Sustainability in the scientific information industry

Micha Moskovic (CERN), 2nd sustainable HEP workshop

Outline

1. Introduction
2. Insider point of view: INSPIRE
3. Outsider point of view: the academic publishing industry
4. Conclusion

1. Introduction

Disclaimer

- I am responsible for the INSPIRE HEP information platform
- I am not a publisher
- I am not a climatologist, sustainability expert, economist, sociologist, etc.
- My views on the academic publication industry are those of an interested outsider, they don’t represent publishers

Spoiler alert: I don’t have all the answers!

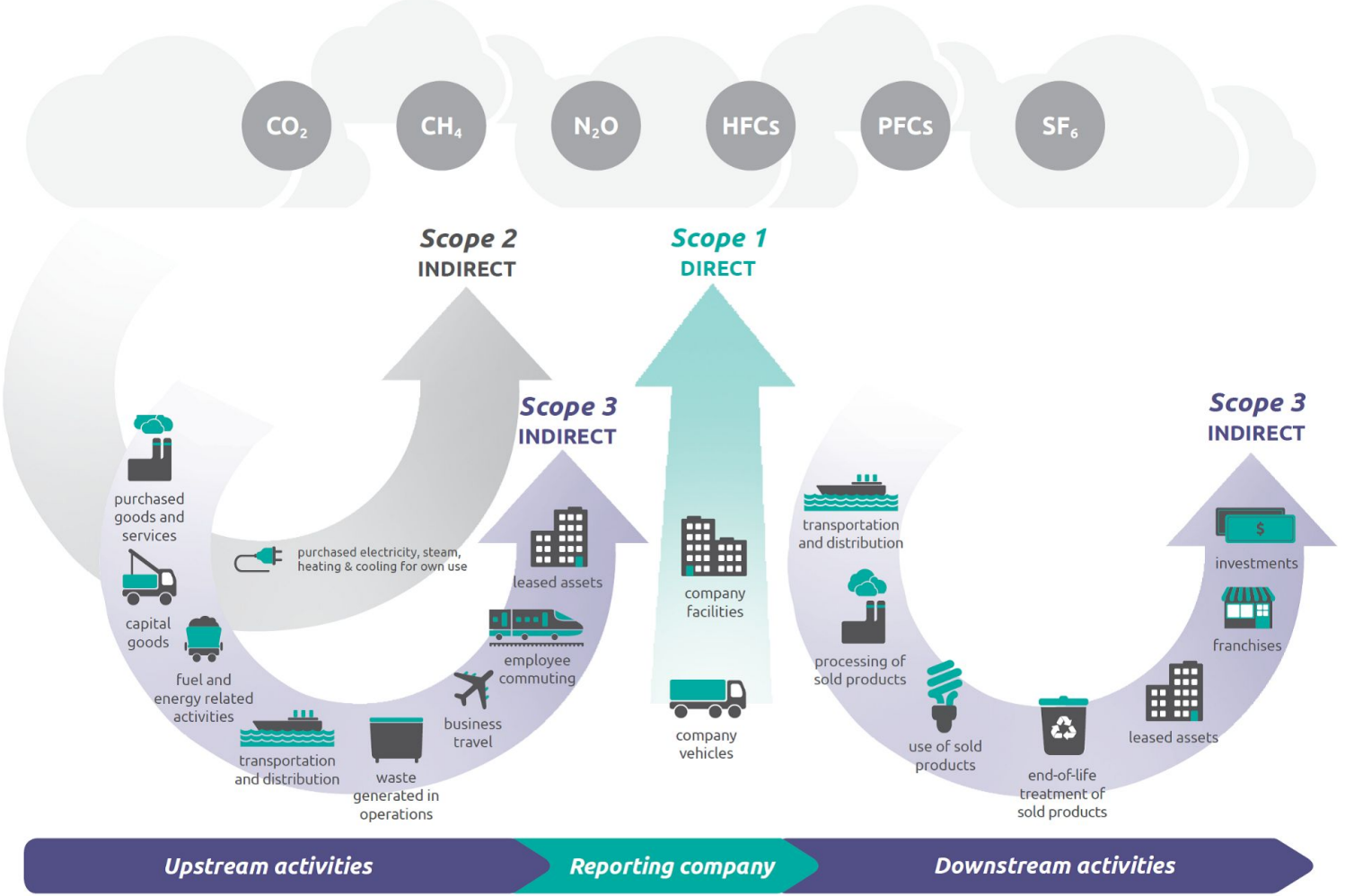
What is the scientific information industry?

All actors that deal with scientific information in a broad sense

- Academic publishers (Springer Nature, Elsevier, American Physical Society, etc.)
- Preprint repositories (arXiv, bioRxiv, etc.)
- Institutional repositories (CDS, OSTI.gov, etc.)
- Other online information platforms (Google Scholar, Web of Science, INSPIRE, ADS, ORCID, HEPData, Indico, etc.)

