How to improve data quality for research assessment in Pure

Marijke Folgering
University of Groningen, the Netherlands
Hi!

Development & Innovation
Overview

› Introduction
› How to improve the quality of data
› How to measure the data entry progress
› How to use Pure for research quality assessment
“Er gaat niets boven Groningen”
University of Groningen

› Founded in 1614
› Top 100 university
11 Faculties

- Arts
- Behavioural & Social Sciences
- Economics & Business
- Law
- Medical Sciences
- Philosophy
- Science & Engineering
- Spatial Sciences
- Theology & Religious Studies
- University College Groningen
- University of Groningen/Campus Fryslân
3,500 fte Academic staff
300 fte Full professors
> 2,000 PhD candidates
800 International PhD students
Pure @University of Groningen

- Implementation in 2014
  - Current Research Information System
  - Repository

- Research portal since 2015
prof. dr. B.L. (Ben L.) Feringa
Professor

Faculty of Science and Engineering

Contact: +31 50 36 34278
+31 50 36 38569 (Secretariat)

E-mail: b.l.feringa@rug.nl

Research

Research units:
Synthetic Organic Chemistry

Overview  Publications  Activities  Prizes  Statistics

Nobel Prize in Chemistry 2016
Bernard Feringa (Recipient), 5-Oct-2016

Tetrahedron Prize 2016
Bernard Feringa (Recipient), 7-Sep-2016
prof. dr. B.L. (Ben L.) Feringa
Professor

Faculty of Science and Engineering

Contact: +31 50 36 34278
        +31 50 36 38569 (Secretariat)

E-mail: b.l.feringa@rug.nl

Research

Research units:
Synthetic Organic Chemistry

Overview Publications Activities Prizes Statistics

All | Journal contributions | Book contributions | Books and reports | Doctoral Thesis

1 - 50 out of 1,040

2018

Pure @University of Groningen

- Cooperation of University of Groningen (UG) & University Medical Center Groningen (UMCG)
## Facts & figures

<table>
<thead>
<tr>
<th>Pure contents</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research output</td>
<td>238,000</td>
</tr>
<tr>
<td>Activities</td>
<td>32,000</td>
</tr>
<tr>
<td>Press/media</td>
<td>6,000</td>
</tr>
<tr>
<td>Prizes</td>
<td>1,500</td>
</tr>
<tr>
<td>Datasets</td>
<td>300</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Research output per year</td>
<td>9,500</td>
</tr>
<tr>
<td>Downloads per month</td>
<td>&gt; 200,000</td>
</tr>
</tbody>
</table>
Pure organization

- Managing board for Research output
- Pure coordinators at the Faculties
- Pure support team and Pure validation team in the University Library
Data quality

> Trust that authors, affiliations and output types are correct
1. Get good data in

› Import from online sources

  or from file

› Create manually?

  • use submission guide
  • or deposit your article
Deposit your article

Use this form to deposit the final author's version of your peer-reviewed article (PDF) in Pure. The UG Library and the Central Medical Library will take care of the registration in Pure, including uploading the publisher's version once it is available. More information about the Policy on 'Green Open Access'.

- P-number or e-mail address *
- Journal name *
- Comments (including your remarks on Open Access/embargo periods)
- Attachment (PDF of the final author's version) *

* = Input is required
2. Show the data

Stimulating data entry (and data corrections!)

› Research portal
› Department websites
› As input for performance interviews
› CV
3. Validation checks

› Library validates
  • Bibliographical metadata
  • Fulltext

› Faculty checks
  • Administrative metadata
    • Affiliation
    • Classification
4. Deduplication

› Publications
› Journals
› Events

› Persons (external)
› Organisations (external)
Something about Matching (1)

What can possibly go wrong?

?
Something about Matching (2)

› What can we do?

