



# Knowledge Exchange consensus on monitoring Open Access publications and cost data

Report from workshop held in Copenhagen 29-30 November 2016



## Acknowledgements

This report is the effort of many. All participants at the workshop have a stake in it. All of them contributed to the lively and constructive discussions in the breakout sessions leading to useful recommendations and we therefore owe a debt of gratitude to those who participated in the workshop in Copenhagen with special thanks, of course, to the four keynote speakers: Stuart Lawson, Rachael Lammey, Kai Geschuhn and Graham Stone and to those giving country presentations from all six Knowledge Exchange countries. Thanks to the KE Open Access expert group, to the KE Task and Finish group for the workshop coordinated by Christian Hagen Thomasen, the KE Office Bas Cordewener and Sarah James, the workshop moderator John Doove and not least to its hospitable host and organizer Michael Svendsen and The Royal Library in Copenhagen. All have fruitfully contributed to this report, although any mistakes and misreporting are the responsibility of the author.

© Knowledge Exchange 2017

Title: Knowledge Exchange consensus on monitoring Open Access publications and cost data

Subtitle: Report from workshop held in Copenhagen 29-30 November 2016

Authored by: Niels Stern

DOI: 10.5281/zenodo.480852

All content published can be shared  
(CC BY 4.0)  
[creativecommons.org/licenses/by/4.0](https://creativecommons.org/licenses/by/4.0)



# Contents

---

<b>Executive Summary</b>	<b>4</b>	<b>Monitoring Open Access cost data</b>	<b>26</b>
Lessons learned from the first Knowledge Exchange workshop on OA monitoring in Utrecht	4	Kai Geschuhn: From Offsetting to Pay-as-you-publish. Let's reintroduce transparency and competition to the scholarly publishing market!	26
The purpose of the second Knowledge Exchange workshop on OA monitoring in Copenhagen	4	Graham Stone: Collecting APC cost data and information	27
<b>Knowledge Exchange</b>	<b>6</b>	Break out session 2: Monitoring cost data – Recommendations from the working groups	28
<b>Setting the Scene</b>	<b>7</b>	1. Working group on Data	29
Keynote by Stuart Lawson: Revealing the true costs of publishing - Towards a public data infrastructure of scholarly publishing costs	7	2. Working group on Workflows	30
<b>Country presentations</b>	<b>10</b>	3. Working group on Standards	30
The Netherlands	10	4. Working group on Policy	31
OA Cost data	10	Some conclusions from the breakout session on monitoring of cost data	32
OA monitoring	11	<b>Next steps – beyond Copenhagen</b>	<b>33</b>
United Kingdom	12	1. Current Research Information Systems (CRIS)	33
Monitor Local	12	2. Publishers	33
Monitor UK	13	3. Libraries	33
Finland	14	<b>Annex A: List of workshop participants</b>	<b>35</b>
Germany	15		
France	17		
Denmark	18		
<b>Monitoring Open Access publications</b>	<b>20</b>		
Keynote by Rachael Lammey: Crossref and OA Metadata	20		
Break out session 1: Monitoring OA publications – Recommendations from the working groups	22		
1. Working group on Data	22		
2. Working group on Workflows	23		
3. Working group on Standards	23		
4. Working group on Policy	24		
Some conclusions from the breakout session on monitoring OA publications	25		

---

# Executive summary

## Lessons learned from the first Knowledge Exchange workshop on OA monitoring in Utrecht

The initial Knowledge Exchange (KE) workshop **Monitoring of Open Access** held in Utrecht in May 2015 concluded that Open Access definitions, workflows and collaboration should be closely linked in order to keep monitoring simple and open, allowing best practice to emerge and be built upon.

The aims of the KE workshop in Utrecht were to explore sustainable open access approaches at policy, service and business levels for all research outputs and information, to the benefit of the communities that KE serves. The workshop brought together 32 participants, representing Open Access policy-makers and practitioners from the KE partner countries (Ireland, Norway, Belgium and Spain) and beyond. They shared ideas and plans for monitoring open access, focusing on methods, tools and workflow, and discussed the questions that arose.

The participants concluded that it would be advisable for the sake of international comparison to foster cooperation between countries, eg on standards and identifiers. Furthermore a strong wish was expressed for enabling comparison and aggregation internationally and for keeping workflows simple and data open.