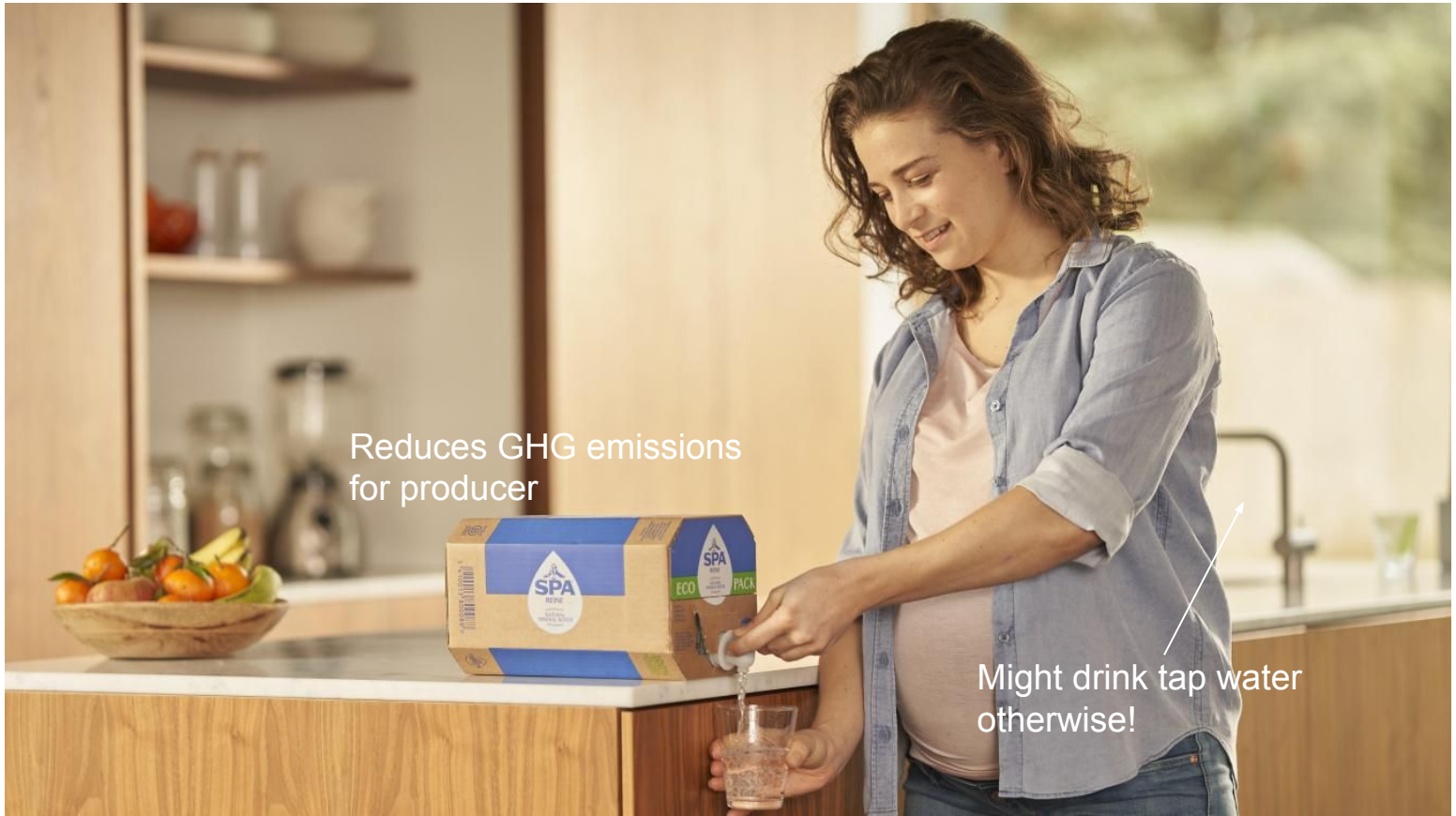
How to think about the climate impact?

- Standard: GHG protocol
  "Greenhouse Gas Protocol provides standards, guidance, tools and training for business and government to measure and manage climate-warming emissions."
- 3 scopes are defined
  - Scope 1: direct emissions from owned or controlled sources
    Example: natural gas heating for offices
  - Scope 2: indirect emissions from the generation of purchased energy
    Example: electricity for in-house computers
  - Scope 3: all other indirect emissions that occur in the value chain of the reporting company
    Example: electricity for cloud services



GHG Corporate Value Chain (Scope 3) Accounting and Reporting Standard

Problem: encourages first order thinking



Alternative approach: consider also induced emissions

When making some change that has a potential impact on emissions, consider

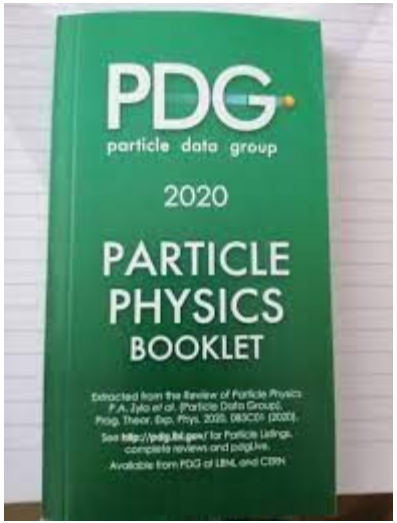
- First order effects: direct changes in emissions due to the change
- Second order effects: global changes in emissions in response to first order effects

Second order effects very hard to assess quantitatively, but can be large given importance of scientific information (“publish or perish”)!

