



Assessing Creative and Scientific Commons

Bronwyn H. Hall
Maastricht University and
University of California at
Berkeley

[The problem]

- How should we evaluate the performance of the scientific and creative commons?
 - What do we mean by performance?
 - Why do we want to know?

Creative and scientific commons

■ Examples:

- Free and open source software
- Open access scientific and engineering preprint databases
- Databases - common use licensing of data contributed to repositories
- Cross-licensing of patented research tools, materials transfer licensing on RAND terms
- Most of the above have some form of contract associated with them, implied or otherwise

[Performance]

- Multi-dimensional:
 - Ease of submission, updating
 - Ease of access, use, search
 - Comprehensiveness
 - Accuracy and quality
 - In some cases, the ability to use the contents for statistical purposes
 - “efficiency” - better outcomes at lower cost

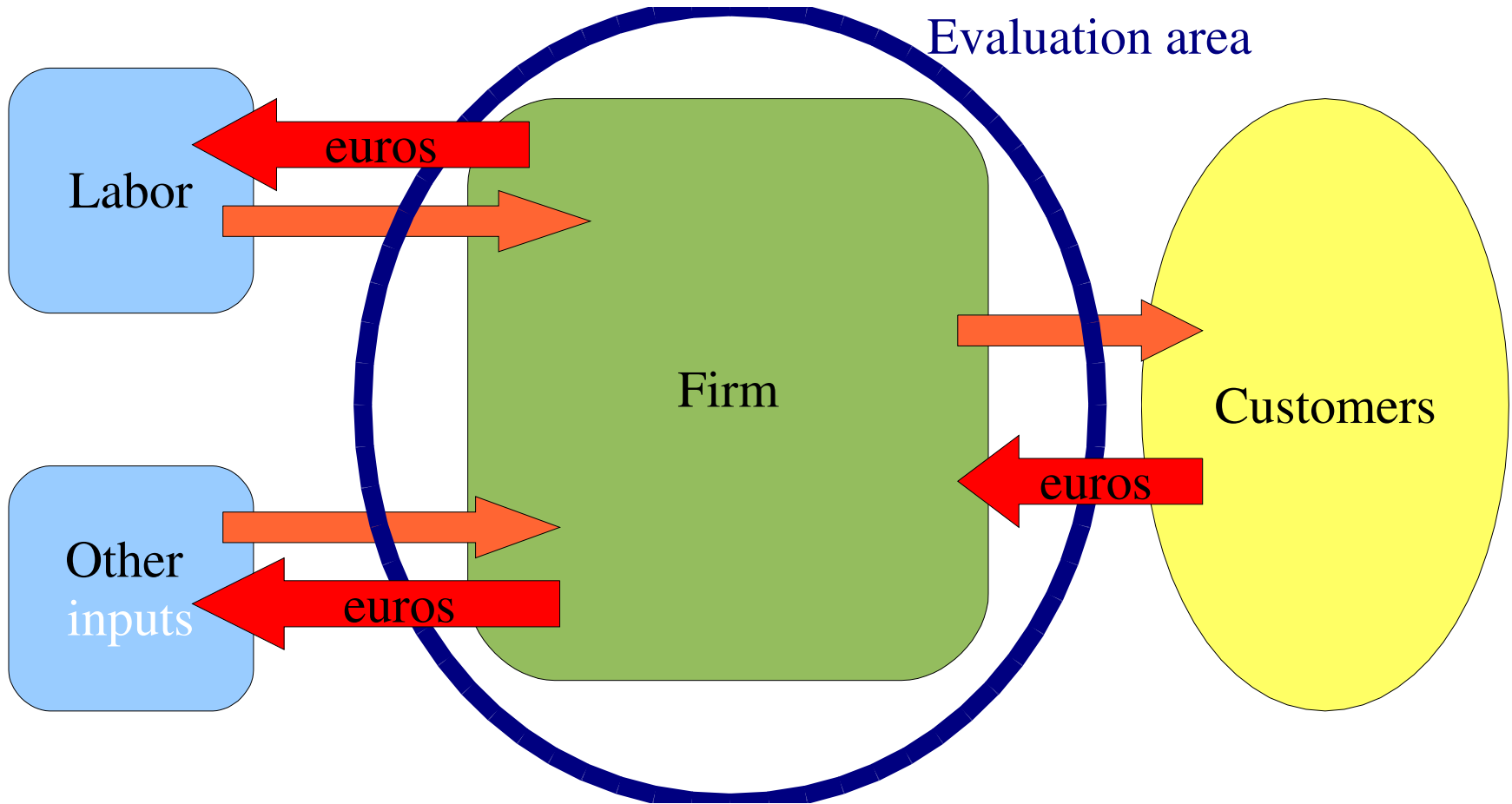
[Why do we want to know?]

