

RESEARCH EVALUATION IN SSH: GOALS, CHALLENGES, LESSONS LEARNED

MARCO MALGARINI – MARILENA MANIACI

What should 21st century cross-disciplinary research evaluation look like?

EASSH Workshop – February 23, 2024

OUTLINE

- 1. Evaluating research in Italy**
 - Evaluation of Institutions
 - Career evaluation and journal classification in SSH
 - Accreditation and periodic assessment of Phd programmes
- 2. The evaluation of research quality (VQR): objectives and methods**
- 3. Towards the new VQR 2020-2024: critical issues and lessons learned (especially – but not only - for the evaluation of the SSH)**

RESEARCH EVALUATION IN ITALY

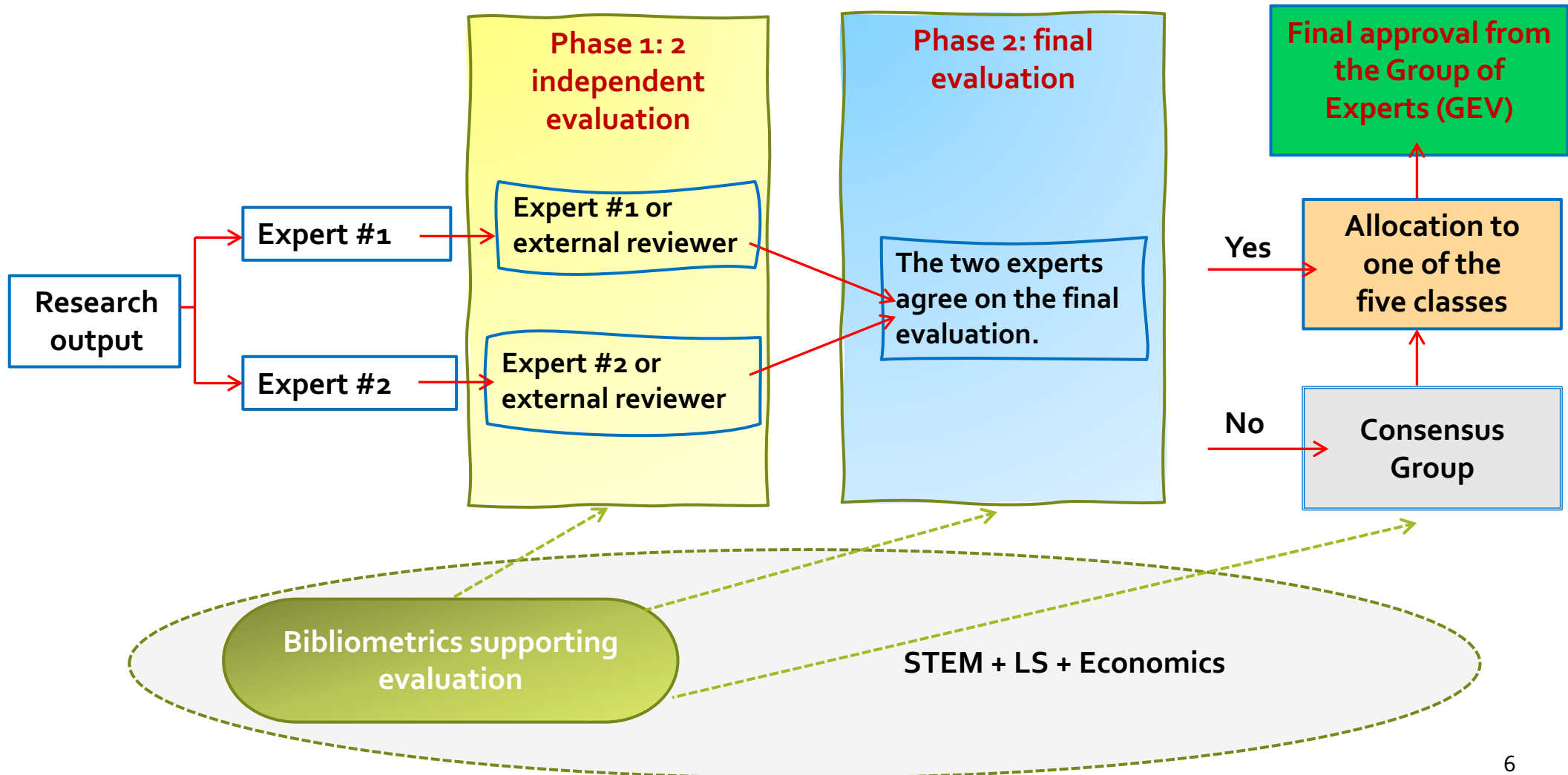
- ✓ **Main lines of ANVUR's activity in the field of research evaluation**
 - Evaluation of the **quality of research conducted by institutions – VQR** (universities, research institutions, other entities voluntarily undergoing evaluation)
 - Initial and periodic **accreditation of PhD programs**
 - Evaluation activities related to the **National Scientific Qualification**
 - Scientific qualification of evaluation committees' members
 - Classification of scientific and Class-A journals in SSH
- ✓ **In these evaluation activities, ANVUR uses peer review supported by indicators, if appropriate to the field**