A (hypothetical?) example: PDG booklet

If PDG stops printing of booklet

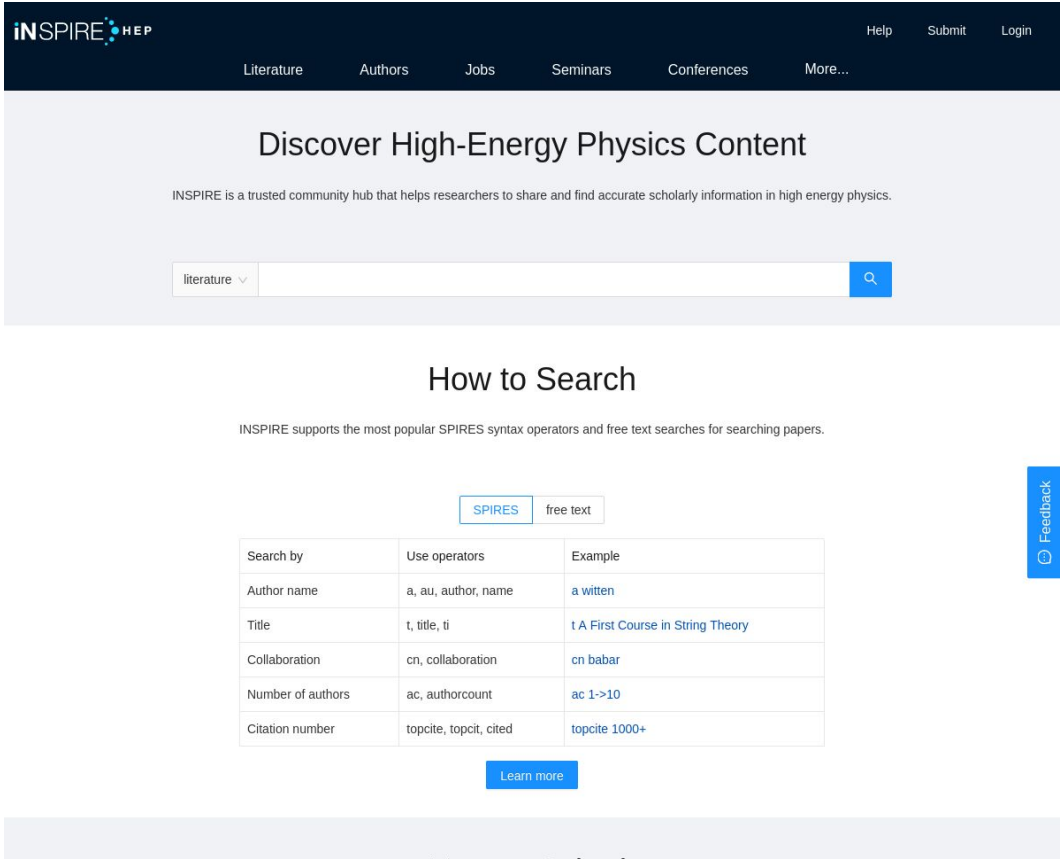
- First order effect: reduced scope 3 emissions (paper production, printing, binding, shipping, etc.)
- Second order effect: some people will print their own, increasing emissions

