Support tools for the VQR Italian Research Assessment Exercise: the Sapienza Experience

Gartner European Education Forum: Universities, Innovation and ICT

Rome April 18, 2013

Camil Demetrescu, Marco Schaerf
Dept. Computer, Control and Management Engineering
Outline

• The Italian VQR research assessment exercise
• The Sapienza experience
• Results
• Conclusions
The VQR National Research Assessment Exercise

- In 2012, public Italian universities and research centers have participated in a major research assessment exercise (VQR)

- **Goal:**
  inform selective funding allocation

- **Coverage:**
  research products published in 2004-2010

- **Evaluation:**
  mix of peer-review and bibliometrics

- **Main challenge for universities & RC:**
  choosing a selection of the best products to submit
VQR in a nutshell (1/2)

• Each researcher/faculty member submitted up to 3 of her/his best products published in 2004-2010

• No duplicate submissions: each product selected by at most one coauthor of the same institution

• Evaluation done by 14 panels (GEV)

• Different evaluation criteria for each panel

• Reference databases:
  – Thomson Reuters Web of Science (WoS) [main]
  – Elsevier Scopus [additional]

• For each submitted product, institutions had to choose:
  – a specific evaluation panel to evaluate the product
  – a subject category
VQR in a nutshell (2/2)

• Mandatory pieces of information to submit:
  – Meta-data (title, authors, etc.)
  – Full text (pdf)
  – Abstract
  – ISSN (journals)
  – ISBN (other publications)

• **Outcome of the evaluation**: a numeric score for each submitted product

• **Total score of the institution** = sum of scores of submitted products (will **determine part of the funding allocation for next years**)

## VQR grades and scores

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Grade</th>
<th>Score</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>A</td>
<td>+1</td>
<td>Publication rank in the top 20%</td>
</tr>
<tr>
<td>Good</td>
<td>B</td>
<td>+0.8</td>
<td>Publication rank in the range 60%-80%</td>
</tr>
<tr>
<td>Fair</td>
<td>C</td>
<td>+0.5</td>
<td>Publication rank in the range 50%-60%</td>
</tr>
<tr>
<td>Limited</td>
<td>D</td>
<td>0</td>
<td>Publication rank in the bottom 50%</td>
</tr>
<tr>
<td>Missing</td>
<td>-</td>
<td>−0.5</td>
<td>Expected submission slot for the institution that is not filled with an eligible publication</td>
</tr>
<tr>
<td>Not eligible</td>
<td>-</td>
<td>−1</td>
<td>Publication not eligible for submission to the evaluation exercise (e.g., wrong category, missing full text attachment, older than 2004, etc.)</td>
</tr>
<tr>
<td>Plagiarism</td>
<td>-</td>
<td>−2</td>
<td>Publication proven to have plagiarized or to be fraud</td>
</tr>
</tbody>
</table>
Products eligible for evaluation

• Articles in **journals** with ISSN

• **Books, book chapters**, and conference **proceedings** papers with ISBN

• Critical editions, translations, scientific comments

• Deposited patents

• Compositions, drawings, design, performance, exhibits and organised exposures, artifacts, prototypes and artworks and their projects, databases and software, and thematic maps (provided that they are supported by accompanying publications)
# VQR evaluation panels (GEV) for Subject Areas

<table>
<thead>
<tr>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area 1</td>
<td>Mathematics and Informatics</td>
</tr>
<tr>
<td>Area 2</td>
<td>Physics</td>
</tr>
<tr>
<td>Area 3</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Area 4</td>
<td>Earth Sciences</td>
</tr>
<tr>
<td>Area 5</td>
<td>Biology</td>
</tr>
<tr>
<td>Area 6</td>
<td>Medicine</td>
</tr>
<tr>
<td>Area 7</td>
<td>Agricultural and Veterinary Sciences</td>
</tr>
<tr>
<td>Area 8</td>
<td>Civil Engineering and Architecture</td>
</tr>
<tr>
<td>Area 9</td>
<td>Industrial and Information Engineering</td>
</tr>
<tr>
<td>Area 10</td>
<td>Antiquities, philology, literary studies, art history</td>
</tr>
<tr>
<td>Area 11</td>
<td>History, philosophy, pedagogy and psychology</td>
</tr>
<tr>
<td>Area 12</td>
<td>Law</td>
</tr>
<tr>
<td>Area 13</td>
<td>Economics and Statistics</td>
</tr>
<tr>
<td>Area 14</td>
<td>Political and Social Sciences</td>
</tr>
</tbody>
</table>
Evaluation criteria

