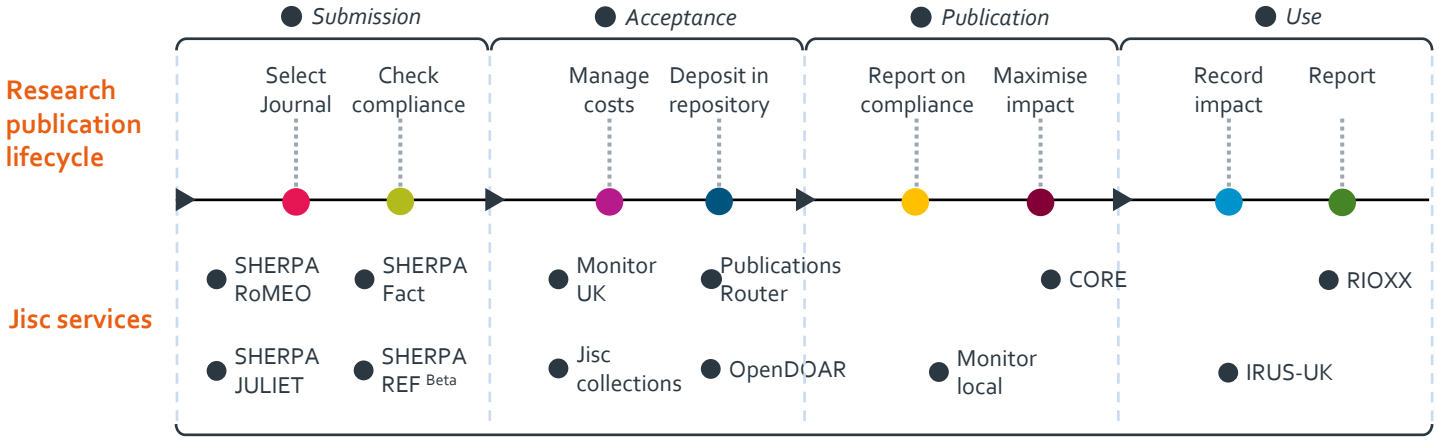


4/12/2018

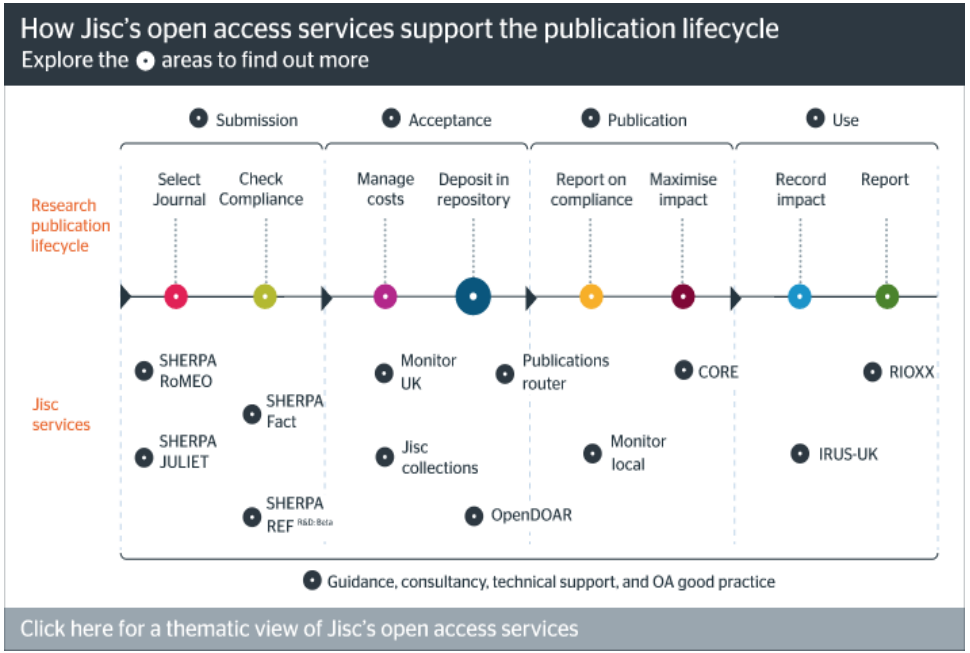
Jisc OA Dashboard

Monitoring OA / Enhancing Discoverability

Sarah Fahmy, Scholarly Communications Services Manager

Jisc OA Dashboard project – Background

- Institutions manage a wealth of data about their OA outputs at each stage of publication.
- **However**, information is stored across disconnected systems, each with different baselines and scope.
- An OA Dashboard ***could*** summarise information from Jisc services (and other open sources) to better inform institutions, funders and Jisc itself on the current status of OA development.



The project at a glance

Phase 1: Defining the use cases

•Dashboard options identified and prioritised

Undertaken by Jisc in partnership with Research Consulting, Pleiade Management and Digerati