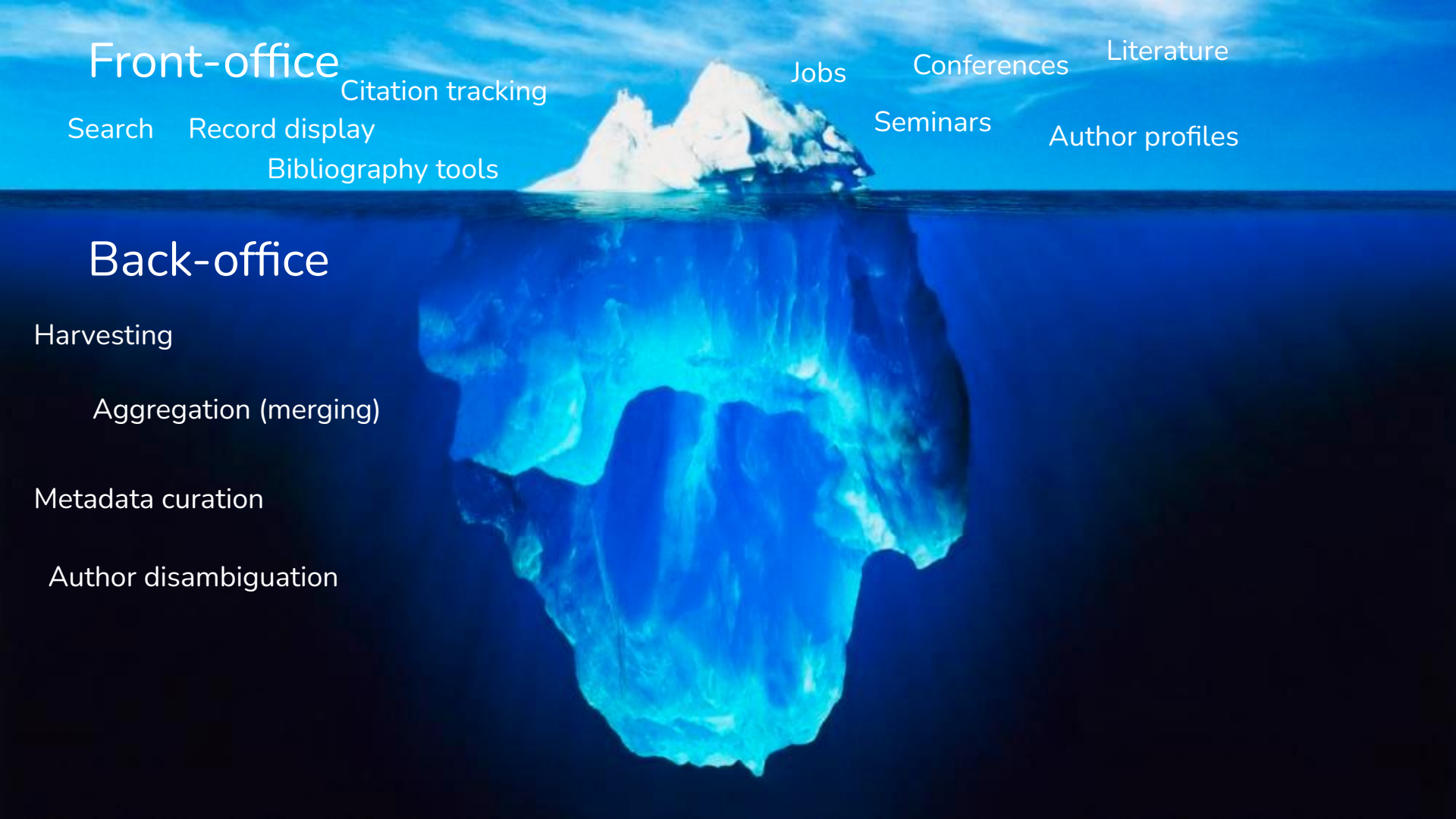


2. INSPIRE

What is INSPIRE https://inspirehep.net

- Main information platform for HEP
- Similar to ADS for astro
- Completely free to use
- Run by collaboration of HEP labs




Front-office
Citation tracking
Search
Record display
Bibliography tools
Jobs
Conferences
Literature
Seminars
Author profiles
Back-office
Harvesting
Aggregation (merging)
Metadata curation
Author disambiguation

