

My repository is being aggregated: a blessing or a curse?

Petr Knoth
CORE (Connecting REpositories)
Knowledge Media institute
The Open University

@petrknoth

Open Repositories 2014
Helsinki, Finland

Some interesting quotes about aggregations

"Aggregators and Google News are, to us, the worst offenders. They make money by living off the sweat of our brow."

<https://www.techdirt.com/articles/20091014/1831246537.shtml>

It seems as though when we like it we call it "curation," and when we don't we call it "aggregation." <https://gigaom.com/2011/07/13/like-it-or-not-aggregation-is-part-of-the-future-of-media/>

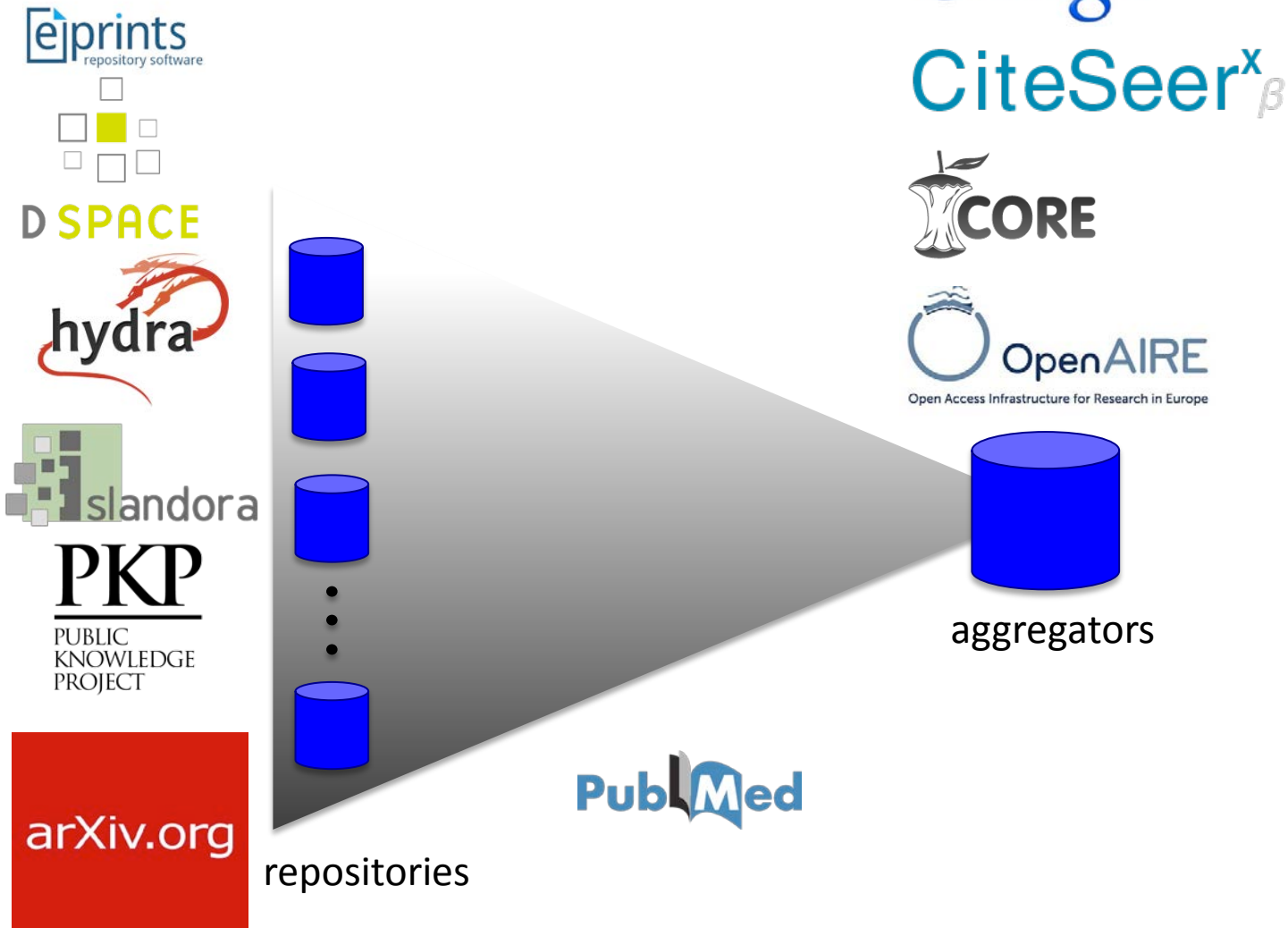


OR



?