Phase 2: Prototype Development

• Development of technical prototypes and feasibility study

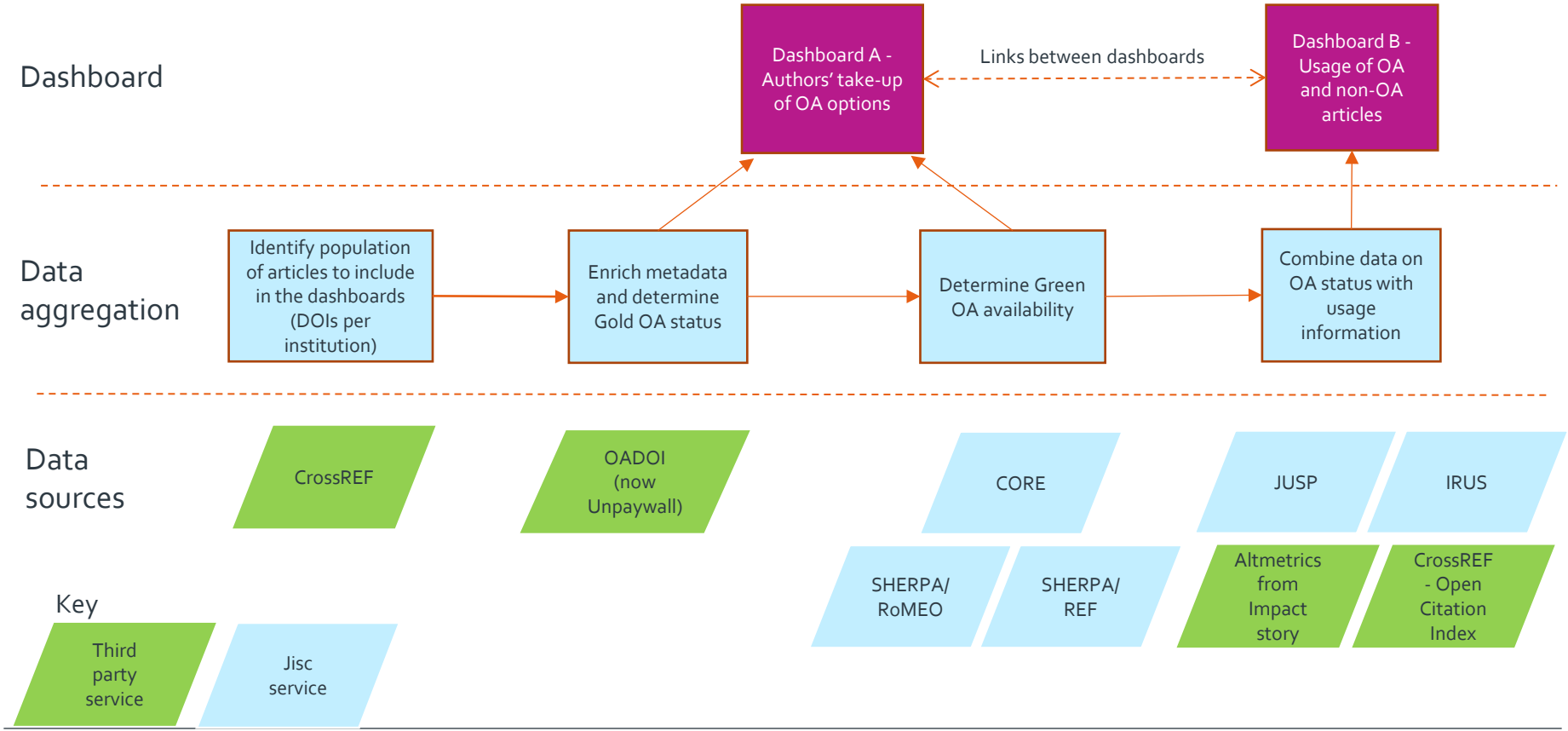
Phase 3: Business case and reporting

• Feedback on prototype and insights on a possible business case.

The selected dashboards

Possible Dashboard	Use case	Sample indicators
A. Informing OA policy effects	Authors' take-up of OA options	% Green OA articles % Pure Gold OA articles % Hybrid Gold OA articles
B. Informing Green and Gold OA policy effectiveness	Usage of OA and non-OA articles Citations and Altmetrics of OA and non-OA articles	# downloads articles in repositories in comparison with publishers' platforms Average citation rate of APC OA Gold articles