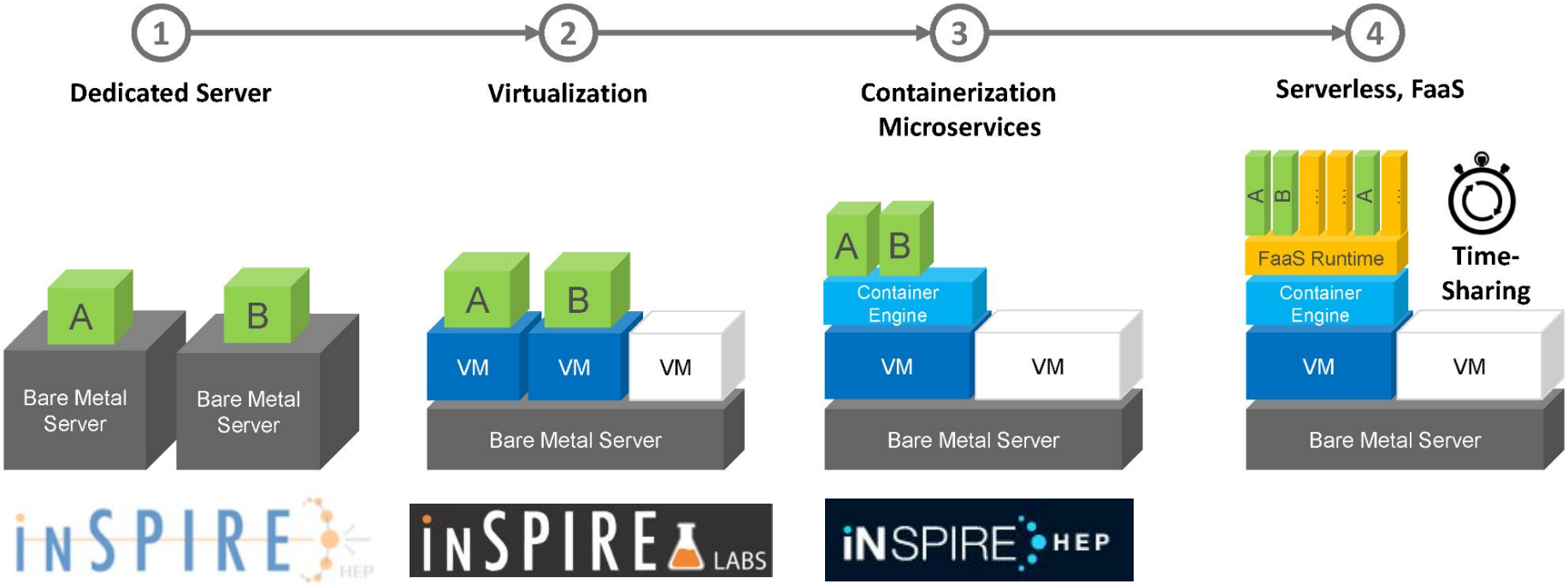
Behind INSPIRE: people

- Developers ~4 FTE
  - Technical operations
  - New features development
- Curators ~11 FTE
  - Content selection
  - Metadata improvements
  - Dealing with user requests

Emissions from INSPIRE human activity

- Hard to estimate quantitatively: geographic diversity and lack of available data
- Geographical specialization: high-bandwidth interactions in person
- Collaboration-level coordination: interactions almost exclusively on-line
- Travel-related emissions are low

Behind INSPIRE: machines



Adapted from [10.3390/app8081368](10.3390/app8081368)

Behind INSPIRE: machines

- The CERN Scientific Information Service (SIS) operates a kubernetes cluster to run containers for various services
  - INSPIRE
  - HEPData
  - CERN Analysis Preservation
  - …
- Almost all of INSPIRE components in cluster
- But search engine (ElasticSearch) operated by CERN IT

Estimating scope 2 emissions for computing

- SIS cluster hosted in CERN data center
- No granular electricity usage available, need to estimate from total
- CERN main data center consumed 37 GWh in 2021 (CERN News)
- Currently has 451k cores in total (CERN IT monitoring)
- SIS cluster has 980 cores = 0.2% of total
- ⇒ Estimated consumption: 80 MWh
- Average emissions for electricity in 2021: 22g CO_2eq/kWh (EDF)
- INSPIRE yearly emissions: 1.8t CO_2eq/year
- Same as single London-New York roundtrip economy flight! (myclimate)

What does this mean for INSPIRE?

- Important to minimize emissions
- INSPIRE scope 2 computing emissions are relatively small
- More promising to focus on second order effects
- Improve tools so the HEP community can reduce its emissions

How we develop new features

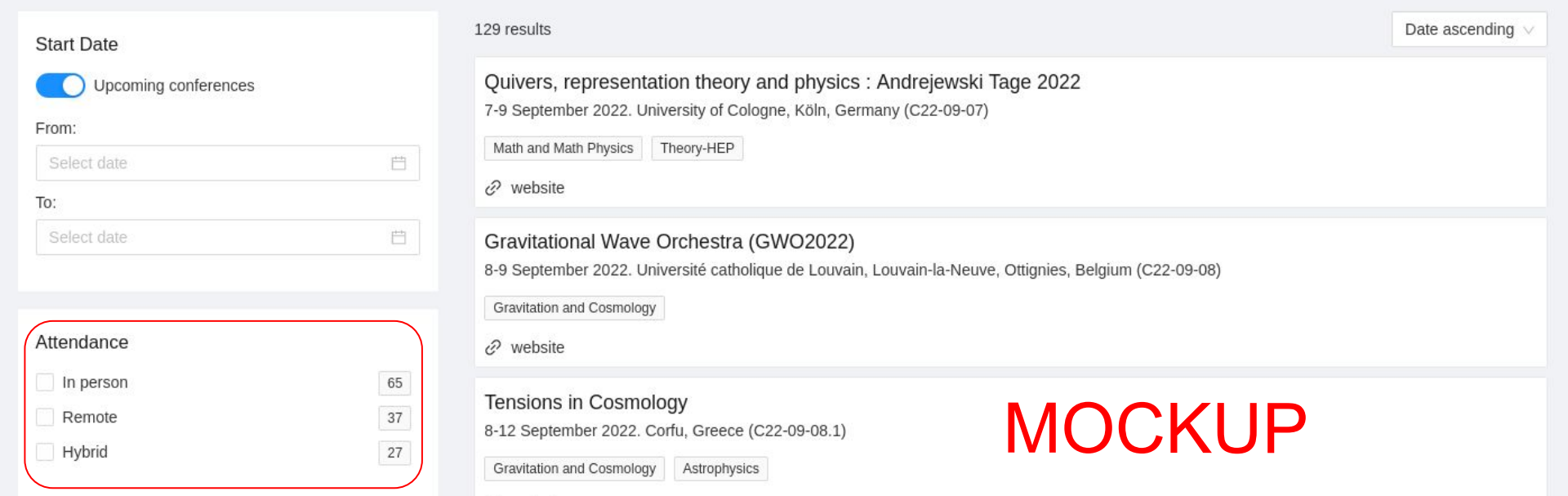
Prioritization and design of a new is a multi-dimensional optimization problem, taking into account

