INFORMATICS EUROPE

Informatics Research Evaluation – Revised Report

ECSS 2024 Workshop, Malta 30 October 2024

IE/NIA Research Evaluation Recommendations Panel

Background

The first Informatics Europe recommendations on Research Evaluation appeared as an IE report in 2008 and as an article in CACM in 2009 (Meyer, Choppy, Staunstrup, van Leeuwen).

A major update was prepared in 2018 (Esposito, Ghezzi, Hermenegildo, Kirchner Ong).

The present revision builds upon the two previous ones and updates them, particularly in view of:

- the broadly recognised *CoARA Agreement on Research Assessment (2022)*,
- five currently topical areas: responsible use of bibliometrics, credit assignment in contributions, assessing artefacts, Open Science, interdisciplinary research, and
- the emerging role of AI in research evaluation.

Outline

- Research evaluation for quality and impact
- Characteristics of Informatics
- Responsible use of indicators
- Credit assignment in contributions
- Assessing artefacts
- Open Science
- Interdisciplinary research
- The role of AI in research evaluations
- Executive summary: Key messages

Research evaluation for quality and impact

The fundamental goal of research evaluation is to assess the quality and impact of research, for the eventual improvement of both.

Quality is an elusive intrinsic characteristic for which a commonly accepted assessment method, even if imperfect, is peer review by a panel of informed experts.

Impact is an observable external characteristic that takes many forms and can to some extent be measured by numerical indicators, but even then only with human expert interpretation.

Quality is mostly a good predictor of impact, and impact is mostly a good indicator of quality, but the two are not coextensive.

Research assessment should primarily assess quality and impact over quantity.

CoARA: "Focus research assessment criteria on quality [and] recognise the contributions that advance knowledge and the (potential) impact of research results."

Characteristics of Informatics

Informatics is a relatively **young science** that is **rapidly evolving** in close connection with technology.

An important characteristic of Informatics is the creation of **artefacts**.

Informatics research is very **methodologically diverse**.

Informatics has an **extremely high societal and economic impact**.

Informatics research, as any other science, must be evaluated according to criteria that **take into account its specificity.**

In the Informatics publication culture, **conference publications have a prominent role**. Journals do not necessarily carry more prestige than conferences.

However, new alternatives are emerging that bridge this dichotomy, e.g. **coupled conferences and journals**, **open archives** and **overlay journals.**

Responsible use of indicators

We all know that bibliometrics (citation count, h-index, impact factor…) are **here to stay**.

Most of us also admit having a look at these numbers as a **first proxy** for classifying a researcher, department, publication venue, etc.

We observe that bibliometrics are **increasingly used**, often tacitly, in **internal** evaluations that claim to rely on peer review.

However, indicators can be manipulated, thus:

- Publication counts **must not** be used to evaluate **research value**!
- Numerical impact measurements (citation counts…) **must not** be used **in isolation** – but must be interpreted by humans!
- They **must not** be used to compare researchers across **different fields**, nor within subfields of Informatics!

CoARA:
"Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators."

Credit assignment in contributions

Assess the individual contributions in multi-author publications!

In Informatics, the order of authorship **rarely reflects** the level of contribution – however, also Informatics researchers face the necessity to produce publications with 'primary authorship' roles (e.g., for publication-based PhDs)

→ researchers should **clearly specify each author's contribution** in their publications and other scientific artefacts

E.g., the CRediT taxonomy (credit.niso.org) classifies contribution roles as

> *Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing*

- **All** contributions should be listed (from authors, from other contributors)
- **Individual contributors** can be assigned **multiple roles**
- A **single role** can be assigned to **multiple contributors**
- The **degree of contribution** can optionally be specified as 'lead,' 'equal,' or 'supporting' (https://credit.niso.org/implementing-credit/)

Assessing artefacts

- Source code, data, models... outcomes of research
  - Used by authors to demonstrate research practicality
  - Used by other researchers to build upon
  - Also by industry if robust enough!
  - Linked to Open Science, reproducibility, …
- Artefacts: often focus on conference submisssion, tied to paper
  - Life afterwards? (documentation, evolution, maintenance...)
  - "Quality badges" do not assess long-term impact
- Outcomes of quality (for e.g. software) require extra efforts, often detracted from usual research resources
- Plus additional caveats: evolution, longevity; authorship; impact measurement; temporal context; industry collaborations

Assessing artefacts: recommendations

- Avoid bean counting
- No penalty for lack of releases
- Reward only research-driven advances
- Recognize long-term projects
- Consider researcher role
- Don't consider short contributions
- Accumulate value, even in different systems
- Contributions to components
- Weigh software releases in isolation
- Best practices for citing products
- Public funding produces public software
- Specific evaluation criteria

Open Science

Open Science concerns a plethora of different dimensions which aims at **incentivising quality in science** and **recognising the diversity of research outputs, activities and missions**

The scholarly community should **retain control and ownership over Open Science infrastructures and services** (e.g. OpenCitations, OpenAIRE, Software Heritage, DBLP) providing the data (i.e. **open research information**) used for devising metrics and indicators that may be used to support peer-review assessment, and those implementing transparent assessment, e.g. by enabling applying **open peer review** evaluation practices (PREreview)

Informatics:

- should **acknowledge Open Science practices** in its research evaluation
- has a prominent role to play in the adoption and development of the **Open Science approaches and infrastructures**, and its support is key to keep them sustainable in the long term

Interdisciplinary research

Status: Informatics has become an important support provider for research in other fields. There are also more ambitiously integrated collaborations, where the Informatics contribution is a core element of the research agenda.

Challenges:

- In interdisciplinary work, credit is often assigned along disciplinary lines, so that recognition goes primarily to the substance area, and also on the Informatics side the contribution is considered "an application".
- In assessments that are based mechanically on publication venues and/or indicators, Informatics contributions to other areas than core Informatics are easily overlooked.

Recommendations:

- Recognise the value of inter/transdisciplinary research in its own terms, not as an "application" of Informatics.
- Assess (i) the depth of the integration and (ii) the novelty and significance of the Informatics contribution to the totality of the work.
- Be wary of numerical indicators and "top venue" lists oriented towards assessing Informatics disciplinary work.

The role of AI in research evaluation

The use of generative AI is rising and we expect it to be used in research evaluations, too.

To make the best of the situation, we have to use AI in a responsible way.

Recommendations towards this end include:

- The use of generative AI in research evaluations should be communicated openly.
- Decisions should still be made by human experts, and the use of AI should be restricted to the lower levels of the decision process.
- Efforts should be made to verify critical data obtained from a computer.
- The use of AI must not be used to reduce the number of human experts in decision panels and their responsibility.

Executive summary: key messages

1. Informatics is an original discipline that combines aspects of mathematics, science, and engineering. Researcher evaluation must recognise and respect its specificity.
2. A distinctive feature of the publication culture in Informatics is the importance of highly selective conferences.
3. Open archives and overlay journals are recent innovations that offer improved tracking in evaluation.
4. The impact of artefacts such as software, open datasets, and other research products such as trained machine learning models can be as great as publications.
5. Open Science and its research evaluation practices are highly relevant to Informatics.
6. Numerical measurements (such as citation and publication counts) must never be used as the sole evaluation instrument.
7. In Informatics, the order of authors often holds little significance and varies across subfields.
8. In assessing institutions, researchers, publications and citations, the use of open research information provided by Open Science infrastructures should be favoured and supported.
9. Any evaluation, especially quantitative, must be based on clear, published criteria.