The workshop in Utrecht closed with three suggestions for future work:

1. Summarize questions, challenges and compare different approaches
2. Make recommendations on best practices and work out some scenarios
3. Agree on a minimum standard and start an ad hoc working group on exchange formats

## The purpose of the second Knowledge Exchange workshop on OA monitoring in Copenhagen

The second workshop on monitoring Open Access publications and cost data related to publications, which is the focus of this report, was a direct response to the suggestions made in Utrecht.<sup>1</sup>

Hence this second workshop was designed to create some tangible recommendations by gathering experts from all the KE countries and beyond to discuss and compare results and ongoing experimentation in the fields of:

- ▶ Monitoring OA publications
- ▶ Monitoring cost data for OA publications

The Copenhagen workshop successfully addressed both topics in a number of ways. Keynote presentations gave useful and general overviews of ongoing initiatives and results while presentations from six different countries offered unique insights into the most prominent activities regarding monitoring of OA publications and related cost data. Finally, and most importantly, two breakout sessions involved all 57 participants in discussions about monitoring, leading to 48 concrete recommendations.

The recommendations are all found in this report. In summary the recommendations on monitoring OA publications are:

- ▶ Standards and common definitions are crucial
- ▶ Standards already exist to a large extent, eg in Common European Research Information Format (CERIF) and Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- ▶ If new standards are needed they should be added to the existing protocols

- ▶ With the standards and definitions in place, policies and agreements can require publishers to deliver data in ways that make the workflows open and transparent
- ▶ Current Research Information Systems (CRIS's) can be used as sources for monitoring OA publications and ensure that the monitoring data is open through open API's (application programming interface) so that monitoring results can be validated, thus ensuring transparency and reproducibility
- ▶ Finally, libraries should play an active role in this area
- ▶ Total costs of publication (TCP) is a key concept. It is important to dissect the costs of publishing carefully. The Article Processing Charge (APC) does not necessarily cover all costs, eg administrative costs, infrastructural costs, special extra charges set by the publishers on a per publication basis etc
- ▶ Transparency and access to the cost data is crucial

In summary, the recommendations relating to monitoring on cost data are:

- ▶ Accounting systems and CRIS's are central to the topic
- ▶ These systems should be interoperable and aligned so that cost data at all levels can be easily retrieved
- ▶ The data should be open and shareable
- ▶ The Digital Object Identifier (DOI) is a key tool for data transferal between systems giving Crossref an important role in the workflow
- ▶ Publishers should be required to enter all funder data in the publication metadata as well as in the publications themselves
- ▶ Such requirements should be embedded in offsetting or licensing contracts with the publishers
- ▶ Non-disclosure regulations in these contracts should be avoided at all times

---

#### Footnotes

- 1 For an overview of selected references on monitoring Open Access, a Zotero bibliography has been set up: [zotero.org/groups/apcbibliografi/items](https://www.zotero.org/groups/apcbibliografi/items). Please feel free to contribute relevant references.
-

# Knowledge Exchange

Knowledge Exchange was founded in 2005 as a partnership to foster cooperation and the exchange of knowledge between four national organisations responsible for IT in higher education and research: the German Research Organisation (DFG), SURF in the Netherlands, Jisc in the United Kingdom, and Denmark's Electronic Research Library (DEFF). They were later joined by CSC from Finland, and in 2017 the network was enlarged with a new member: the National Center for Scientific Research (CNRS) from France.

The organisations work together to support the use and development of ICT infrastructure for higher education and research. Although the organisations differ in the range and scope of their work, each has a nationwide responsibility and influence on national policy, operates at the cutting edge level of IT development, and can mobilise resources to make a difference.

Since its creation Knowledge Exchange has focused on open access and research data with open scholarship as the most recent extension. The mission of Knowledge Exchange to enable open scholarship by supporting an information infrastructure at an international level. The members do this by:

- ▶ Inspiring each other with new approaches
- ▶ Enhancing current practice in research and higher education using IT
- ▶ Improving the infrastructure and services available to scholars and researchers
- ▶ Stimulating productive networks
- ▶ Creating and sharing knowledge collaboratively
- ▶ Exploring better ways to exchange knowledge

Knowledge Exchange carries out a number of activities in order to achieve all of this, for example, by commissioning surveys, studies and reports on focus areas or emerging topics and by arranging workshops, meetings and seminars for experts and stakeholders.

