New Data and Code for Customised Research Evaluation
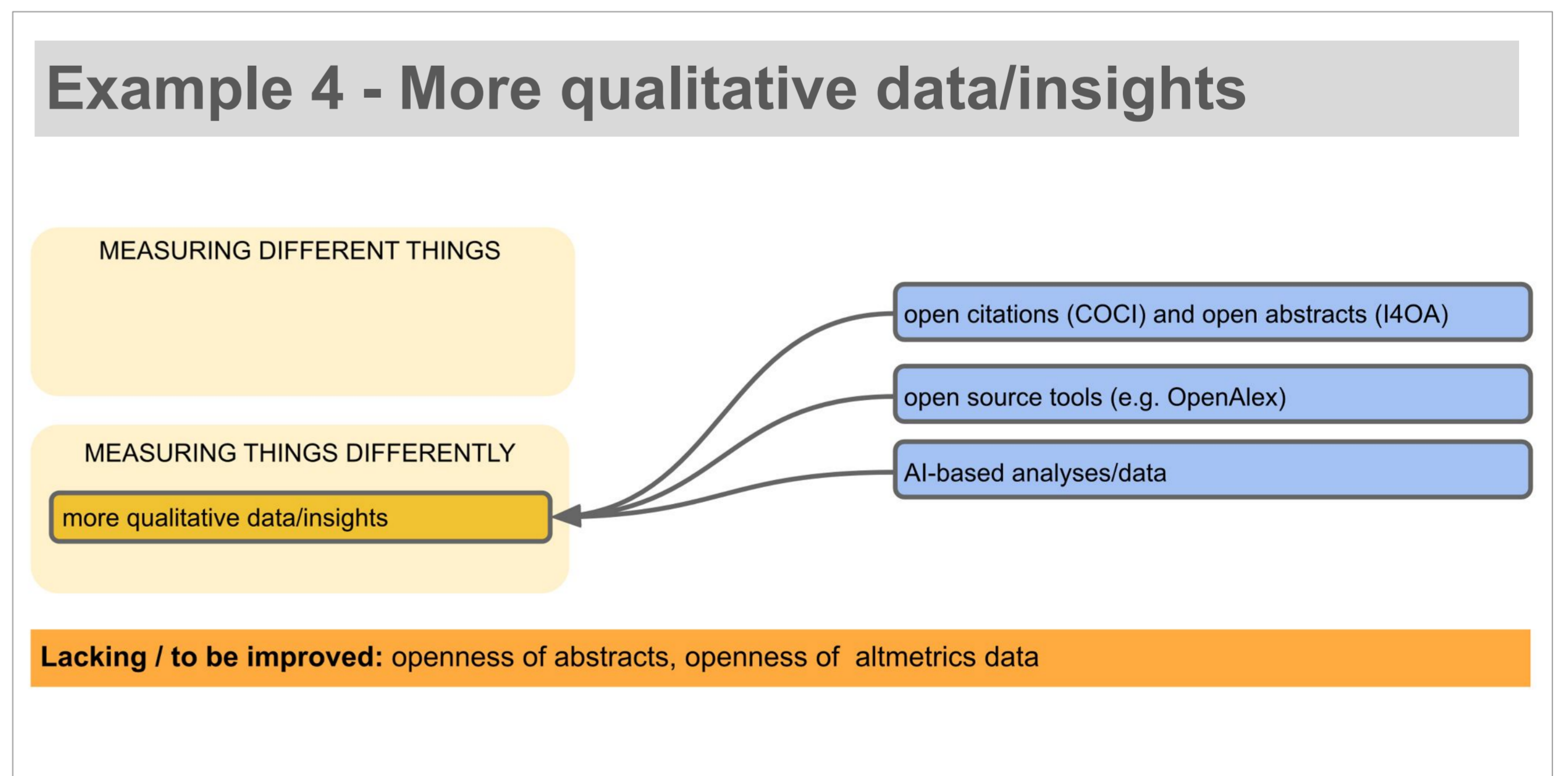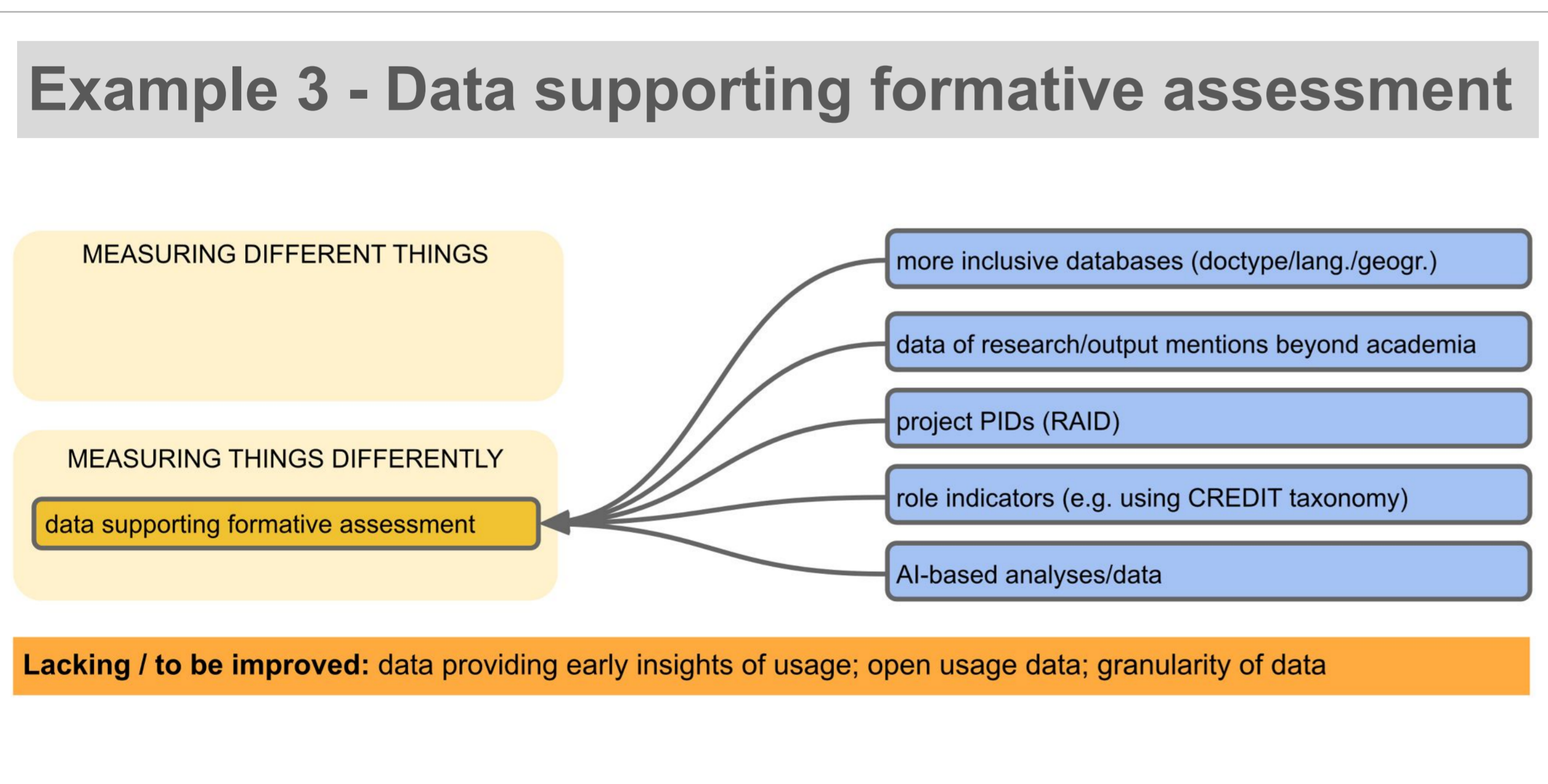