- Impact on users (number of users, importance)
- User experience
- Development complexity
- Operational complexity
- Curation needs

Ideally: significant positive user impact, simple implementation, no operational or curation burden

Idea for second order reductions: conference travel

- Physicists travel a lot, can we make it easier for them not to travel that much?
- “Incentivize” remote & hybrid conferences? (better discoverability)
- But some controversy following talk yesterday by Gabriela and Daniel



What about computing efficiency?

- Efficiency of data centers is increasing (compute/W)
- On our side, energy efficiency is a secondary concern
- Often positively correlated with user experience (efficient ~ fast ~ good UX)
- Negatively correlated with development complexity
- Interactions with other systems might make things hard

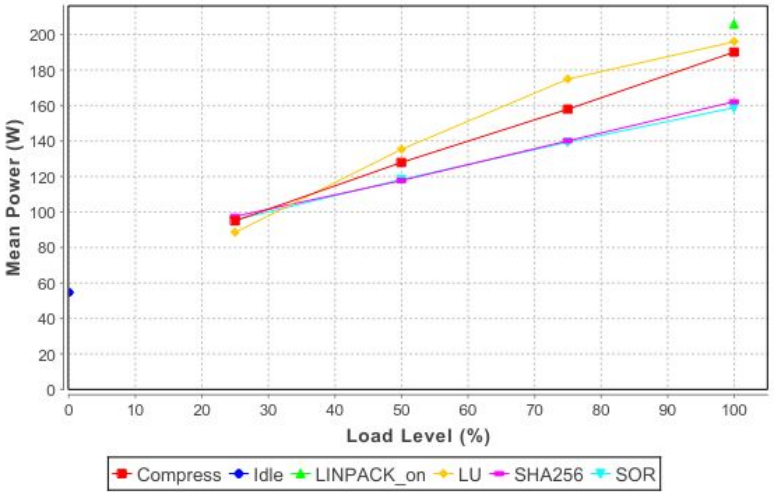
Examples:

- Efficiency is secondary: fulltext search
- Fast is not efficient: pre-computing citation formats
- Efficient is complex: citation links computed only in one directional
- Boundary conditions: arXiv updates have day granularity

Increasing efficiency: auto-scaling

(Not currently in use for SIS cluster)

- Idle power usage is significant
- More efficient to run fewer servers at higher load
- Combine 2 approaches
  - Horizontal Pod Autoscaling: start/stop containers based on needs (e.g. #requests)
  - Cluster Autoscaling: add/remove nodes depending on resources required by containers



ICPE 2016 Proceedings, p.147

3. Academic publishing

How big is academic publishing?

In 2020:

- Global scholar publishing market: 26.5B USD (STM report)
- Global GDP: 84.91T USD (World Bank)
- Global R&D: 2.63% of GDP (World Bank) = 2.23T USD
- ⇒ Scholar publishing marked = 1.25% of R&D spending

Main sustainability initiative: SDG publishers compact

Signatories to the SDG Publishers Compact commit to:

1. **Committing to the SDGs**: Stating sustainability policies and targets on our website, including adherence to this Compact; incorporating SDGs and their targets as appropriate.

2. **Actively promoting and acquiring content** that advocates for themes represented by the SDGs, such as equality, sustainability, justice and safeguarding and strengthening the environment.

3. **Annually reporting on progress towards achieving SDGs**, sharing data and contribute to benchmarking activities, helping to share best practices and identify gaps that still need to be addressed.

4. **Nominating a person who will promote SDG progress,** acting as a point of contact and coordinating the SDG themes throughout the organization.

5. **Raising awareness and promoting the SDGs among staff** to increase awareness of SDG-related policies and goals and encouraging projects that will help achieve the SDGs by 2030.

6. **Raising awareness and promoting the SDGs among suppliers,** to advocate for SDGs and to collaborate on areas that need innovative actions and solutions.

7. **Becoming an advocate to customers and stakeholders** by promoting and actively communicating about the SDG agenda through marketing, websites, promotions and projects.

8. **Collaborating across cities, countries, and continents** with other signatories and organizations to develop, localize and scale projects that will advance progress on the SDGs individually or through their Publishing Association.

9. **Dedicating budget and other resources towards accelerating progress** for SDG-dedicated projects and promoting SDG principles.