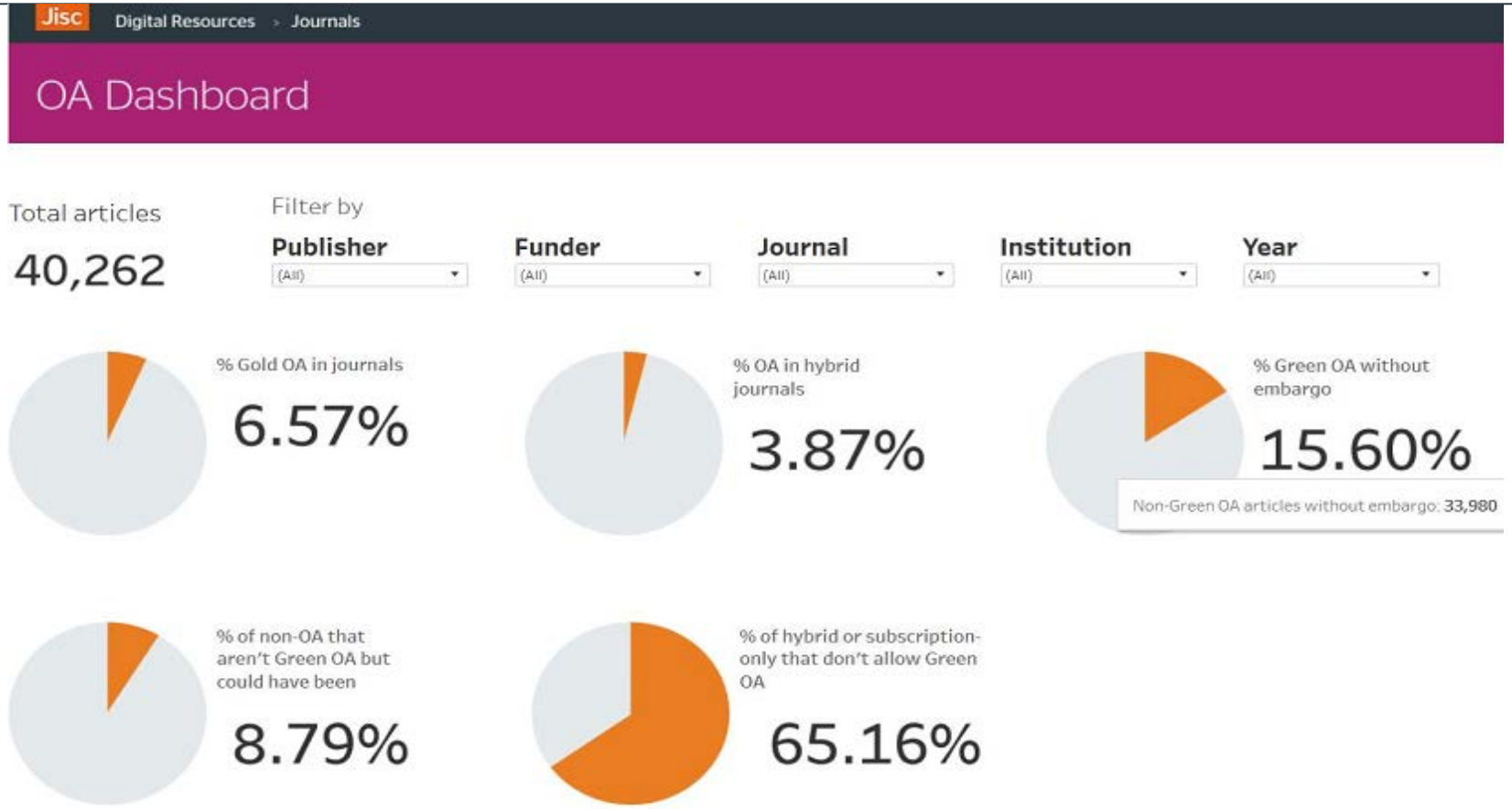
Data sources for Dashboards A and B



Data source assessment

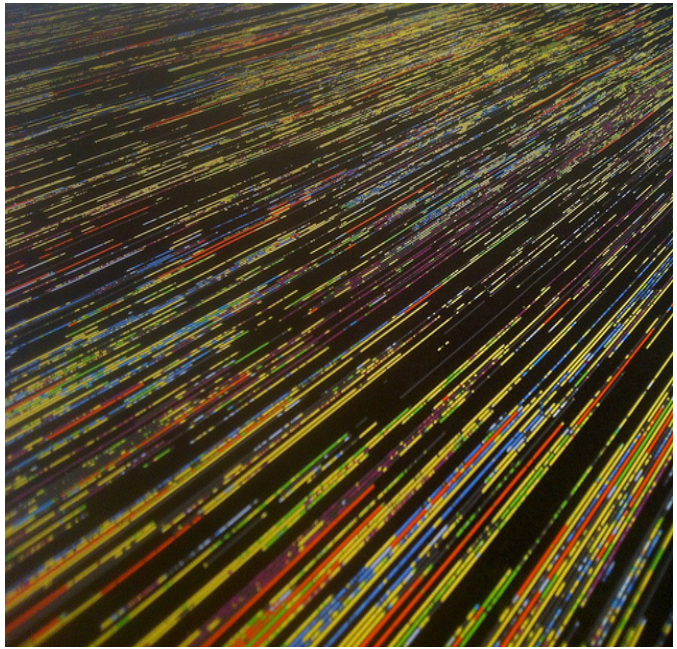
Data Source	Used to identify	Benefits	Reasons for selection	Challenges
Crossref	Universe of articles	• Robust data on publisher/ journal • Comprehensive record of new articles	• Disciplinary coverage	• Incomplete data for author/ institutional affiliation
oaDOI (Unpaywall)	OA Status	• Gathers data from range of aggregators • Handles 1 million call per minute • Relatively high level of accuracy	Inclusion of licence information plus data from other sources to determine OA Status	• Does not retrieve Green OA articles under embargo • Not all OA articles identified (estimated 77%)
Sherpa Romeo	Green OA availability	• Extensive coverage of publisher policies	Provision of good (not perfect!) coverage	• Data quality- some journals classified incorrectly • Difficult to determine licensing information • Scalability of API