• Hard sciences:
  – Citations
  – Impact factor or Scopus SJR
  – Lists of area-specific journal rankings (A, B, C, D)
  – Informed peer-review (IR)
  – Peer review for non-journal articles

• Soft sciences: peer review

• Countless details:
  – Different evaluation for survey articles
  – Different thresholds for different panels, etc. etc.
### Example: GEV 03 (Chemistry)

#### 2004-2008

<table>
<thead>
<tr>
<th>Citations grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>IR</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>IR</td>
</tr>
<tr>
<td>B</td>
<td>IR</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>IR</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>

#### 2009-2010

<table>
<thead>
<tr>
<th>Citations grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>IR</td>
<td>IR</td>
<td>IR</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>IR</td>
<td>IR</td>
<td>IR</td>
<td>D</td>
</tr>
</tbody>
</table>

**Example**: article published in **2005**, with:
- Citations grade A (top 20%)
- Impact Factor grade C (top 50%)

Overall grade: A
Score: +1
### Example: GEV 03 (Chemistry)

**2004-2008**

<table>
<thead>
<tr>
<th>Citations grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>IR</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>IR</td>
</tr>
<tr>
<td>C</td>
<td>IR</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>IR</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>

**2009-2010**

<table>
<thead>
<tr>
<th>Citations grade</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>IR</td>
<td>IR</td>
<td>IR</td>
<td>IR</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>IR</td>
<td>IR</td>
<td>IR</td>
<td>D</td>
</tr>
</tbody>
</table>

**Example**: article published in **2010**, with:

- Citations grade **A** (top 20%)
- Impact Factor grade **C** (top 50%)

Informed peer-review
VQR Timeline

- November 7, 2011: call for participation published
- February 29, 2012: (incomplete) evaluation criteria published
- June 15, 2012: product submission deadline for institutions

- Selection process for institutions: 3 months (+ 2 weeks last-minute extension)
Outline

- The Italian VQR research assessment exercise
- **The Sapienza experience**
- Results
- Conclusions
Sapienza in a nutshell

• One of the largest universities in Europe
• 129,500 students in 2010, 1st in Europe, 43rd in the world as number of students
• One of the oldest in Italy, founded in the 14th century
• Over 4,000 researchers from 63 departments
• 21 museums and more than 50 libraries
• Research catalog including 250,000 publications
• ~75,000 considered for the VQR
Selection approach

- **Top-down**: central coordination for all departments based on a sofware system especially designed for the VQR
- **Goal**: use *optimization algorithms* to maximize the expected total score of Sapienza
- **Same product** may have **different scores** depending on:
  - **Panel** to which the product is submitted
  - **Subject category** in which the product is classified
- Our *software simulated all possible* panel/subject category combinations, computing the expected score
- **Human validation** selected “reasonable” combinations
### Example: journal article in physics

<table>
<thead>
<tr>
<th>Expect. grade</th>
<th>Chosen panel</th>
<th>Relev.</th>
<th>Chosen subject category</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>C B C B C C A</td>
<td>02 02 03 03 04 07 07</td>
<td>3 3 0 0 0 0 0</td>
<td>METEOROLOGY &amp; ATMOSPHERIC SCIENCE</td>
<td>Scopus</td>
</tr>
</tbody>
</table>

**“Reasonable” choice**

**Maximization choice**

- **Physics**
- **Agricultural and Veterinary Sciences**

- **Reasonable choice**
  - MEETEROLOGY & ATMOSPHERIC SCIENCES
  - ATMOSPHERIC SCIENCE

- **Maximization choice**
  - MEETEROLOGY & ATMOSPHERIC SCIENCES
  - ATMOSPHERIC SCIENCE
Surviving “big” data…

• **Problem:** manually choosing the best reasonable panel/subject category combination for all eligible products would have been overwhelming!

• **Our solution (for hard sciences):**

1. Initial **automatic choice of tentative panel/subject category** to each journal article based on a maximum “relevance” metric we designed
   => yields initial “tentative” grade for each article

2. Automatic selection of the best 3 and 6 products for each author based on the tentative grades

3. **Manual validation of selected products only**

4. Optimization algorithm re-executed every night
### VQRselect Web interface

- **Best 3 products per author**
- **Best 6 products per author**
- **All eligible products**
- **Excluded products**