10. **Taking action on at least one SDG goal**, either as an individual publisher or through your national publishing association and sharing progress annually.

https://www.un.org/sustainabledevelopment/sdg-publishers-compact/

SDG Compact Members

Big commercial publishers

- Springer Nature
- Elsevier
- Taylor & Francis
- Wiley

Learned societies

- IOP Publishing

Open access publishers

- MDPI

… and many more

Notable omissions

- APS (not exclusively a publisher)

SDG13 in practice

HOW WE SUPPORT CLIMATE ACTION

It has been a pivotal year for climate action at Springer Nature. We:

2020

Became carbon neutral for offices, fleets and flights

April 2021

Signed The Climate Pledge

August 2021

Appointed a Climate Action Officer and launched a Group-wide Environmental Policy

October 2021

Committed to set science-based targets

November 2021

Engaged in COP26 in partnership with others and hosted our second company-wide Green Week

December 2021

Completed Green Building Checklists across 14 of our largest sites

Our progress in 2021[1]

37% decrease in gross emissions

47% decrease in net emissions

9% decrease in total energy use

84% decrease in emissions from business flights

https://sustainablebusiness.springernature.com/2021/environment-managing-our-impacts.html

[1] When compared with 2020.

Research Insights

Increasingly, governments, funders, the community and more, are demanding to understand the value of university research. A significant way universities show commitment to sustainable development and societal impact is by using their powerful research capabilities in these critical areas. Explore our curated SDG13 Research Insights.

Climate Action hub

SSRN's Climate Action hub highlights curated, early-stage research related to Earth's climate changes due to global warming and the subsequent large-scale shifts in weather patterns, across many disciplines that may inform the ongoing conversation about climate change and global warming.

Climate change showcase

Explore a selection of Elsevier research relating to SDG13, climate change and global warming, explained in plain language by professional science writers.

Cell One Earth

One Earth, a Cell journal that seeks to understand and address today's environmental grand challenges, has partnered with Gold Standard , a standards body that certifies initiatives that support the SDGs. Together, they offset three tons of carbon emissions for every research paper the journal publishes.

Environmental Standards

To help achieve our climate targets, the RELX Group Environmental Standards program sets benchmark performance and inspires green competition between offices. RELX has also published a Climate Change Statement , and adheres to a Global Environment Policy .

Inside Elsevier

Sustain is a global network of Elsevier staff aiming to create a global network to embed sustainability into Elsevier's work at all levels. Sustain works with the Green Teams, supporting and promoting sustainable action within our different locations, including Elsevier's headquarters in Amsterdam.

https://www.elsevier.com/about/corporate-responsibility/climate-action

A parallel: Open Access adoption

A brief history of Open Access

- 1989: www invented by Tim Berners-Lee at CERN
- 1991: arXiv founded by HEP researcher Paul Ginzparg
- 2001: Budapest Open Access Initiative

  “An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds.”

A brief history of Open Access

- 2018 status: 28% of articles with DOIs are OA
- What happened in the meantime?