# Setting the scene

The host and organiser of the conference, Michael Svendsen (The Royal Library, DK) gave a warm welcome to this second Knowledge Exchange workshop on monitoring of open access (OA) publications and cost data. It was a truly international workshop with 57 participants from eleven different countries representing 41 institutions. He set out the goal of the workshop as a push for transparency with the aim of influencing evidence based policy making and promoting better outcomes in negotiations with publishers.

**Figure 1: Opening of the workshop**



The goal of the workshop is to come up with recommendations that can:

- ▶ Push for transparency
- ▶ Influence evidence based policy making
- ▶ Promote better outcomes in negotiations with publishers

## Keynote by Stuart Lawson: Revealing the true costs of publishing - Towards a public data infrastructure of scholarly publishing costs

Stuart Lawson is a doctoral researcher at Birkbeck, University of London, and is affiliated to the Open Knowledge Foundation. He has previously worked as a research analyst for Jisc. In April 2016 he co-published the article “Opening the Black Box of Scholarly Communication Funding: A Public Data Infrastructure for Financial Flows in Academic Publishing”<sup>2</sup> on which this keynote was partly based. He opened his talk by summarising the serials crisis and other crucial developments in journal publishing over the past two decades.<sup>3</sup>

In his talk Lawson gave special attention to the aggregation of journals by a few large publishers and addressed the negative consequences of an increasingly oligopolistic and dysfunctional market dominated by a handful of very big scientific publishers. This market has partly fuelled the development of hybrid journals where APC's are on average considerably higher than those for full OA journals. At the same time there is an increased risk of publisher ‘double dipping’, ie when the publisher collects an APC for a given article and continues to charge the same subscription fee for the journal in which the article sits. In other words, the publisher gets paid twice since the APC expenditure is not offset against the subscription price. This has attracted a lot of attention from libraries and funders and is part of the reason why these actors are hugely engaged in offsetting arrangements with publishers, as we shall see in a number of ways throughout this report.

### Footnotes

<sup>2</sup> <https://olh.openlibhums.org/articles/10.16995/olh.72>

<sup>3</sup> The serials crisis refers to the fact that the prices of the institutional or library subscriptions to scholarly journals have been rising much faster than the Consumer Price Index for several decades, while the funds available to libraries have remained static or declined in real terms.

In his introduction Lawson also demonstrated the complexity of APC pricing due to the number of different models. These include:

- ▶ Prepaid APC's
- ▶ Discounts on list prices
- ▶ Membership models with upfront payments for all publications
- ▶ Additional charges connected to page count or colour print
- ▶ Sensitivity to an institution's provenance
- ▶ Prices varying across journals
- ▶ APC's paid for in bundles – so called 'big deals' – with individual universities

Subsequently, the true costs of publishing are hard to get at since according to Lawson our knowledge of the APC expenditures are based on patchy samples of the full costs of Gold OA. Full transparency in the Gold OA market would require disclosure of how much each institution pays to each publisher for each journal, ideally relating these payments to public funds. Financial opacity has negative effects on the libraries' ability to effectively negotiate, for instance, offsetting deals with publishers.

*“The true costs of publishing are hard to get at since our knowledge of the APC expenditures are based on patchy samples of the full costs of Gold OA.”*

Stuart Lawson

Figure 2: Keynote by Stuart Lawson



So, what prevents us from gaining the full picture of the costs associated with scholarly publishing? According to Stuart Lawson there are a number of reasons:

1. Institutions have multiple income sources. This complicates our understanding of the extent of publicly funded Gold OA
2. The institutional financial management may obfuscate how income is spent on subscription and APC's respectively, making it even harder to aggregate information across institutions, hard to get a systemic analysis
3. Payment by individual researchers may go unreported. For instance, studies at the University of Nottingham have shown that up to 50% of the funding at this institution remains unseen
4. Policies mandate the reporting of payments differently

5. Policies foreground payments in some disciplines leaving others in the dark. For example, in the UK most funding from Research Councils UK (RCUK) and Wellcome Trust goes to Science Technology Engineering and Maths (STEM) research, whereas funding for the humanities and social sciences is scarce
6. Non-disclosure agreements prevent transparent cost evaluations. However, Freedom of Information requests have been a way to shed light on the payments. In this way Lawson has successfully gathered substantial information from Higher Education Institutions (HEI's) about their subscription expenditures and shared this information openly online
7. Comparisons between list prices and those APC's that were actually paid suggest that the actual costs can be significantly higher
8. For hybrid journals it is difficult to disentangle revenue from subscription fees and APC's – especially if the journal is part of a big deal which themselves generally obscure costs

Considering these obstacles one could get rather pessimistic. Fortunately, however, Lawson concluded his presentation by pointing to several routes towards a public data infrastructure of scholarly communication costs. One way of doing this could be through co-creation of public databases like the Open APC initiative.