#### Faculty members and products assigned to them

<table>
<thead>
<tr>
<th>Peso</th>
<th>Titolo</th>
<th>Autori Sapienza</th>
<th>Autori</th>
<th>SSD</th>
<th>Area</th>
<th>Dipartimento</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>Wild Ungulates as Babesia Hosts in Northern and Central Italy.</td>
<td>CANCRINI Gabriella</td>
<td>TAMPIERI MP; GALUPPI R; BONOLI C; G. CANCRINI; MORETTI A; PIETROBELLI M</td>
<td>07</td>
<td>VET/06</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>Initial experience with magnetic resonance fluoroscopy in the evaluation of oesophageal...</td>
<td>TOMEI Ernesto</td>
<td>PANEBIANCO V; HABIB FI; E. TOMEI; ET AL</td>
<td>06</td>
<td>MED/36</td>
<td></td>
</tr>
</tbody>
</table>
Manual validation of panel/subject category combination

Scelta non effettuata

Selezione classe di merito, area VQR, SC, ASJC

Classi di merito: A (18)

Area VQR: 11: SCIENZE STORICHE, FILOSOFICHE, PEDAGOGICHE E PSICOLOGICHE [M-EDF, M-PSI] - 80 (18)

Categoria ISI/Scopus: PSYCHOLOGY(ALL) [Scopus] (0)

Altra categoria ISI/Scopus: 60 (B) - PSYCHOLOGY, MULTIDISCIPLINARY [ISI] (0)

Prodotto escluso dalla valutazione

Salva
Optimization algorithm

Author 1
Slot 1
Slot 2
Slot 3

Product 1 +1
Product 2 +0.8
Product 3 0
Product 4 +0.8
Product 5 0
Product 6 +0.5
Product 7 +1

Author 2
Slot 1
Slot 2
Optimization algorithm

Author 1
- Slot 1
- Slot 2
- Slot 3

Product 1: +1
Product 2: +0.8
Product 3: 0
Product 4: +0.8
Product 5: 0
Product 6: +0.5
Product 7: +1

Author 2
- Slot 1
- Slot 2
Optimization algorithm

Author 1

- Slot 1
- Slot 2
- Slot 3

Product 1: +1
Product 2: +0.8
Product 3: 0
Product 4: +0.8
Product 5: 0
Product 6: +0.5
Product 7: +1

Author 2

- Slot 1
- Slot 2
Critical aspects (1/2)

- Extremely tight time frame for selecting the research products
- Large-scale coordination: 63 departments
- (Incomplete) evaluation criteria known 3.5 months until the submission deadline
- Different evaluation criteria for different panels
- Critical data not publicly available (e.g., thresholds for determining if a product is in the top 20% ecc.)
Critical aspects (2/2)

• Extensive data quality problems in our research catalog:
  – Duplicates
  – Wrong classification (e.g., proceedings as Article)
  – Missing or wrong fields
  – Missing coauthors
  – Missing or wrong codes (DOI, PUBMED, ISBN, ….)

• Data quality problems also in WoS and Scopus (e.g., incorrect subject categories)
Sapienza timeline (3.5 months)

- **Phase 0:** March 1 – April 11 [42 days]
  - **What:** VQRselect software development
  - **Who:** Sapienza publications group + Exaltech Srl

- **Phase 1:** April 12 – May 6 [25 days]
  - **What:** product selection
  - **Who:** department heads

- **Phase 2:** May 7 – June 22 [16 days]
  - **What:** additional info, upload of PDFs
  - **Who:** faculty members, department heads

- **Phase 3/4:** May 23 – June 15 [24 days]
  - **What:** linking with WoS/Scopus, error corrections
  - **Who:** VQR task force
Timeline of product selection

Number of selected products

-166  Phase 1
-19  Phase 2
-42  Phase 3
-49  Phase 4

April 12  May 7  May 23  June 1  June 15
Outline

- The Italian VQR research assessment exercise
- The Sapienza experience
- Results
- Conclusions
Selected products over 92% of expected

Selected (10019) 92.4%

Missing (823) 7.6%
Selected products: soft vs. hard sciences

- Hard sciences: 56%
- Soft sciences: 44%
% selected products by type

- Journal article: 074%
- Book chapter: 013%
- Monograph: 008%
- Conference Proceedings: 005%
- Curatorship: 000%
- Patent: 000%
- Other: 000%
Estimated scores for submitted journal articles (hard sciences)

- A/B: 2%
- A: 55%
- B: 17%
- B/C: 4%
- C: 7%
- C/D: 4%
- D: 11%
Conclusions

- Sheer size of Sapienza, large number of products, data quality issues, incomplete evaluation criteria, and short time frame made the process extremely critical
- Top-down approach, using IT methods
- Optimization algorithms used to maximize the expected score of Sapienza
- IT infrastructure was crucial for the success of the process
- Role of IT for research assessment will increase in the future
Thanks