- To compare and evaluate different methods of organizing such commons
- To allocate funds to help in the provision and maintenance of such commons
- To establish best practices in organizing commons
- ...other reasons?

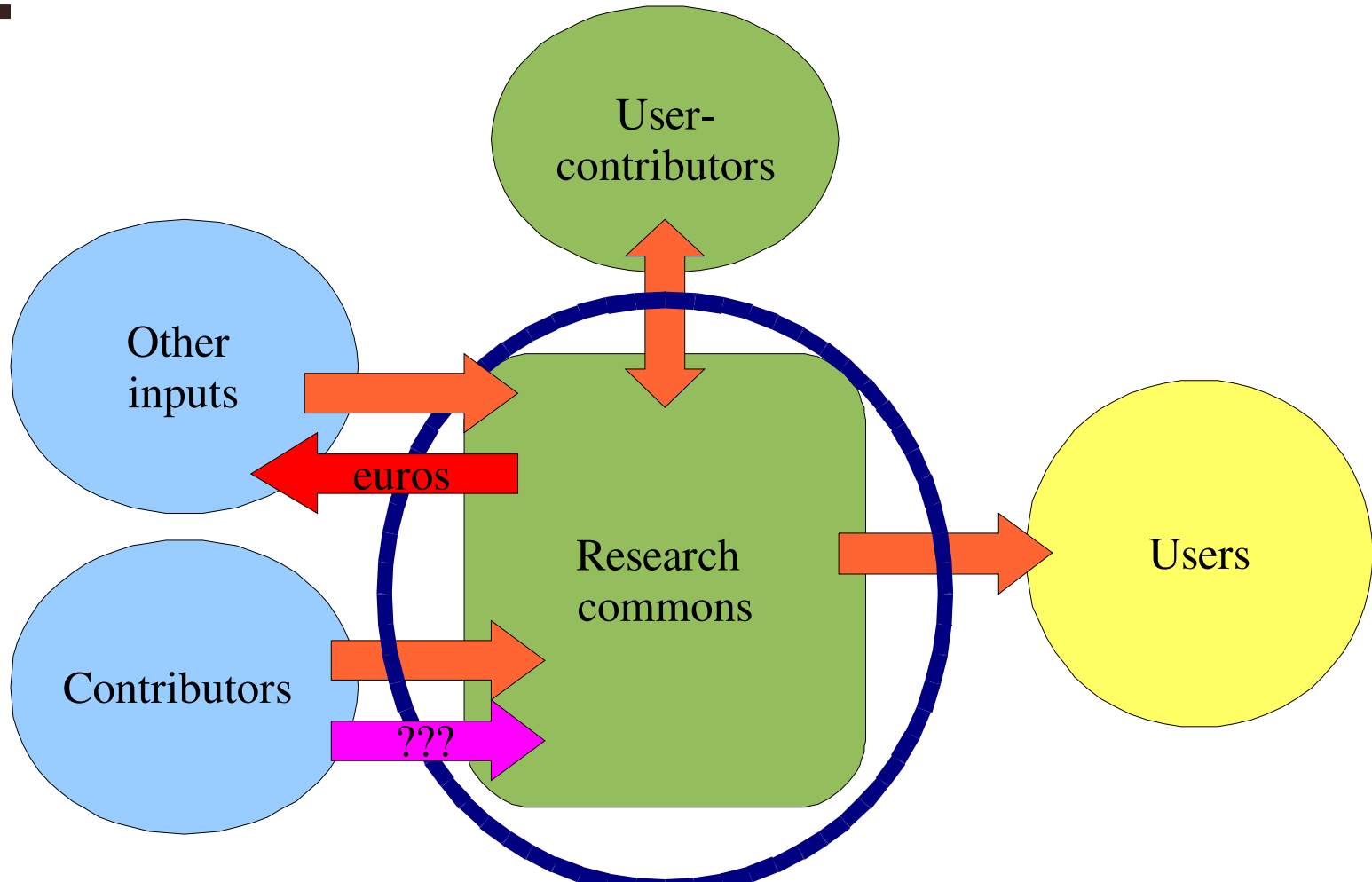
Conventional economic evaluation

- Productivity or profitability of a “closed” system (firm, line of business, etc.)
 - Measurement principle: output less input
- Aggregation over different types of outputs and inputs performed by measuring them in terms of monetary units
 - Feasible and appropriate in a market system because of the willingness-to-pay test

[For-profit firm]



[Open access commons]



Three differences

- Lack of market-mediated transactions that would provide appropriate aggregation.
 - Input suppliers may incur costs.
 - But not always (survey evidence)
 - Some participants are both customers and contributors.
- => Suggests that we need a different approach to measurement.