PARTICIPATION OF ANVUR IN THE INTERNATIONAL DEBATE ON RESEARCH ASSESSMENT

- ✓ **CoARA** – ANVUR is among the first members of **CoARA – Coalition for Advancing Research Assessment**, and one of the Agency's Governing Board members sits in the Steering Board of the coalition.
- ✓ **AGORRA** – ANVUR participates in the activities of **AGORRA – A Global Observatory of Responsible Research Assessment**
- ✓ **ENQA WG on QA of Research** – ANVUR is one of the **14 QA Agencies** that participate in the **ENQA WG on QA of Research**, aimed at exploring the state of the art of the assessment of research-based learning, institutions' research policy and quality assurance processes of research

THE EVALUATION OF RESEARCH QUALITY (VQR): OBJECTIVES AND METHODS

THE VQR – EVALUATION PROCEDURE



THE IMPACT OF RESEARCH EVALUATION ON THE FINANCING OF ITALIAN HEI (1)

- ✓ VQR results are used by the Ministry of Universities and Research for **allocating the performance-based share of the main university funding** (FFO – *Ordinary Financing Fund*). For 2023 the premium share amounts to **EUR 2.5 billion**
- ✓ **80%** of the competitive share is assigned according to **VQR results**, of which
 - **60% on the basis of the quality of research of all the researchers**
 - 90% of which on the basis of research quality
 - 5% of which on the basis of research training
 - 5% of which on the basis of “third mission” activities
 - **20% on the basis of recruitment policies**
- ✓ **20%** is assigned based on other factors (teaching activities; internationalisation; research environment)

THE IMPACT OF RESEARCH EVALUATION ON THE FINANCING OF ITALIAN HEI (2)

- ✓ VQR results are also used by the Ministry of Universities and Research for **selecting the 180 excellent departments** of Italian universities that will obtain extra financial support for 5 years (varying between EUR 1,620 and 1,080 million euros annually for five years)
- ✓ The initiative is aimed at supporting innovation in Italian universities
- ✓ The selection process is made up of two steps:
 1. ANVUR assesses **the departments' performance** and ranks the **best 350 departments** according to their ***Standardised Indicator of Departmental Performance (ISPD)*** score. The ISPD is calculated on the basis of the **VQR** ₈ **outcomes**.
 2. A committee appointed by MIUR **selects the best 180 departments** based on the **ISPD** (weight 70%) and a **strategic development program** proposed by the department (weight 30%)