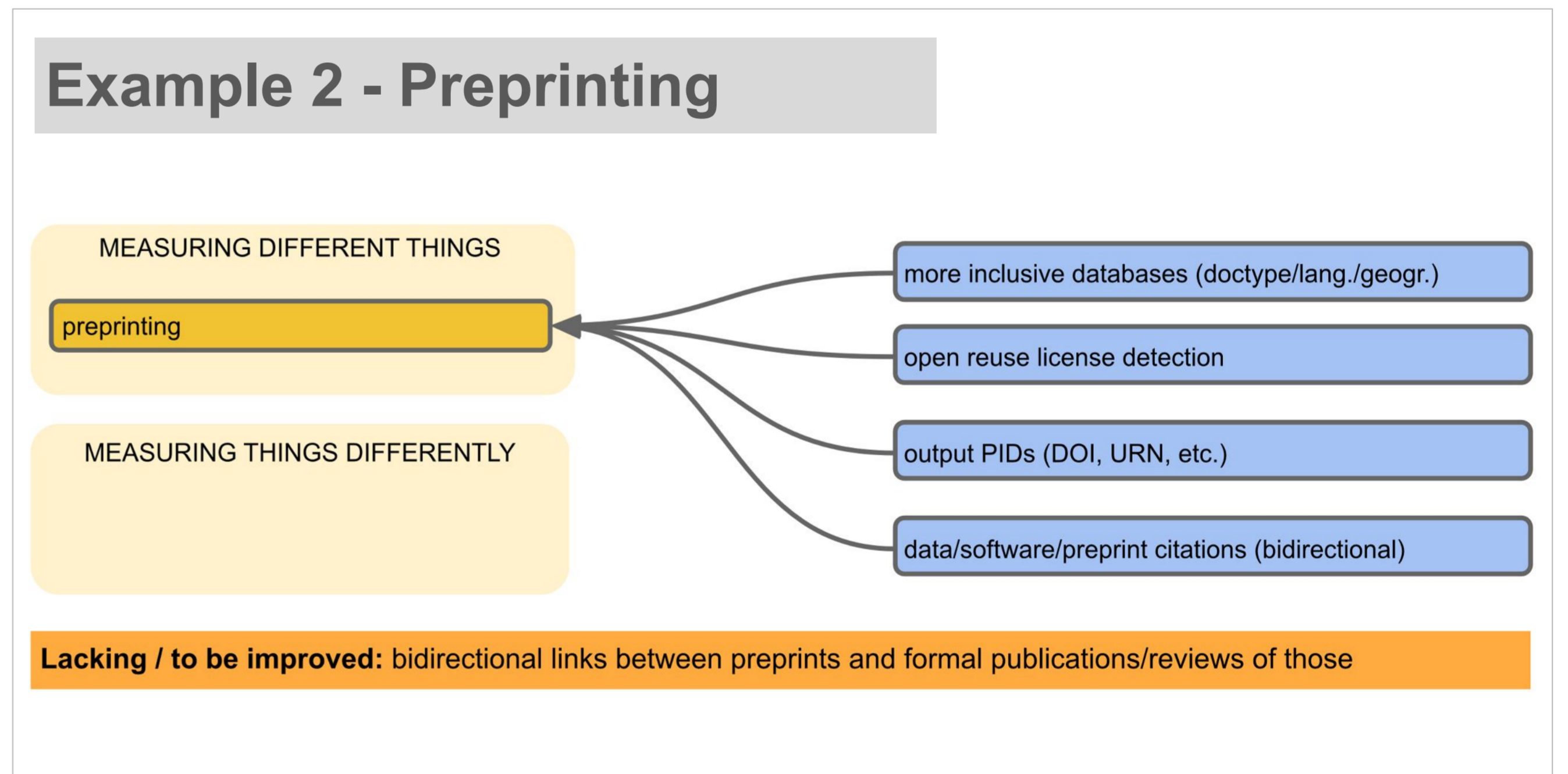