- Initially (@implementation)
  - Enrich current metadata
  - Apply own matching rules
  - Apply manual matching of organisations

- Future
  - Use ID’s for matching -> ORCID
  - Use aliases for internal organisations
5. Plausibility checks

› Some logical thoughts

› When a match on author doesn’t seem right
Support staff with > 25 publications
Age at publication date?
Plausibility checks

Publication with > 25 authors

And author *not* from Science & Engineering
Or Medical Sciences
6. “Trawl” actions

Percentage of WOS publications that were not in Pure:

- 2014: 12%
- 2015: 16%
- 2016: 7%
How to measure the data entry progress?
How to use Pure for Research Quality Assessment
Dutch Frameworks

SEP = Standard Evaluation Protocol

KUOZ = Key figures for University Research (Onderzoek)
## Req’s part I: research output

### Table D3b Main categories of research output

<table>
<thead>
<tr>
<th>Research unit</th>
<th>#</th>
<th>#</th>
<th>#</th>
<th>#</th>
<th>#</th>
<th>#</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refereed articles</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Non-refereed articles (1)</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Books</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Book chapters</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>PhD theses</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Conference papers</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Professional publications (2)</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Publications aimed at the general public (3)</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Other research output &lt;specify&gt; (4)</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td><strong>Total publications</strong></td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>
SEP category can be deducted

1. Pure type
2. Publication category
3. Peer reviewed
Interpretations...

Patent – always professional or can be academic?

Special issue in Contribution to journal

Refereed article?

Other research output?
<table>
<thead>
<tr>
<th>Type</th>
<th>Sub type</th>
<th>Publication category</th>
<th>Peer review</th>
<th>SEP output type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book/Report</td>
<td>Book</td>
<td>Academic</td>
<td>-</td>
<td>Books</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professional</td>
<td>-</td>
<td>Professional Publications</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Popular</td>
<td>-</td>
<td>Other Research output</td>
</tr>
<tr>
<td></td>
<td>Book Editing</td>
<td>-</td>
<td>-</td>
<td>Other Research output</td>
</tr>
<tr>
<td></td>
<td>Report</td>
<td>Professional</td>
<td>-</td>
<td>Professional Publications</td>
</tr>
</tbody>
</table>
XML to import in Pure

<!-- Books -->
- <ResearchOutputMappingRule>
  <!-- input -->
  <researchOutputTypeUri>/dk/atira/pure/researchoutput/researchoutputtypes/book</researchOutputTypeUri>
  <researchOutputCategoryUri>/dk/atira/pure/publication/category/academic</researchOutputCategoryUri>
  <!-- output -->
  <mappedResearchOutputTypeUri>/dk/atira/pure/assessmenttype/book</mappedResearchOutputTypeUri>
</ResearchOutputMappingRule>

- <ResearchOutputMappingRule>
  <!-- input -->
  <researchOutputTypeUri>/dk/atira/pure/researchoutput/researchoutputtypes/book</researchOutputTypeUri>
  <researchOutputCategoryUri>/dk/atira/pure/publication/category/professional</researchOutputCategoryUri>
  <!-- output -->
  <mappedResearchOutputTypeUri>/dk/atira/pure/assessmenttype/professional_publication</mappedResearchOutputTypeUri>
</ResearchOutputMappingRule>
Results

new fields available for reporting
... and shown in SEP/KUOZ module

<table>
<thead>
<tr>
<th>SEP output type</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refereed articles</td>
<td>5249</td>
<td>5283</td>
<td>5490</td>
<td>5662</td>
<td>5381</td>
<td>3060</td>
</tr>
<tr>
<td>Non-refereed articles</td>
<td>283</td>
<td>158</td>
<td>167</td>
<td>160</td>
<td>142</td>
<td>67</td>
</tr>
<tr>
<td>Books</td>
<td>162</td>
<td>131</td>
<td>95</td>
<td>96</td>
<td>93</td>
<td>30</td>
</tr>
<tr>
<td>Book chapters</td>
<td>666</td>
<td>675</td>
<td>691</td>
<td>656</td>
<td>660</td>
<td>284</td>
</tr>
<tr>
<td>PhD theses</td>
<td>435</td>
<td>444</td>
<td>506</td>
<td>535</td>
<td>501</td>
<td>376</td>
</tr>
<tr>
<td>Conference papers</td>
<td>247</td>
<td>308</td>
<td>271</td>
<td>254</td>
<td>215</td>
<td>65</td>
</tr>
<tr>
<td>Professional publications</td>
<td>940</td>
<td>885</td>
<td>860</td>
<td>803</td>
<td>677</td>
<td>264</td>
</tr>
<tr>
<td>Publications aimed at the general public</td>
<td>157</td>
<td>130</td>
<td>128</td>
<td>131</td>
<td>93</td>
<td>42</td>
</tr>
</tbody>
</table>
For selection:
- **Period** = 2018
- **SEP output type** = Books