The ecosystem



The use cases

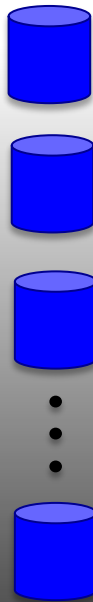
Mutually beneficial ecosystem!



Data input

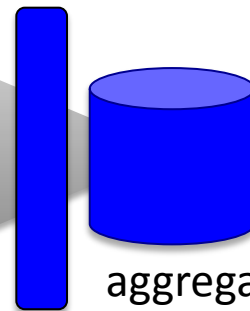


Data management



repositories

Enrichment & harmonisation



aggregators



Programmable
(machine-to-machine)
access

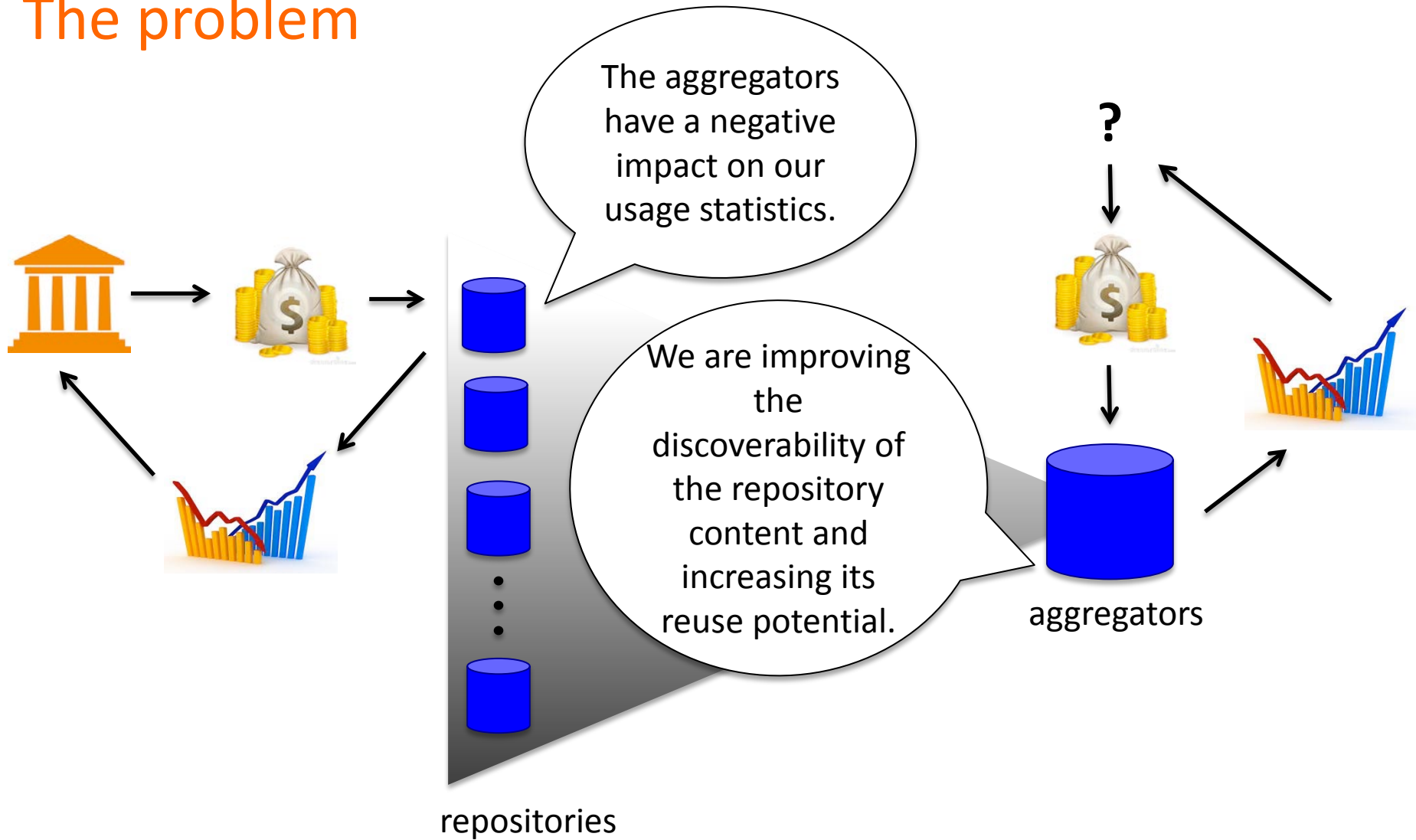


Search & discovery

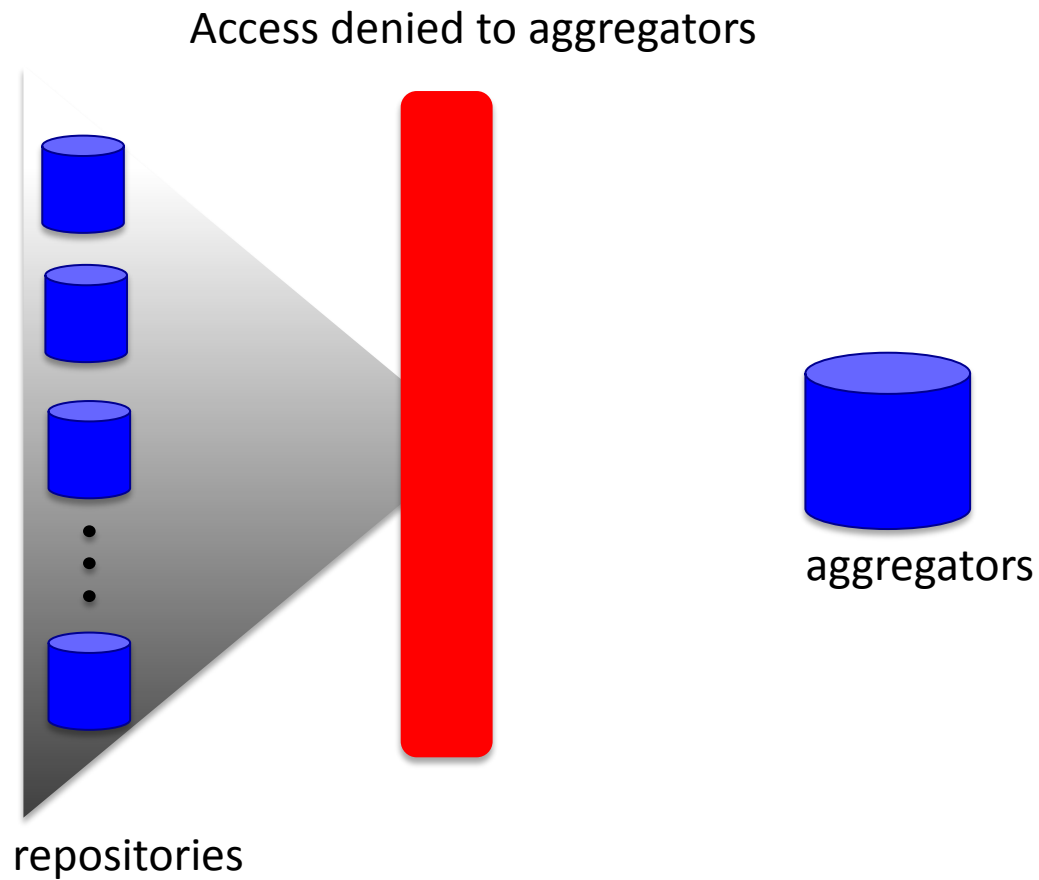


Analytics

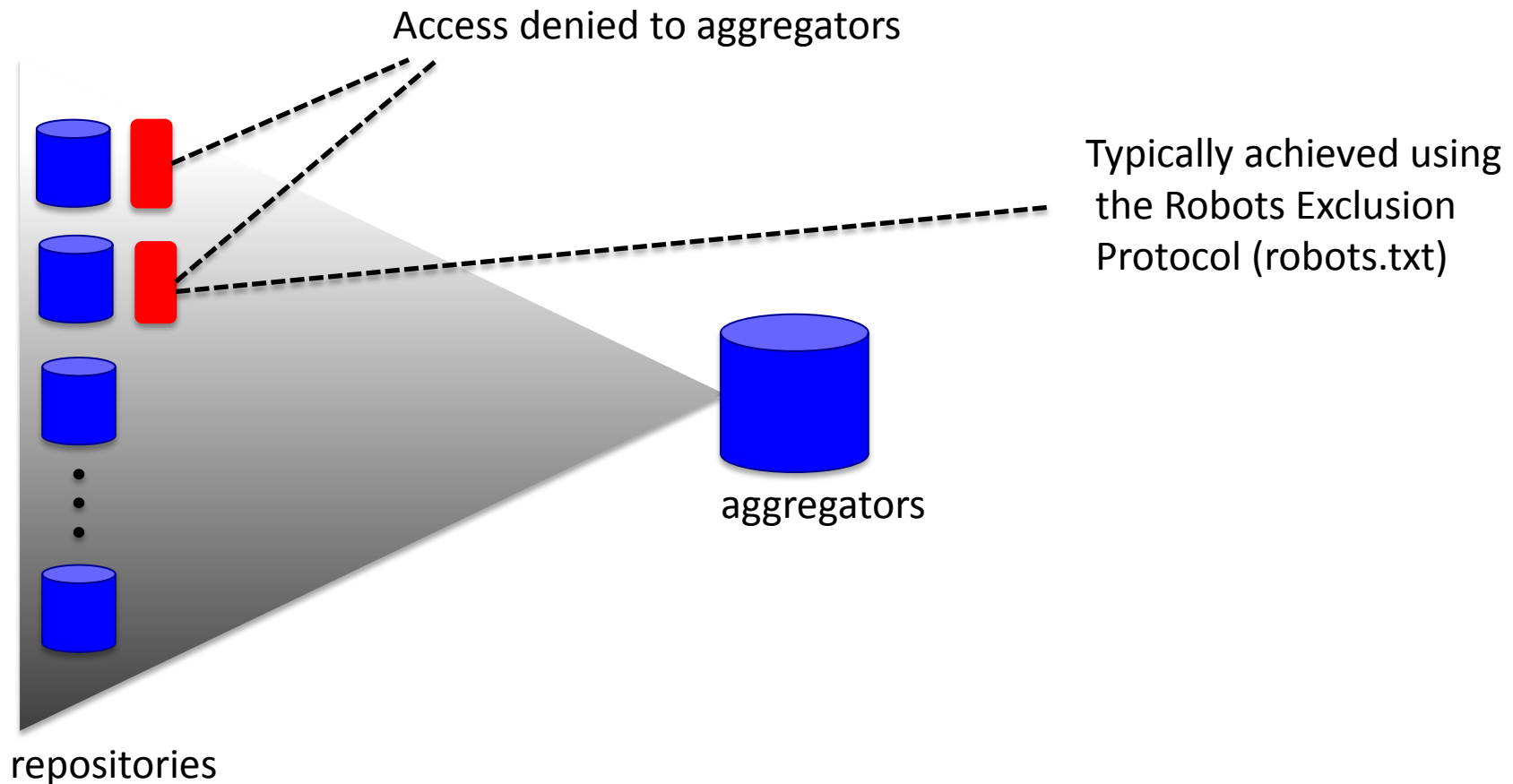
The problem



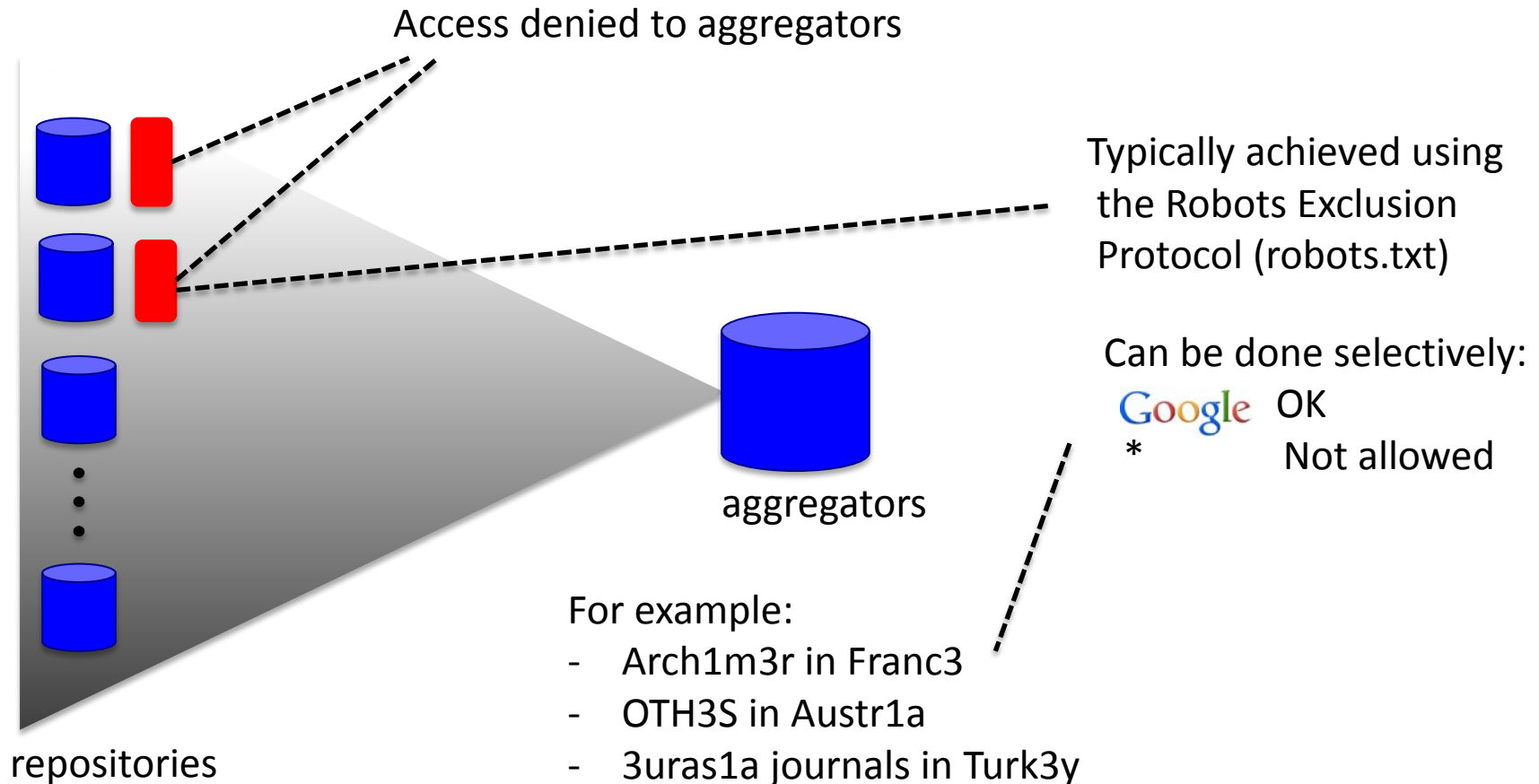
A shortsighted solution to the problem



A shortsighted solution to the problem



A shortsighted solution to the problem



The open access paradox

“Open access content is more open for exploitation by commercial services than by not for profit public services.”