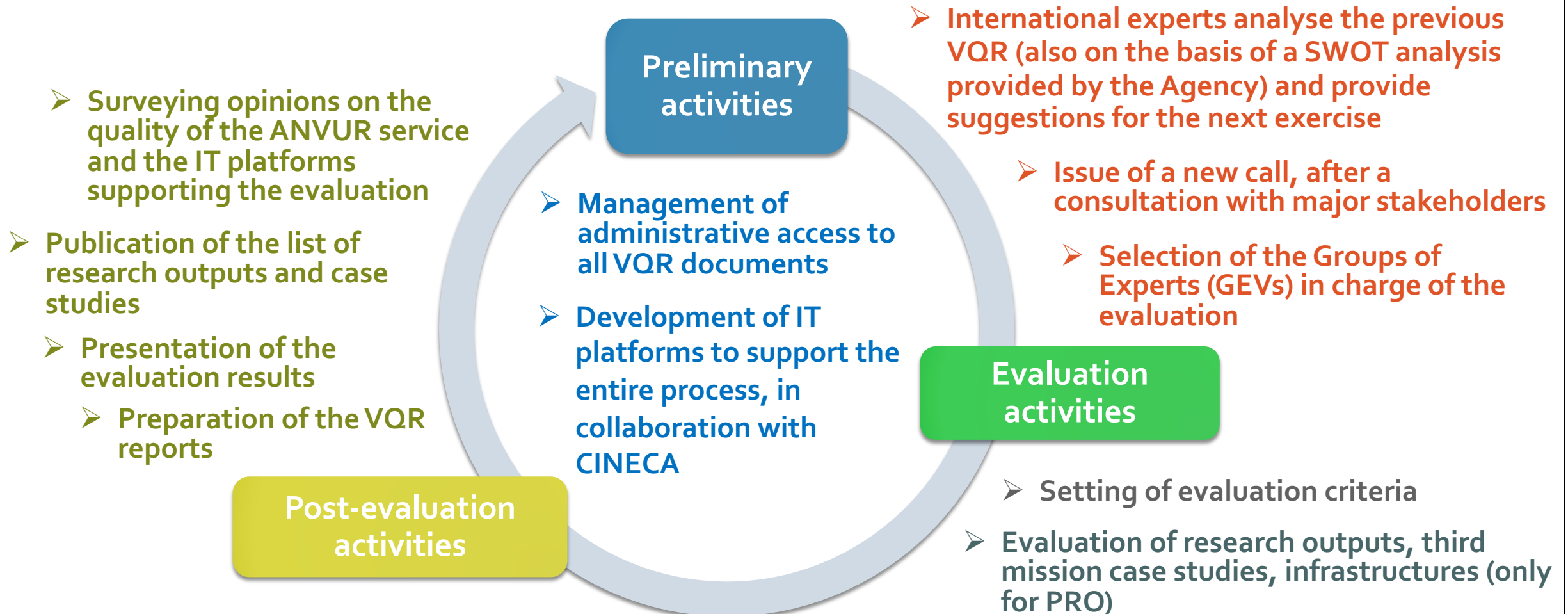
THE IMPACT OF RESEARCH EVALUATION ON THE INTERNAL QUALITY MANAGEMENT OF INSTITUTIONS

VQR results may help HEIs and research institutions in several ways:

- ✓ **Data-driven decision making** – The results of the VQR can help determine the strengths and weaknesses of research areas within the institution, leading to informed strategic planning
- ✓ **Setting benchmarks and goals** – Research evaluation helps in setting benchmarks and goals for future performance, aligning them with the institution's overall strategic objectives
- ✓ **Resource allocation** – Evaluation outcomes can influence where institutions choose to allocate their resources, such as funding, personnel, and infrastructure
- ✓ **Faculty development and recruitment** – Evaluations can identify gaps in expertise or performance, shaping faculty development programs and informing recruitment strategies to enhance institutional research capabilities
- ✓ **Enhancing collaboration and competitiveness** – Policymaking can be influenced to promote interdisciplinary research and collaboration if evaluations indicate these are areas leading to impactful research outcomes
- ✓ **Fostering innovation** – Insights from research evaluations can help reveal emerging trends in research and lead to policies that encourage innovation, such as by supporting experimental methodologies or novel areas of inquiry that could yield significant benefits

CRITICAL ISSUES AND LESSONS LEARNED (SSH vs OTHER RESEARCH FIELDS)

THE VQR CYCLE



POSSIBLE BIAS IN RESEARCH EVALUATION: AN OVERVIEW

The internal self-assessment report following the completion of the VQR 2015-19 identified **two main sources of bias** potentially affecting reviewers' evaluation:

➤ **Author bias**, related to

- ✓ Gender
- ✓ Status (full professors vs other categories)
- ✓ Type of institution (Schools of advanced Studies vs State universities vs non-State universities vs online universities)
- ✓ Geographical location
- ✓ Seniority (newly recruited vs. stable researchers)
- ✓ Scientific discipline (STEM+LS vs SSH)

➤ **Publication bias**, related to

- ✓ Type of publication (articles vs book chapters vs monographs vs other publications and outputs)
- ✓ Language of publication
- ✓ Open access availability

AUTHOR BIAS

With respect to possible sources of author bias, we find that:

- ✓ **Gender** – The probability of obtaining an excellent evaluation is **not influenced** by gender
- ✓ **Status** – The probability of obtaining an excellent evaluation is **significantly higher for full professors** than for other categories
- ✓ **Type of institution** – The probability of obtaining an excellent evaluation is **significantly higher for professors at Schools of advanced Studies** than for those at State universities, and for the latter compared to the non-State non-online universities
- ✓ **Geographical location** – The probability of obtaining an excellent evaluation is **significantly higher for professors at universities in the North-West and North-East** than for those in the Centre, and for the latter than for those in the South and the Islands
- ✓ **Seniority** – The probability of obtaining an excellent evaluation is significantly higher for professors who have been recruited or promoted during the five-year assessment period than for permanent researchers
- ✓ **Scientific discipline** - The probability of obtaining an excellent evaluation is significantly higher for researchers working in STEM and life sciences fields (where bibliometrics are used to support peer review) compared to those in SSH (who adopt “pure” peer review). Moreover, in some SSH fields (in particular, Social sciences and Economics), further differences may stem from controversies between academic schools or scholarly communities that can produce biased evaluations

PUBLICATION BIAS

With respect to possible sources of publication bias, we find that:

- ✓ **Type of publication** – The probability of obtaining an excellent evaluation is **higher for journal articles** than for other types of publication, with the only exception of monographs, which have a more favourable assessment even compared to journal articles. A significant difference emerges between STEM-*like* and SSH fields, since in the latter monographs are more likely to receive an excellent evaluation than journal articles, whereas the opposite is true in the STEM-*like* fields
- ✓ **Language of publication** – In SSH fields only, the probability of obtaining an excellent evaluation is **higher for products written in English** than for those in languages other than English and Italian, and for the latter than for those written in Italian
- ✓ **Open access availability** – In SSH fields, outputs available in open access are **less likely** to obtain an excellent evaluation

SSH fields show a different behaviour regarding publication language and open access availability

NOISE IN REVIEWERS' EVALUATION

Evaluation may also be affected by noise, in the sense that different reviewers may provide different assessments on the same research output

With respect to noise, it is possible to observe that:

- ✓ **Correlation among reviewers is higher** when all the evaluations are performed by experts that are also members of the GEVs
- ✓ **Correlation among “internal” and “mixed” reviewers is usually higher** (with some exceptions) in **SSH (in yellow)** than in **STEM+LS**

GEV	Correlations rev1 vs rev2 (external)	Correlations rev1 vs rev2 (internal)	Correlations rev1 vs rev2 (mixed)
1	0,220	0,702	0,327
2	0,195	0,653	0,287
3	-0,033	0,392	0,124
4	0,904	0,793	0,224
5	0,996	0,682	0,191
6	0,635	0,690	0,353
7	0,161	0,708	0,160
8a	0,247	0,876	0,312
8b	0,248	0,655	0,387
9	0,384	0,733	0,394
10	0,280	0,627	0,362
11a	0,290	0,495	0,303
11b	0,432	0,850	0,441
12	0,251	0,788	0,433
13a	0,377	0,803	0,315
13b	0,456	0,913	0,583
14	0,236	0,729	0,405
All	0,310	0,722	0,392