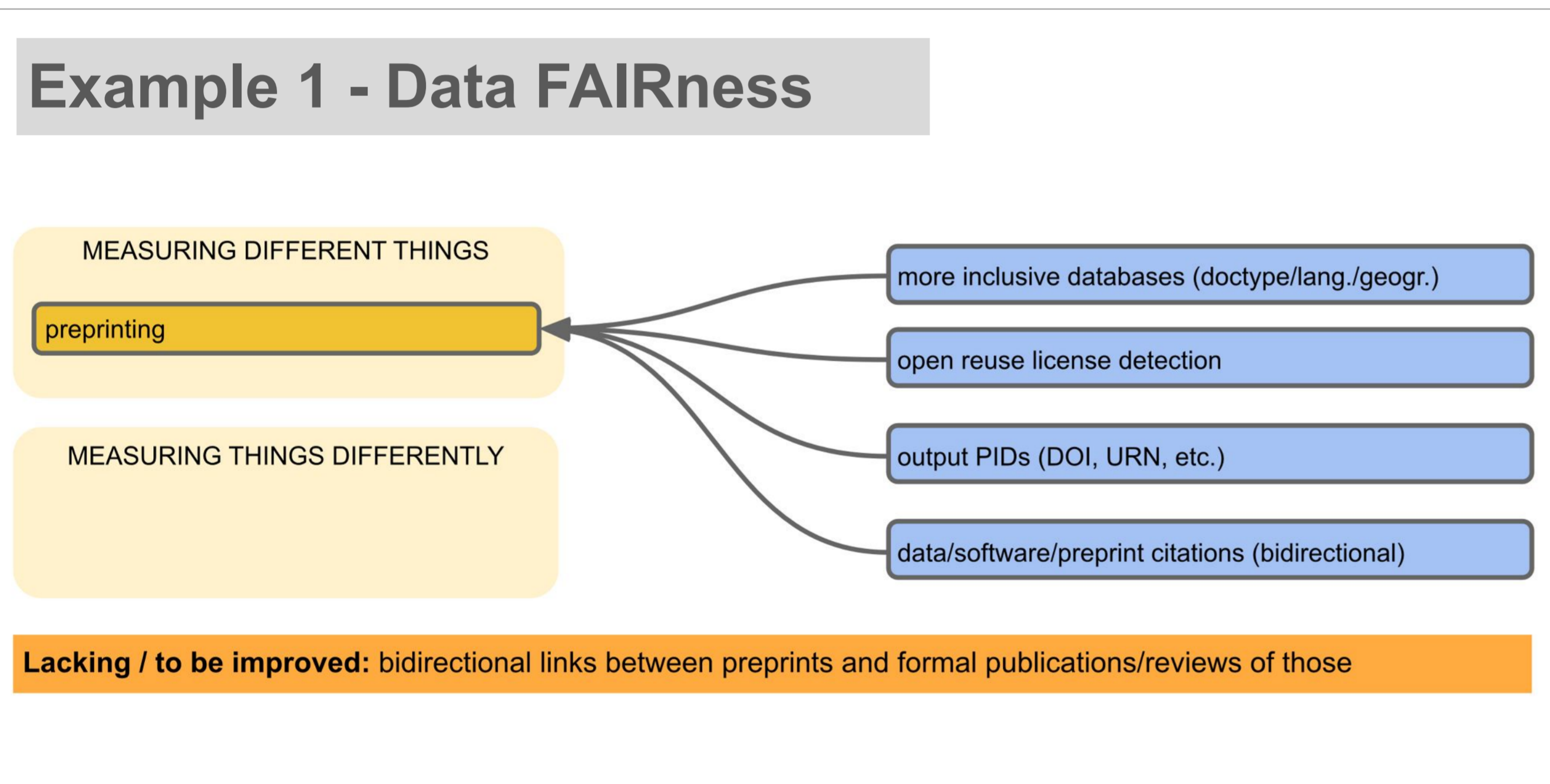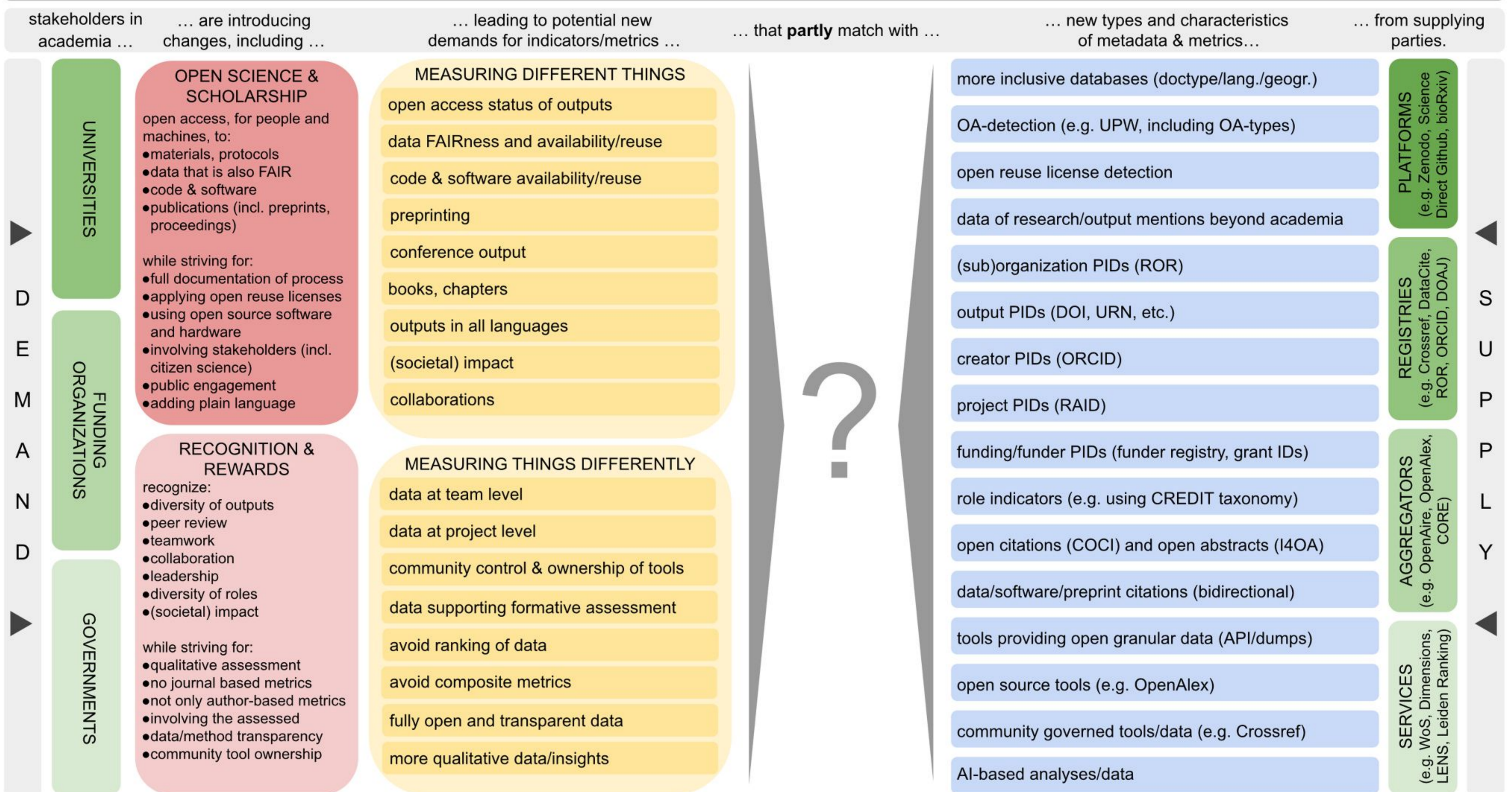
Jeroen Bosman - Utrecht University
iD 0000-0001-5796-2727