Is protectionism legal?

Groom (2004) suggests it might be illegal as it, among other things, triggers concerns of unfair competition.

The mission of repositories according to SPARC (Crow, 2002)

“... the primary goal of repositories is to open and disseminate research outputs to a worldwide audience ...”

SPARC's position paper on IRs

*“For the repository to provide access to the broader research community, **users outside the university must be able to find and retrieve information from the repository.** Therefore, institutional repository systems must be able to support interoperability in order to provide access via multiple search engines and other discovery tools. **An institution** does not necessarily need to implement searching and indexing functionality to satisfy this demand: it **could simply maintain and expose metadata, allowing other services to harvest and search the content.** This simplicity lowers the barrier to repository operation for many institutions, as it only requires a file system to hold the content and the ability to create and share metadata with external systems.”*

COAR: About harvesting and aggregations ...

“Each individual repository is of limited value for research: the real power of Open Access lies in the possibility of connecting and tying together repositories, which is why we need interoperability. In order to create a seamless layer of content through connected repositories from around the world, Open Access relies on interoperability, the ability for systems to communicate with each other and pass information back and forth in a usable format. Interoperability allows us to exploit today's computational power so that we can aggregate, data mine, create new tools and services, and generate new knowledge from repository content.”

[COAR manifesto]

What is Open Access exactly?

By “**open access**” to [peer-reviewed research literature], we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, **crawl them for indexing, pass them as data to software**, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.