OA Dashboard prototype



Data source issues – Summary

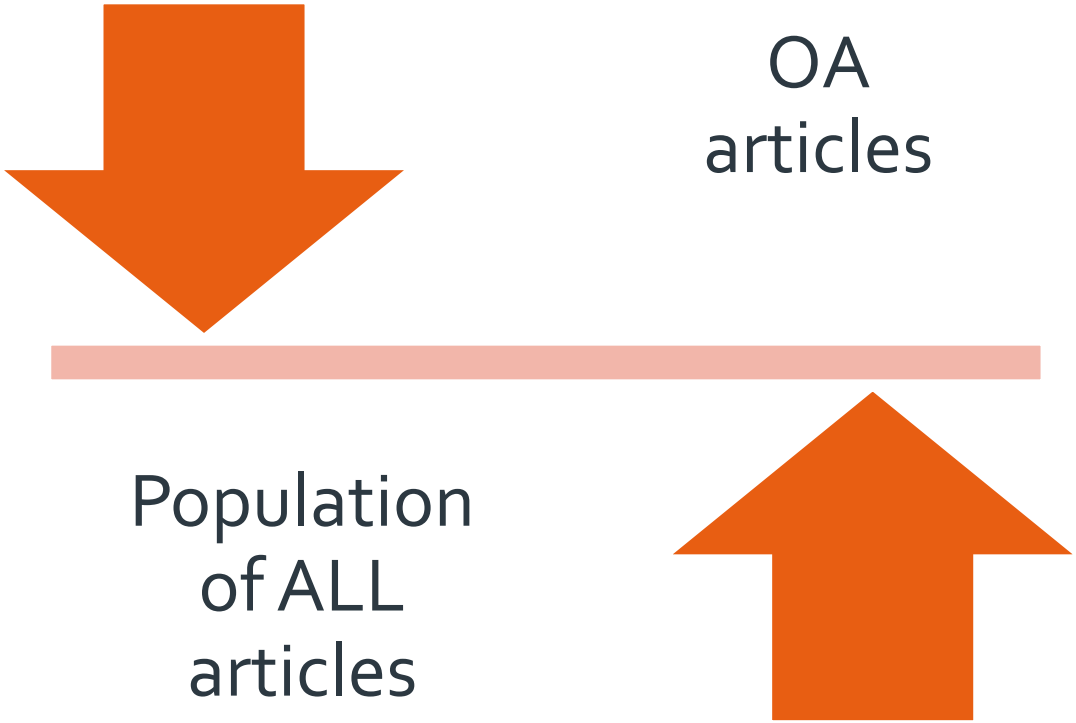
- Intrinsic limitations of the data sources:
  - Incompleteness of data
  - Issues re: APIs
  - Data quality
- No data model to suit requirements



Dashboard A (determining OA status) would require significant effort to obtain and normalise data

The wider question

Think of it as a fraction…



Denominator – 'Universe of articles'

The entire body of research literature is currently estimated at over 100 million publications (Khabsa & Giles, 2014)

Annual increase of around 2.5 million published each year (Jinha, 2010) and an estimated 10% year on year increase in the annual number of these outputs (Bornmann & Mutz, 2014).

On this basis, in 2018, could we put this figure between 110-140 million?

Numerator – Proportion of OA

- Unpaywall is an open online service that determines OA status

- **The State of OA** (Piwowar & Priem,2017) study looked across of 100,000 articles, to investigate OA in three populations:
  - All journal articles assigned a Crossref DOI
  - Recent journal articles indexed in Web of Science
  - Articles viewed by users of Unpaywall

Findings:

- Of 67mill articles Unpaywall searches across, 28% of the scholarly literature is OA (19M in total)
- The most recent year analysed (2015) also has the highest percentage of OA (45%).

The issue:

Licence information is NOT being applied consistently

OA Discoverability

- The numbers are estimates..
- At best, we can say that approx. 50% of all articles are OA
- But does this matter?
- Discoverability is arguably the key here…isn't it?

OA Discoverability

- This is a domain in which there is a great deal of very agile innovation (Unpaywall, OAButton, Kopernio)
- More established public infrastructure (OpenAIRE, CORE)
- Well-used library services (Primo, Summon, Ebsco, OCLC)
- Long-standing commercial services (Web of Science, Scopus, Google Scholar).