Bianca Kramer - Sesame Open Science
iD 0000-0002-5965-6560

Changes in demand and supply of metrics for research evaluation in the context of open science and new R&R

stakeholders in academia …	… are introducing changes, including …	… leading to potential new demands for indicators/metrics …	… that **partly** match with …	… new types and characteristics of metadata & metrics…	… from supplying parties.

DEMAND

UNIVERSITIES

FUNDING ORGANIZATIONS

GOVERNMENTS

OPEN SCIENCE & SCHOLARSHIP
open access, for people and machines, to:
- materials, protocols
- data that is also FAIR
- code & software
- publications (incl. preprints, proceedings)

while striving for:
- full documentation of process
- applying open reuse licenses
- using open source software and hardware
- involving stakeholders (incl. citizen science)
- public engagement
- adding plain language

RECOGNITION & REWARDS
recognize:
- diversity of outputs
- peer review
- teamwork
- collaboration
- leadership
- diversity of roles
- (societal) impact

while striving for:
- qualitative assessment
- no journal based metrics
- not only author-based metrics
- involving the assessed
- data/method transparency
- community tool ownership

MEASURING DIFFERENT THINGS
- open access status of outputs
- data FAIRness and availability/reuse
- code & software availability/reuse
- preprinting
- conference output
- books, chapters
- outputs in all languages
- (societal) impact
- collaborations

MEASURING THINGS DIFFERENTLY
- data at team level
- data at project level
- community control & ownership of tools
- data supporting formative assessment
- avoid ranking of data
- avoid composite metrics
- fully open and transparent data
- more qualitative data/insights

?

- more inclusive databases (doctype/lang./geogr.)
- OA-detection (e.g. UPW, including OA-types)
- open reuse license detection
- data of research/output mentions beyond academia
- (sub)organization PIDs (ROR)
- output PIDs (DOI, URN, etc.)
- creator PIDs (ORCID)
- project PIDs (RAID)
- funding/funder PIDs (funder registry, grant IDs)
- role indicators (e.g. using CREDIT taxonomy)
- open citations (COCI) and open abstracts (I4OA)
- data/software/preprint citations (bidirectional)
- tools providing open granular data (API/dumps)
- open source tools (e.g. OpenAlex)
- community governed tools/data (e.g. Crossref)
- AI-based analyses/data

PLATFORMS (e.g. Zenodo, Science Direct Github, bioRxiv)

REGISTRIES (e.g. Crossref, DataCite, ROR, ORCID, DOAJ)

AGGREGATORS (e.g. OpenAire, OpenAlex, CORE)

SERVICES (e.g. WoS, Dimensions, LENS, Leiden Ranking)

SUPPLY

Example 1 - Data FAIRness

MEASURING DIFFERENT THINGS
- preprinting

MEASURING THINGS DIFFERENTLY

- more inclusive databases (doctype/lang./geogr.)
- open reuse license detection
- output PIDs (DOI, URN, etc.)
- data/software/preprint citations (bidirectional)

Lacking / to be improved: bidirectional links between preprints and formal publications/reviews of those

---

Example 2 - Preprinting

MEASURING DIFFERENT THINGS
- preprinting

MEASURING THINGS DIFFERENTLY

- more inclusive databases (doctype/lang./geogr.)
- open reuse license detection
- output PIDs (DOI, URN, etc.)
- data/software/preprint citations (bidirectional)

Lacking / to be improved: bidirectional links between preprints and formal publications/reviews of those

---

Example 3 - Data supporting formative assessment

MEASURING DIFFERENT THINGS

MEASURING THINGS DIFFERENTLY
- data supporting formative assessment

- more inclusive databases (doctype/lang./geogr.)
- data of research/output mentions beyond academia
- project PIDs (RAID)
- role indicators (e.g. using CREDIT taxonomy)
- AI-based analyses/data

Lacking / to be improved: data providing early insights of usage; open usage data; granularity of data

---

Example 4 - More qualitative data/insights

MEASURING DIFFERENT THINGS

MEASURING THINGS DIFFERENTLY
- more qualitative data/insights

- open citations (COCI) and open abstracts (I4OA)
- open source tools (e.g. OpenAlex)
- AI-based analyses/data

Lacking / to be improved: openness of abstracts, openness of altmetrics data

---

Conclusion

While there are many promising developments in (meta)data provision that facilitate more diverse, fair and relevant research evaluations, there are also qualities with lacking or limited availability. The need for new code that can match (meta)data supply with indicator demands to support customized evaluations is presented as both a challenge and opportunity for the scientometric and research intelligence communities. The example cases illustrate this.