THE REPORT OF THE INTERNATIONAL EXPERT GROUP

- ✓ Following the internal SWOT analysis, ANVUR appointed a group of international experts to provide an external assessment of the VQR 2015-19
- ✓ In their report, the experts stated that **assessment in SSH fields deserves special attention**. In particular, the **increased diversity of research outputs** in these disciplines should be taken into account in the assessment
- ✓ Main experts' recommendations included the needs for
 - strengthening the effectiveness of the evaluation process, by implementing **targeted measures to enhance reviewers' selection**
 - establishing **training sessions for panels members**, in order to ensure that the purpose, procedures, and desired output of the evaluation is well understood
 - **normalizing scores** according to disciplinary practices
 - exploring the **use of AI tools** to facilitate the assignment of experts to research outputs
 - encouraging the **submission of different research outputs** beyond standard publications
 - encouraging fair **evaluation of research outputs regardless of the language** in which they are produced

TOWARDS THE NEW VQR 2020-2024 (1)

Following the SWOT analysis and the international experts' report, the main innovations in the new *Call for the VQR 2020-2024* include

- ✓ A **wider definition of scientific output** to be considered in the evaluation, responding to the international recommendation for a proper recognition of the diversity of research activities and practices
- ✓ A **new formulation of evaluation criteria concerning the scientific methodology**, with the goal of promoting, where applicable, the reproducibility of the results, transparency with regard to the methods and procedures adopted, in order to exploit the entire process that led to the realization of the research product
- ✓ A **strengthening of the importance of peer review evaluation**, emphasizing that the use of indicators can never substitute for a proper peer evaluation of outputs
- ✓ A formal commitment for the **evaluation to depend exclusively on research quality**, regardless of the language or the type and place of publication
- ✓ A formal commitment to a **mandatory training phase aimed at experts and reviewers** to promote full and consistent application of evaluation rules

TOWARDS THE NEW VQR 2020-2024 (2)

Possible improvements will also consist in:

- ✓ building up a **database of experts**, possibly also taking into account past performance in VQR peer review activities (average time per evaluation, rejection rate, etc.)
- ✓ **exploiting AI tools** to obtain a better match between research outputs to be evaluated and appropriate reviewers

THE VQR TEAM

THE ANVUR BOARD

Proff. Antonio Uricchio (President);
Alessandra Celletti (Vice-President);
Marilena Maniaci;
Menico Rizzi;
Massimo Tronci.

Director General Dr. Daniele Livon

Administrative Manager

Dr. Valter Brancati

Research Evaluation Manager:

Dr. Marco Malgarini

Evaluation officers Dr.:

Brigida Blasi

Paola Costantini

Vittorio Leproux

Francesca Macrì

Carmen Nappi

Irene Mazzotta

Francesca Pentassuglio

Sandra Romagnosi

Scipione Sarlo

Cristiano Trani (CINECA)

• **CINECA**, technical assistance, dr:

Michele Avellino

Pierluigi Bonetti

Roberto Gori

Giulio Racale

GEV	Coordinatore
Area 1 - Mathematics and Computer Sciences	Prof. Giovanni Federico Gronchi
Area 2- Physics	Prof. Sabino Matarrese
Area 3 . Chemistry	Prof. Roberto Paolesse
Area 4 - Earth Sciences	Prof. Massimiliano Barchi
Area 5 – Biology	Prof.ssa Valeria Poli
Area 6 – Medicine	Prof. Alessandro Padovani
Area 7 - Agricultural and veterinary sciences	Prof.ssa Stefania De Pascale
Area 8a - Architecture	Prof. Alessandro Balducci
Area 8b - Civil Engineering	Prof. Marco Marani
Area 9 - Industrial and Information Engineering	Prof.ssa Sara Rainieri
Area 10 - Ancient History, Philology, Literature and Art History	Prof. Carlo Giovanni Cereti
Area 11a - History, Philosophy, Pedagogy	Prof.ssa Lina Scalisi
Area 11b – Psychology	Prof.ssa Rosalinda Cassibba
Area 12 – Law	Prof.ssa Marina Brollo
Area 13a - Economics and Statistics	Prof.ssa Emanuela Marrocu
Area 13b- Economics and Management	Prof.ssa Maria Rosaria Napolitano
Area 14 - Political and Social Sciences	Prof. Maurizio Ambrosini
Impact/Third mission	Prof. Sauro Longhi

Thank you!

marco.malgarini@anvur.it
marilena maniaci@anvur.it