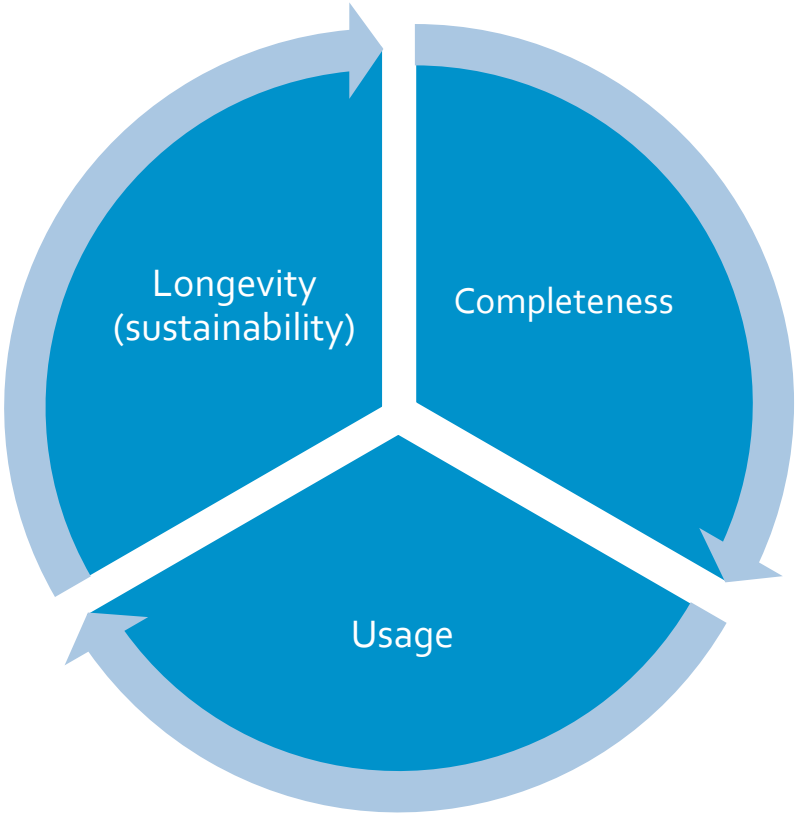
OA Discoverability

» Arguably, there is no one panacea due to issues of:

› **Completeness -** est. 140 million articles
  – BASE: 123 million metadata records
  – CORE: 126 million metadata records (10.5 million full text)