[BOAI, 2002]

Open Access = Access + **Reuse**

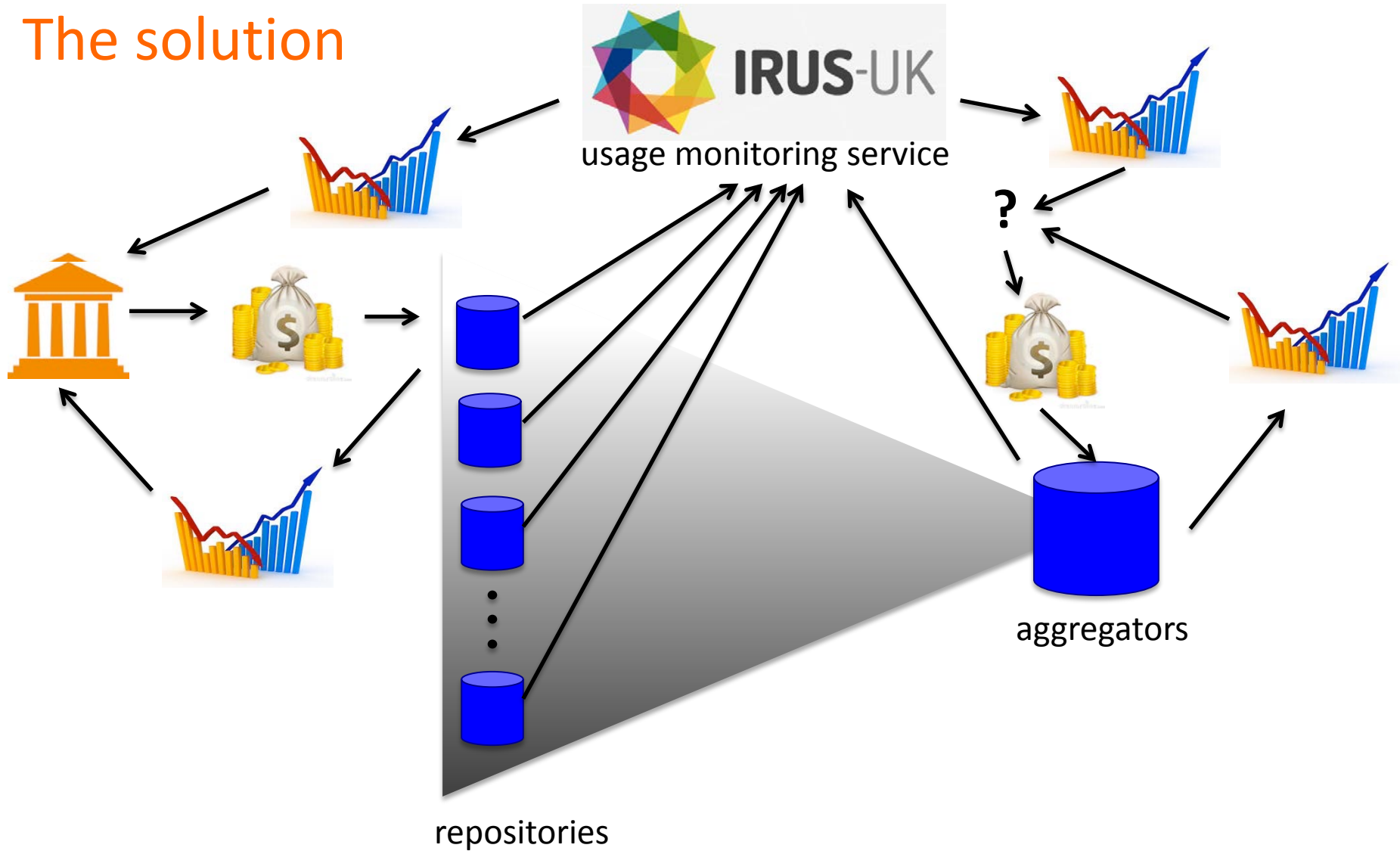
Multiple copies of content

- It would not be right to stop copying of content, as multiple copies mean:
 - Better preservation
 - Higher availability
 - Lower network latency
 - Increased visibility
 - Higher re-use opportunities
 - Keeping the market free from monopoly
 - Researchers like copying of content

