	RI	%T10	%T1	%NC
Plant & Animal Science	73	1427	19.6	4.08	4.29	49%	19%	4%
Molecular Biology & Genetics	38	686	18.1	1.03	1.17	18%	0%	0%
Biology & Biochemistry	21	365	17.4	1.85	2.69	19%	10%	14%
Environment/Ecology	6	16	2.7	0.54	0.78	0%	0%	0%
Agricultural Sciences	1	4	14.0	12.96	12.96	100%	100%	0%
All research fields	139	2508	18.0	2.01	3.10	35%	12%	4%

Mutual advantages

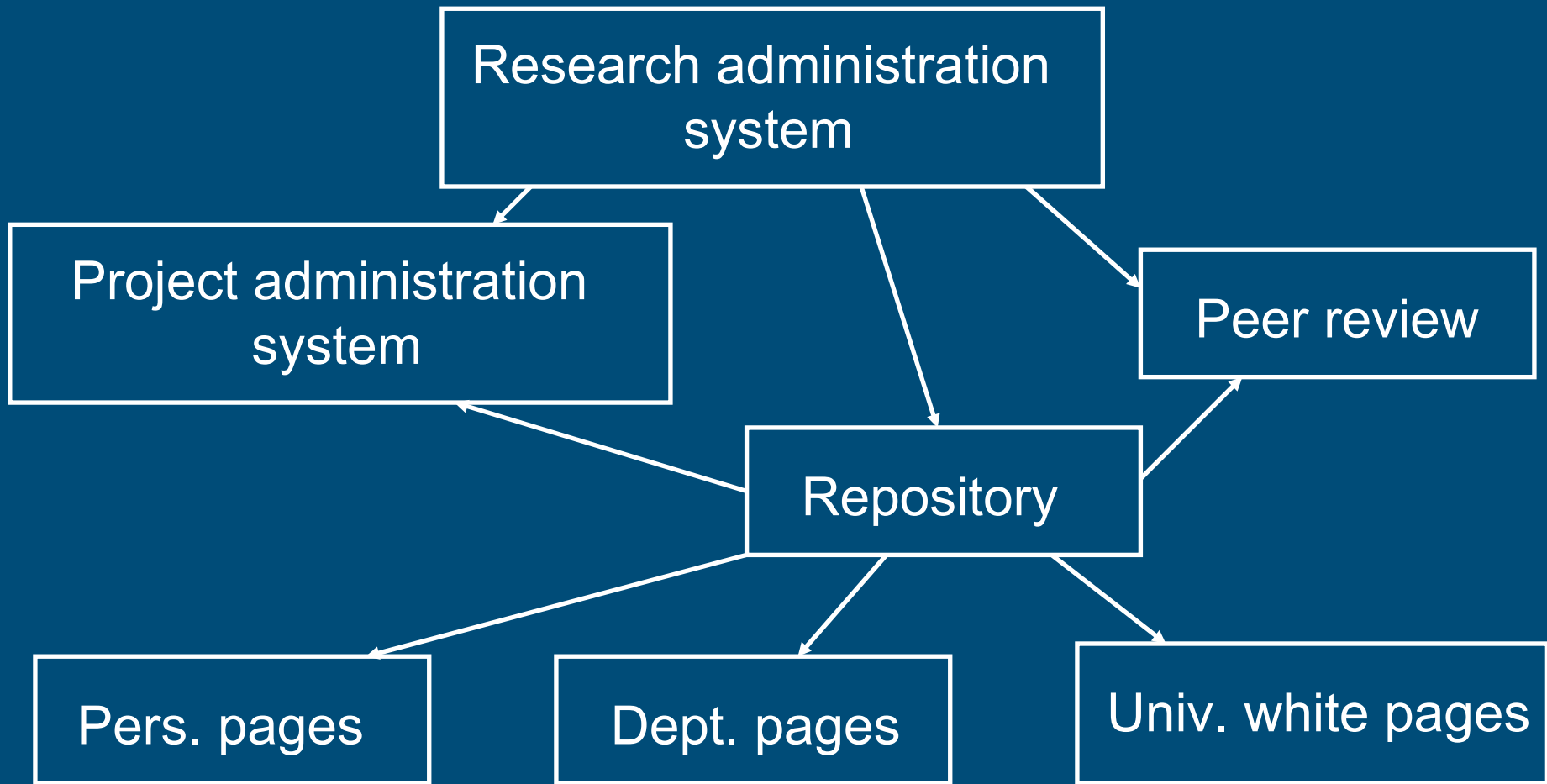
■ Bibliometric analyses

- More precise selection of the right articles
 - Reliable and updated affiliation data and work relations in repository
- Custom-made analyses
 - Repository provides different aggregation levels
- Efficiency
 - Errors in meta data are recorded for future analyses
- Awareness of library services
 - Consultations by library to research for publication strategies

Mutual advantages

- Repository Wageningen Yield
 - Stimulation of registration of new publications
 - Inclusion in bibliometric analyses
 - Various checks on errors
 - At data entry
 - In repository by researcher and library
 - When matching records with Web of Science

Is this the only thing you do with the repository?



Conclusions

- In the beginning the service concentrated on data collection, data entry and calculation of the advanced bibliometric indicators.
- Now, the data entry and calculations are largely automated, and we add value to our services by giving advice and consultations to many groups of researchers
- For publication advice they come to the library!

Thank you !