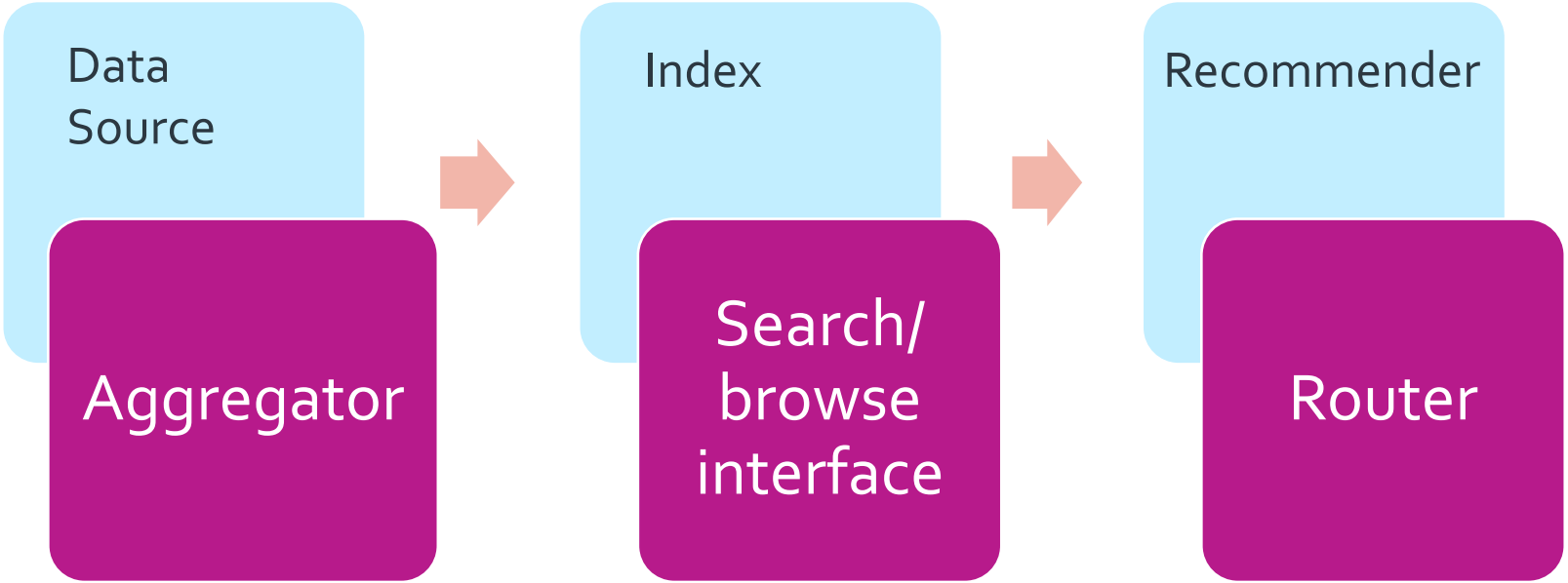
› **Longevity** (sustainability)
  – Usage stats
  – Business model