Before inviting questions, Lawson posed a question himself: Would one global database of subscription expenditure be possible or desirable? There were several reactions to this. Some questioned the need to collect information on subscriptions when the subscription based model is rapidly disappearing. Rather, it was stated, the focus should be on collecting APC data. Lawson did not quite agree on this point. According to him subscriptions will be in play for at least another ten years and therefore we need the data. It gives a better understanding of the

market and it empowers libraries when negotiating with publishers. These views were supported by others, raising the point that a lot of people outside libraries don't know how much libraries are paying for subscriptions. When these figures are disclosed people begin to react. Finland is a good example of how the disclosure of journal subscription costs at Finnish libraries has had an impact amongst the researchers leading to big protests (see the country presentation for Finland). So, indeed, revealing the true costs of subscriptions and APC's definitely has an impact.

# Country presentations

After the keynote presentation by Stuart Lawson, presentations were given for each of the six Knowledge Exchange countries. The focus of the presentations was the workshop themes, ie monitoring of OA publications and cost data. In all countries activities around monitoring are ongoing, albeit at different levels and in uneven ways.

The presentations, therefore, offered a helpful overview of the situation in the KE member countries. As such they also gave a solid basis for the formulation of recommendations through the breakout sessions that followed on the second day of the workshop.

## The Netherlands

The Dutch presentation was given by Robert van der Vooren (Association of universities in the Netherlands - VSNU) and Just de Leeuwe (UKB). The VSNU is formed by the 14 Dutch research universities. UKB is the Dutch consortium of the 13 university libraries and the National Library of the Netherlands.

### OA Cost data

The political ambition in the Netherlands is to achieve 100% Gold Open Access by 2024. This goal is backed up by the 14 universities that constitute VSNU. The strategy to achieve this goal is based on offsetting negotiations with the publishers. VSNU has been negotiating big deals with primarily the big publishers. For the Netherlands the eight biggest publishers (Elsevier, Springer, Wiley, Taylor & Francis, Sage, Oxford University Press (OUP), American Chemical Society (ACS) and Kluwer) are responsible for 70% of the published articles in the country. It was therefore decided to begin with these publishers. Later VSNU will negotiate with the long tail of smaller publishers.

Four parameters were highlighted as essential during the negotiations:

1. The unique bargaining model (all Dutch universities forming one block)
2. A powerful negotiation delegation
3. Consistent adherence to principles
4. Clear political support

The deals that have come out of the negotiations vary quite a lot. With Springer and Taylor & Francis, two-year big deals have been made including all publications, with Wiley and ACS these are four-year big deals, whereas with Elsevier and Sage only partial deals have been signed. Deals with OUP and Kluwer have not yet been agreed. The Springer deal is due for re-negotiation and the Sage deal will soon be due, too. Robert van der Vooren said that he expected all the Dutch big OA deal contracts to be disclosed by the end of 2016. These would then form a transparent basis for further conversations about OA publication costs. In a second wave VSNU will be negotiating with the smaller publishers. The belief and strategic approach of the VSNU is to focus on the total costs of publication (TCP) and not just APC's. In fact, smaller publishers like Amsterdam University Press would not be able to survive in an APC model. They are deeply dependent on including all costs in the publishing process. Another reason for the VSNU not to support APC based models is that the VSNU does not think that researchers should be bothered with market-related issues like APC's, but instead should focus on research.

### OA monitoring

Alongside the offsetting deals, UKB has begun systematic monitoring of OA publications in the Netherlands. A baseline report is expected to be available around summer 2017. When monitoring OA publications UKB has categorised open access into Gold, Hybrid, and Green. Gold is defined as an OA article in an OA journal listed in the Directory of Open Access Journals (DOAJ). Hybrid is defined as an article in a closed/restricted journal which is not listed in DOAJ, and Green is defined as articles in trusted repositories. Articles outside of these categories are considered non-OA. With these working definitions in place UKB has created a framework for monitoring OA (figure 3).