Inputs and outputs (1)

- Costing the inputs has two pieces:
 - The usual methodologies apply to inputs like computing power, website maintenance, telecommunications, etc.
 - More important – the willingness of users to contribute and the quality of their contributions
 - May be useful to subsume these into the output measures (assume that input cost is uniform across quality)

Inputs and outputs (2)

- Similar to the evaluation of basic research - a large share of benefits produced as “externalities” or “public goods”
 - see David, Mowery, Steinmueller 1992 on the supercollider
- Output has two parts:
 - enabling of future research or increasing its productivity
 - input to a variety of private profit-making activities
- Valuing the output - “willingness to use”

“Willingness to use”

- **Willingness to pay** assigns a value to a transaction by observing that at a certain price, the transaction takes place.
 - Allows comparison of “apples and oranges”
- **Willingness to use** assigns “value” from the observation that an individual finds the database or repository useful enough to access it.

Measures of willingness to use

- Website hits
 - Website downloads (“click-throughs”)
 - Citations to included papers, databases, etc.
 - Willingness to contribute
-

next few slides review these measures for
different types of commons

[Open source software]

- Tracking contributions:
 - Code is usually signed (but not always by employing firm)
 - Comes in units (lines) that are measurable
 - Information on re-use available
 - Quite a bit now known on this topic
- Tracking use and quality:
 - Require registration before download
 - Speed of bug correction
 - Growth of firms producing complementary outputs

Science and engineering preprint databases

- Measuring input
 - Number of contributions (relative to discipline)
 - Geographical spread?
 - Time lags?
- Measuring output
 - Downloads
 - Citations to papers in the database – changes in citation practice – see next slide

Does open access increase citation? – Yes

- Computer science conference articles openly accessible on the Web cited more often (150%). (Lawrence 2001)
- Citations rates of (EE, Math, Poli Sci, Philosophy) articles freely available on the web substantially higher than those that were not. (Antelman 2004)
- Open access oceanography articles downloaded more often (+180%). (Walker 2004).
- Open access articles from Proceedings of the National Academy of Science receive a twice as many citations controlling for funding source, field, author's record, etc. (Eysenbach 2006)

Does open access increase citation? - No

- Davis et al. 2008 – random experiment using journals of the American Physiological Society
 - no open access citation advantage after one year.
- Gaulé and Maystre 2008 - attempt to distinguish quality from diffusion:
 - PNAS experimented with authors' paying for open access - authors chose open access for higher quality articles and they are more highly cited.
 - But not after instrumenting by funding availability
 - However, standard errors are too large for a definitive test.

Databases and research tools

- Performance measures would be similar
- Costs of provision are higher
- Quality and accuracy more to the forefront
 - Track corrections?
 - User evaluations?
- License counts, citation counts
 - Citation more clearly measures knowledge flow
 - Need to encourage a citation norm
- More difficult: trace from research output to downstream output

Citations

- We know quite a bit about patent citations, somewhat less about paper or data citations.
Summary:
 - They are correlated with economic and spillover value and with the resources spent to obtain the invention, but still explain less than half of the variability
 - They are very skew (also in the case of paper citations) but so is invention value
 - Getting good measures requires waiting
 - Practices change over time and across disciplines
 - Most of this carries over to other types of citations

[Some suggestions]

- Effectiveness of a collection (contents plus distribution):
 - Web views or downloads (possibly adjusted for database or repository size)
- Value of a collection:
 - Citations per hit (adjusted for time period and type of collection)
 - Need a variety of statistics – mean is not enough, may also want median or top 10 per cent

[Caution]

- Comparing firm-level productivities is difficult due to heterogeneity of activities and intertemporal fluctuations
- The problem is even more difficult here due to the lack of a uniform measure (monetary value)
- As in the case of various research assessment exercises, it is essential to supplement numerical evaluation with qualitative evaluation.

[Conclusion]

- This paper only a conceptual start
- Leave the real work to others.....