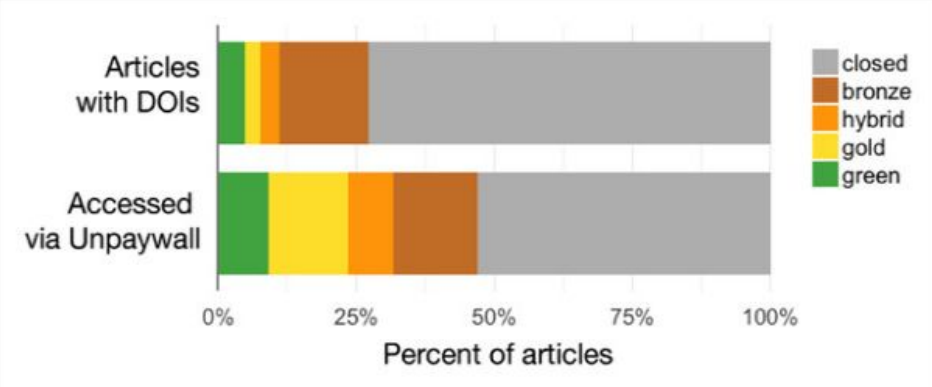
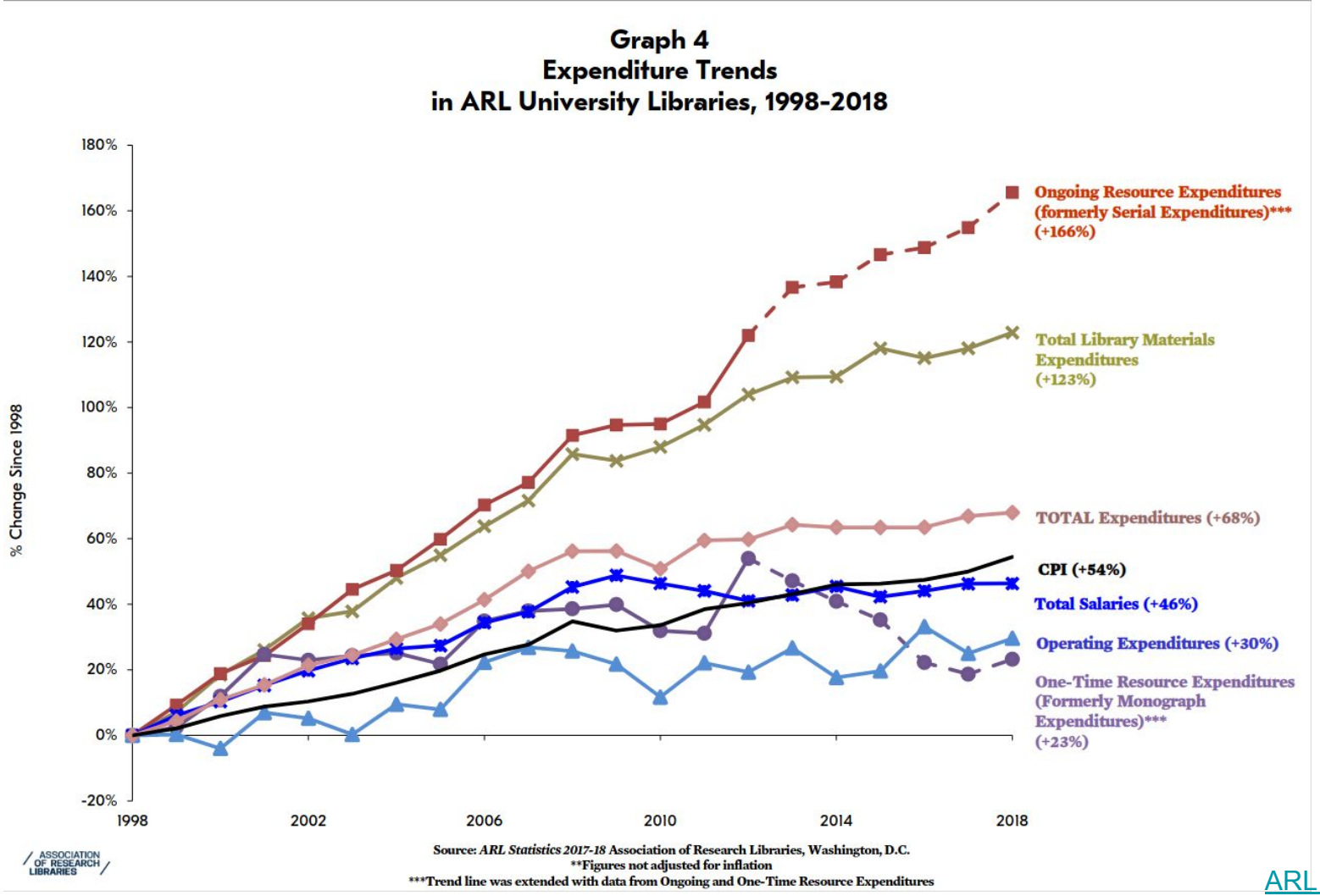


Figure 1: Percent of articles by OA status, Crossref-DOIs sample vs Unpaywall-DOIs sample.

Download full-size image

DOI: 10.7717/peerj.4375/fig-1



ARL statistics

Monetizing Open Access

- Most publishers are commercial entities
- Gold OA: authors pay article processing charges (APC) for article to be released
- Green OA: authors put article (or final manuscript) in institutional repository, often embargo before public release
- Institutions often pay 2x for publishing costs
  - Subscription cost to journal
  - APC for publishing
- Many researchers can’t afford to pay APCs

About

What is cOAlition S?

On 4 September 2018, a group of national research funding organisations, with the support of the European Commission and the European Research Council (ERC), announced the launch of cOAlition S, an initiative to make full and immediate Open Access to research publications a reality. It is built around Plan S, which consists of one target and 10 principles.

cOAlition S signals the commitment to implement the necessary measures to fulfil its main principle:

"With effect from 2021, all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo." https://www.coalition-s.org/about/

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Dr. Alondra Nelson
Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.

White House memo

What do we learn?

- Economic goals of publishers not necessarily aligned with greater good
- Publishers can take good ideas and find a business model around it
- Going against financial interest requires regulations from large research funders

What could publishers do?

- Be more transparent about sources of emissions by activity
- Go fully paperless
- Reduce in-person meetings
- Lower computing related emissions

See also Scholarly Kitchen

4. Conclusion

Summary

- Useful to consider second order effects given importance of scientific information in research
- INSPIRE GHG emissions are relatively low, focus mostly on second order
- Looking for ideas: feedback@inspirehep.net (or contact me personally)
- Publishers are working towards SDG as part of SDG compact
- Implementations differ widely in practice
- Going against financial interests requires institutional action
- Lack of detailed data makes decision-making difficult