OA Discovery



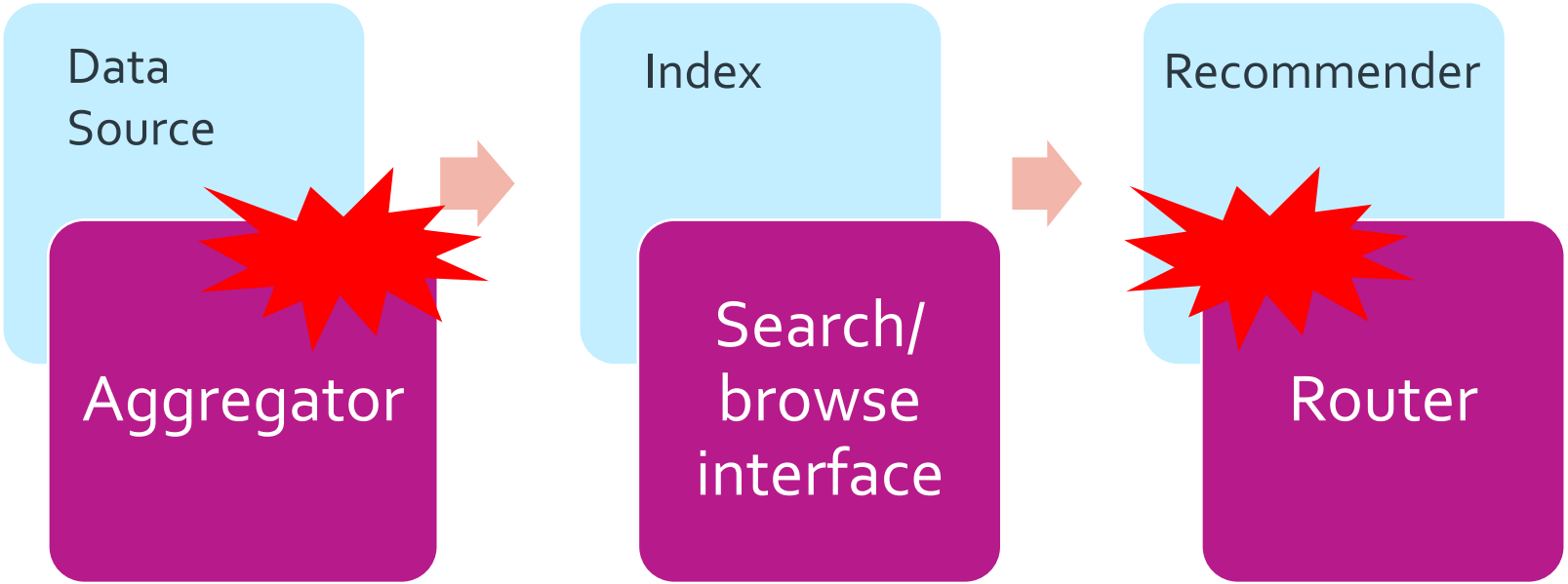
Understanding the OA Discovery chain

Identify relevant actors and how they interact
(for illustrative purposes only!)



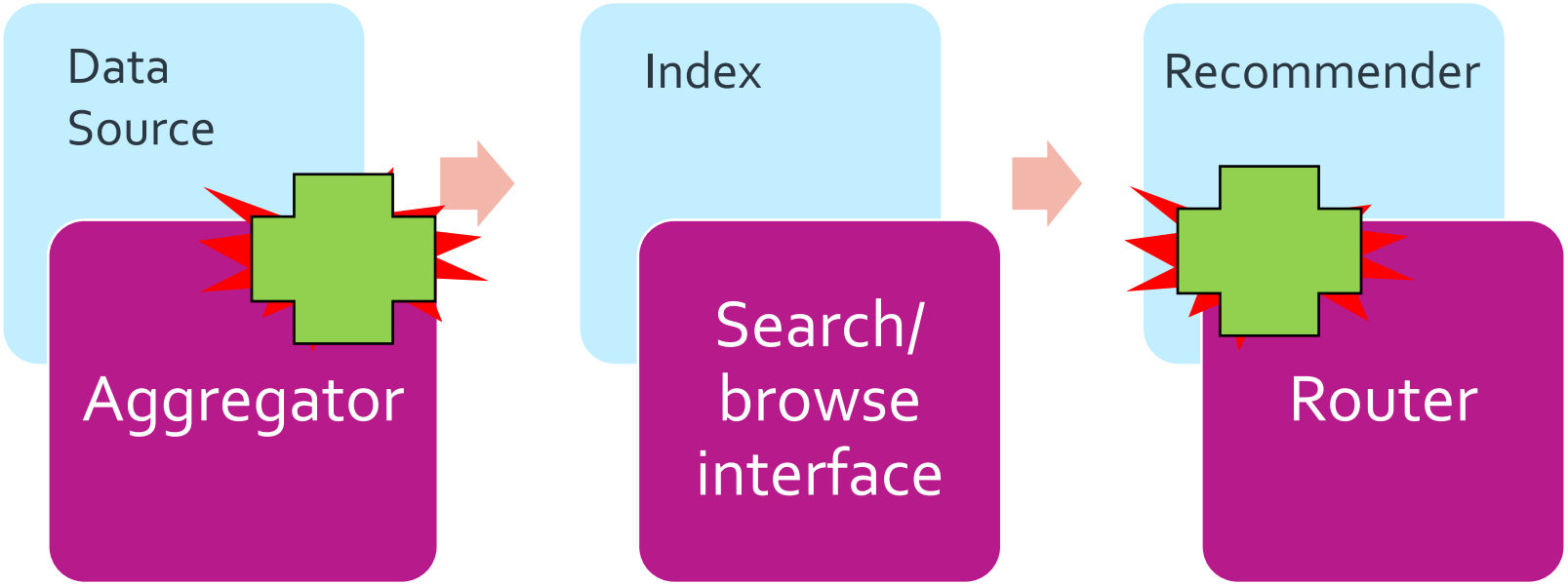
Understanding the OA Discovery chain

Identify the pain points…



Understanding the OA Discovery chain

Identify and develop solutions…



Thank you for listening
Any questions?


Sarah Fahmy
Scholarly Communications services manager
sarah.fahmy@jisc.ac.uk
jisc.ac.uk