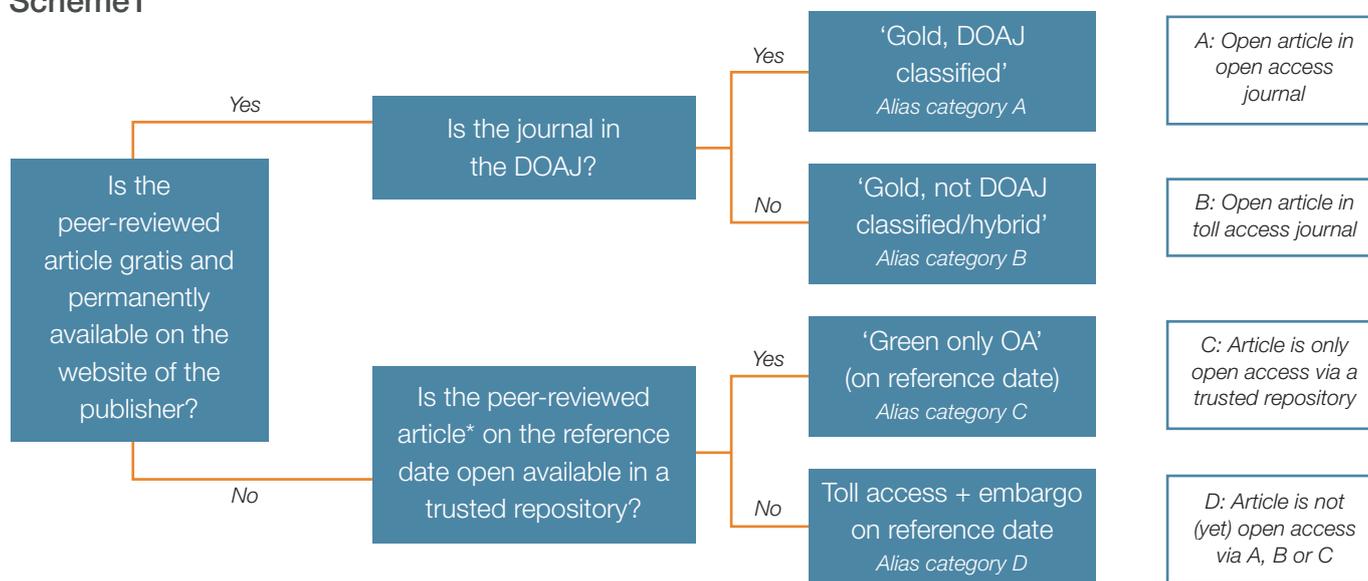
Certain requirements have been attached to the big deals in order to ensure high levels of OA. In the article submission process OA is the default, so the author does not have to opt-in or to care about vouchers, discounts etc. Authors are verified through their ORCID or through IP domain or email affiliation. Combined with clear information from the publishers UKB hopes to see a high level of OA output. However, we will have to wait until summer 2017 to ascertain the baseline open access percentage for the Netherlands. Until then UKB will be busy improving their existing workflows and setting up workflows for new contracts, eg with Taylor & Francis, ACS and others.

*“VSNU does not believe in APC’s. We believe in Total Costs of Publication (TCP).”*  
Robert van der Vooren

*“In the article submission process OA is the default so the author does not have to opt-in.”*  
Just de Leeuwe

Figure 3: OA monitoring workflow in the Netherlands

#### Scheme1



## United Kingdom

Frank Manista from Jisc in the UK presented two new monitoring services: Monitor Local and Monitor UK. They are standalone but connected products that can monitor APC's and funder policy compliance. Where Monitor Local provides an institutional level view of APC data, Monitor UK provides a national level view of APC data. Monitor Local feeds Monitor UK and both products are cloud-based services hosted by Jisc.