+ default filter:
- **Origin** = With internal affiliation
- **State** = Published
Req’s part 2: FTE’s and funding

<table>
<thead>
<tr>
<th>Research unit</th>
<th>Year 5</th>
<th>Year 4</th>
<th>Year 3</th>
<th>Year 2</th>
<th>Year 1</th>
<th>Current year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific staff(1)</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
</tr>
<tr>
<td>Post-docs (2)</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
</tr>
<tr>
<td>PhD students (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total research staff</strong></td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
</tr>
<tr>
<td>Support staff</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
</tr>
<tr>
<td>Visiting fellows</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
</tr>
<tr>
<td><strong>Total staff</strong></td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
<td>/ FTE</td>
</tr>
<tr>
<td>Research unit</td>
<td>Year 5</td>
<td>Year 4</td>
<td>Year 3</td>
<td>Year 2</td>
<td>Year 1</td>
<td>Current year</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>Funding:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct funding ¹</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
</tr>
<tr>
<td>Research grants ²</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
</tr>
<tr>
<td>Contract research ³</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
</tr>
<tr>
<td>Other ⁴</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
</tr>
<tr>
<td><strong>Total funding</strong></td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
<td>FTE / %</td>
</tr>
</tbody>
</table>
# Import file for FTE’s and funding

| ID | Classified ID type | Classified ID | Source ID type | Source ID | Name       | Org       | Org name    | Job title       | Employment type | Year | FTE | % Direct funding | % Research grants | % Contract research | % Other |
|----|--------------------|---------------|----------------|------------|------------|-----------|------------|----------------|----------------|-----------------|------|-----|-----------------|-------------------|-------------------|--------|
| 2  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | professor        | academic          | 2016 | 0.50 | 100              |                   |                   |        |
| 3  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | professor        | academic          | 2017 | 0.50 | 100              |                   |                   |        |
| 4  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | professor        | academic          | 2018 | 0.50 | 100              |                   |                   |        |
| 5  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | asst_Professor   | academic          | 2016 | 0.25 | 100              |                   |                   |        |
| 6  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | asst_Professor   | academic          | 2017 | 0.25 | 100              |                   |                   |        |
| 7  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | asst_Professor   | academic          | 2018 | 0.25 | 100              |                   |                   |        |
| 8  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | assoc_Professor  | academic          | 2016 | 0.25 | 100              |                   |                   |        |
| 9  | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | assoc_Professor  | academic          | 2017 | 0.25 | 100              |                   |                   |        |
| 10 | employee           | 12345         |                |            | Kalin, Bob | 6791      | Civil Engineering | assoc_Professor  | academic          | 2018 | 0.25 | 100              |                   |                   |        |
## Bernard Feringa

**Person**

### Research FTE and Funding

#### Synthetic Organic Chemistry

**Scientific Staff – Professor**

**SEP/KUOZ Staff type: Full professor**

<table>
<thead>
<tr>
<th></th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FTE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% Direct Funding</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% Research Grants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% Contract Research</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% Other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results in the SEP/KUOZ module
Results in the SEP/KUOZ module
In summary

› To improve data quality
  • Facilitate getting good data in
  • Show the data
  • Use ID’s
  • Validate and check

› Keep track of data entry progress

› “Enrich” your reporting data to ease essential reports

Think out of the box 😊
Thanks!

› Marijke Folgering
› m.e.folgering@rug.nl