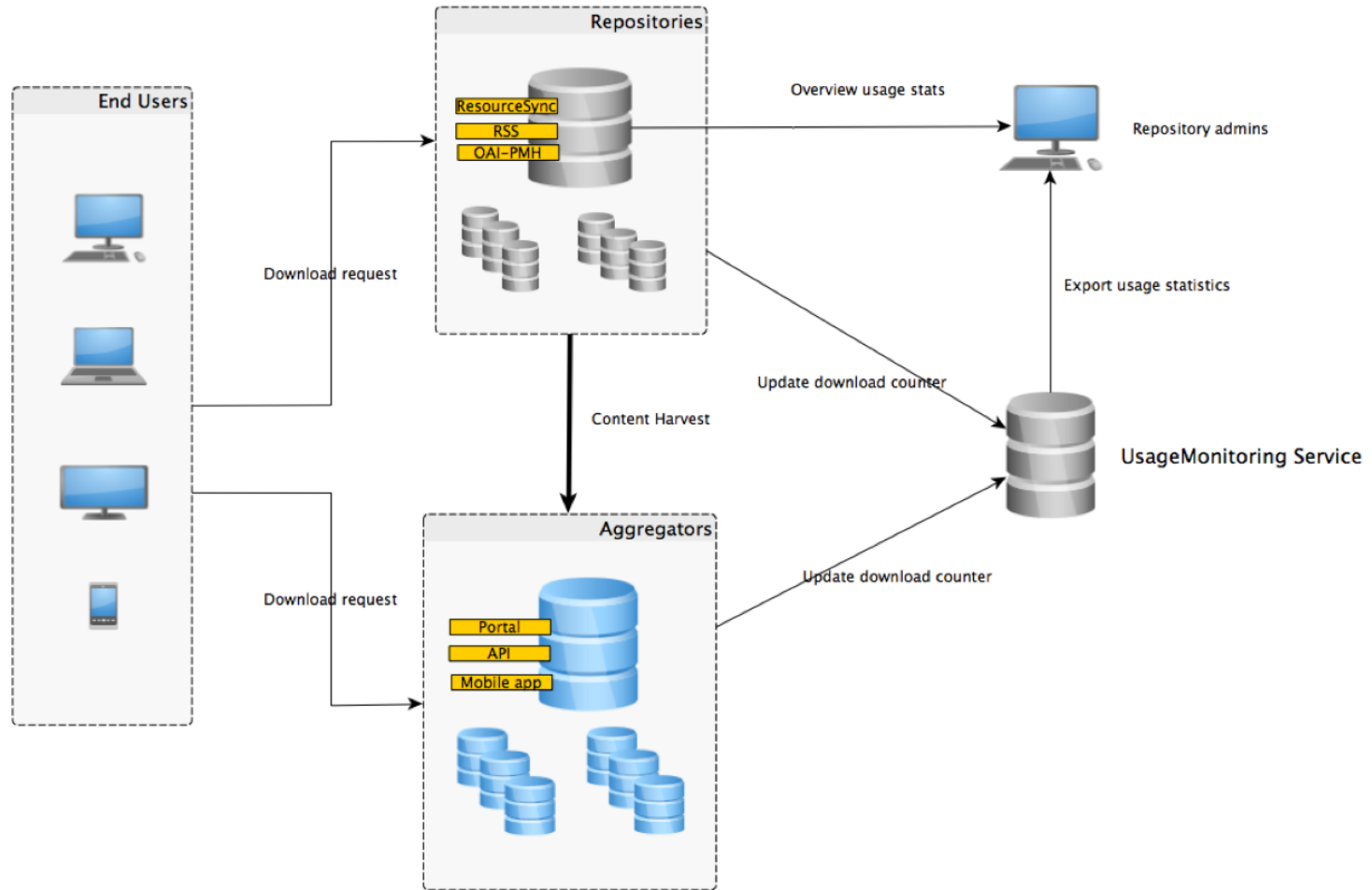
Solution

- Aggregators must support repositories and help them to fulfill their mission
- Repositories must stop believing they are the only access point for open access content (this includes both gold and green OA)
- Aggregators must implement reasonable measures to help repositories get accurate benchmarks.

The solution



IRUS-CORE implementation



Conclusions

- It is possible to create a **mutually beneficial ecosystem** for both repositories and aggregators
- Open Access is not just about access, but also **reuse** - encouraging multiple copies of content.
- The primary role of repositories is to **disseminate** not to become a single access point
- Repositories and aggregators each serve largely a **different audience**.
- Aggregators should implement mechanisms to **give credit** to repositories.

Thank you!