### Monitor Local

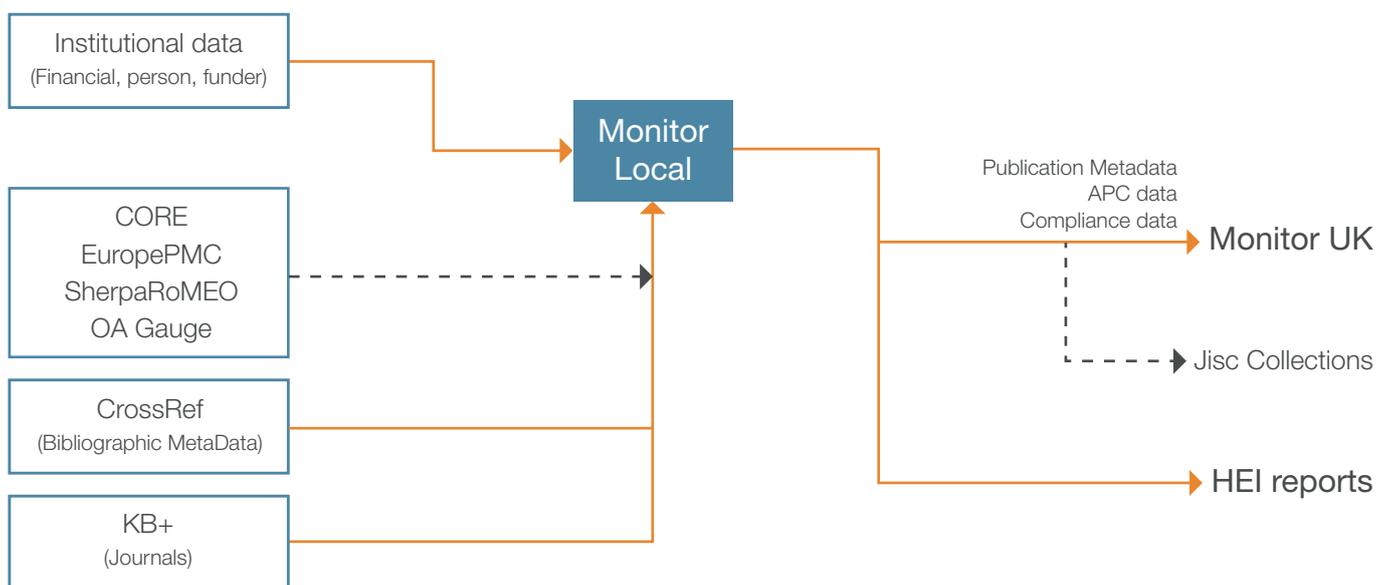
Monitor Local is fed by a variety of data sources:

- ▶ Institutional data (financial, person, funder)
- ▶ COncecting REpositories (CORE) – if materials are in a UK repository

- ▶ EuropePMC for full text, bibliographic and OA information
- ▶ Sherpa RoMEO for embargo and deposit policies
- ▶ Open Article Gauge for licensing and article version information - data from a variety of sources, including the targets of DOIs, plus APIs of the major publishers

It works for librarians, repository managers, research offices etc. as it enables them to record on data relating to the publication of open access outputs by their academics. This could be done for reporting purposes, for instance on APC's, policy compliance, Gold and Green publication routes. Besides these local purposes, Monitor Local also feeds into Monitor UK (see figure 4).

Figure 4: Jisc Monitor Local: data flows and outputs

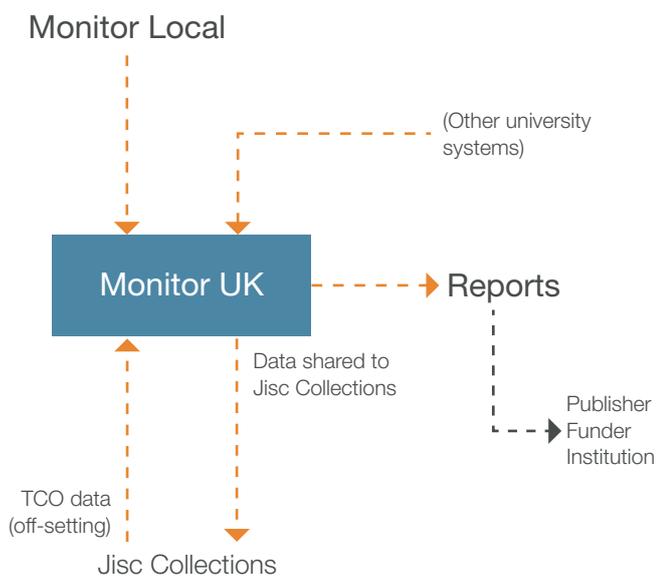


### Monitor UK

Monitor UK aggregates nationwide APC data. Theoretically the monitor could be made global but Jisc wants to see how it works in the UK before setting up a business case. It is fed by Monitor Local but it can also absorb data from other sources. This remains necessary since some institutions have 'home-grown' APC/ Compliance solutions which means they do not want to adopt Monitor Local until it surpasses their highly customised solutions. It provides expenditure reports per funder, institution or publisher, eg total number of APC's, total APC expenditure and average APC cost. Therefore it is of interest to UK HEI's and funders for the benchmarking and evaluation of cost and compliance data. The data is also utilised by Jisc Collections and is easy to retrieve through a web interface (see figure 5).

The development of the two monitors ended in August 2016. During the pilot 23 HEIs took part. The pilot was followed by a transition period with the goal of bringing the monitors into service and transforming the pilot participants into early adopters. At the time of presentation Frank Manista could report that fewer than the 23 pilot participating HEIs were currently using Monitor Local. Therefore the next steps for the monitor would have to focus on converting the 23 pilot HEIs into early adopters, approaching the non-pilot HEIs and continuing discussions with funders around compliance monitoring and the role both products could play. Finally Frank Manista outlined some of the developments in the pipeline for the Monitors, eg User interface and user experience enhancements for the front end and back end integration with the Jisc Total Cost of Ownership (TCO) spreadsheet.

Figure 5: Jisc Monitor UK



*“Funders want to see how institutions are compliant with their policy. This is now possible with our Monitor.”*  
 Frank Manista

## Finland

In Finland an ambitious open science agenda – the Open Science and Research Initiative – is ongoing and due to end in 2017. According to Jyrki Ilva of the National Library of Finland the agenda has been a top-down initiative with a primary focus on research data and less focus on OA publications. However, the goals of the Open Science and Research Initiative are more ambitious than those 19% of peer-reviewed articles produced by the Finnish research organisations' OA in 2015. It has been announced that OA could become a factor in the funding model for Finnish universities which could then lead to an increase in the extent of OA. Still, there have been issues with the definitions and categories for counting OA publications in Finland. As of 2016 improvements have been made in this field and henceforth the figures – collected by CSC from national CRIS's – should give a more accurate picture of the status in Finland. However, the improved definitions include Green and Gold OA (including Hybrid), and exclude papers available on personal/project websites or at services like ResearchGate or Academia.edu.

In parallel with these attempts, the Finnish consortium for licence negotiations, FinELib, is undertaking offsetting negotiations with Elsevier and other major publishers. These negotiations are unexpectedly fuelled by bottom-up protests from angry researchers who have proclaimed that they will abstain from refereeing and editorial duties for the journals of the publishers involved in the offsetting negotiations if the goals of the Finnish negotiators are not realised. More than 2,500 researchers have supported the statement so far. This situation coincides with the disclosure of scientific journal subscription costs in Finland from 2010-2015. The disclosure of annual subscription costs was the result of a court decision that these costs are public information, according to the Finnish Act on Openness these costs are made for public information. Open Knowledge Finland activists have previously sued one of the Finnish universities, which had declined to provide information on the costs of its licensing deals, and OKF were supported by the courts.

*“Finnish researchers who are angry about the substantial price increases on scholarly journals have fuelled FinELib’s bargaining power vis-à-vis publishers.”*

Jyrki Ilva

The increasing transparency of cost data in Finland does not, however, include APC data. The universities have had no centralised tracking system for APC's and the financial administration often has no category for APC's, since they are paid from different accounts, and invoices will typically be sent directly to researchers. Furthermore the monitoring of APC's has been out of scope for the CRIS's. Many of the universities have now started to adjust their systems and processes to enable the monitoring of APC's since reliable OA data would be very useful for the national offsetting negotiations as well. Among libraries there have been conversations about taking part in the Open APC project but the Ministry for Education and Culture is still undecided on whether it would be willing to fund an Open APC project.

Finally Jyrki Ilva mentioned a new initiative of building a consortium for Finnish OA monographs influenced by the Knowledge Unlatched model. As for journals that migrate from Toll Access (TA) to OA, a consortium based funding model is envisioned. Several studies are currently taking place in Finland.

*“It has been announced that OA could become a factor in the funding model for Finnish universities which then could lead to an increase in the extent of OA.”*

Jyrki Ilva

**Figure 6: Finnish academics are protesting against the high cost of scientific journals**



## Germany

Dirk Pieper of Bielefeld University Library gave insights into the German OA monitoring landscape by focusing on the DFG supported project INTACT – Transparent Infrastructures for Article Charges. But first he quoted Richard Walker who stated in a blog post<sup>4</sup> (25 November 2016) how important especially the Max Planck Society is for the development of OA in Germany, eg by hosting the Berlin Conferences, publishing the white paper on disrupting journals' subscription models<sup>5</sup>, having the biggest APC spend in Germany, entering the first offsetting agreement in Germany, starting the OA2020 initiative and so on. However, several other German actors, not least the DFG and the Alliance of Science Organisations, do also contribute significantly to the development of further monitoring of OA publications and cost data according to Dirk Pieper.

In this context and supported by the newly published German strategy for Open Access<sup>6</sup> which also addresses the need for and intention of creating a German OA monitor, the INTACT project was presented. It consists of three initiatives:

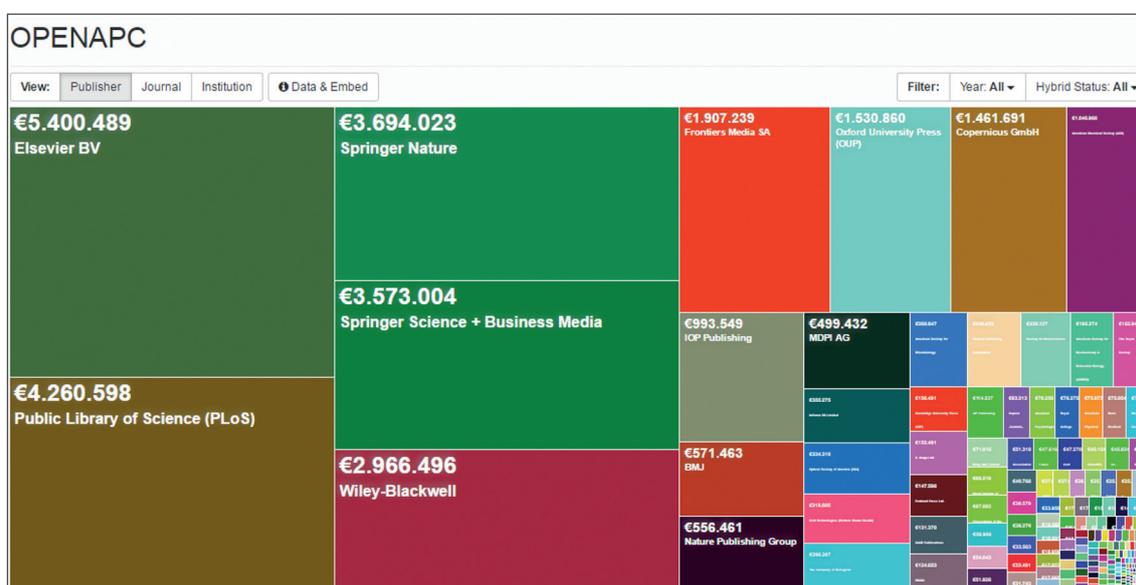
1. Open APC for reporting and transparency
2. Efficiency and Standards for Article Charges (ESAC) for workflows, efficiency, and publisher communication
3. OA analytics for publication data and bibliometrics

The INTACT project is truly international, collecting data on a global scale, eg APC data from Harvard University, the Austrian Science Fund (FWF), OpenAIRE, FP7, Wellcome Trust, and offsetting data from Dutch Universities, FWF, Jisc, Max Planck Society (MPG), and Swedish Universities. All the data, the documentation and reporting codes are collected open source on Github<sup>7</sup> and visualised using treemaps.

## Footnotes

- 4 <https://blog.frontiersin.org/2016/11/25/open-access-in-germany-improving-all-the-time>
- 5 Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015): Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. <http://dx.doi.org/10.17617/1.3>
- 6 [bmbf.de/pub/Open\\_Access\\_in\\_Deutschland.pdf](http://bmbf.de/pub/Open_Access_in_Deutschland.pdf)
- 7 <https://github.com/OpenAPC>

Figure 7: An example of a treemap from the German OpenAPC project.



On the day of the presentation (29 Nov 2016) the INTACT project showed the following cost data for pure Gold OA articles based on 38 German institutions: 9,778 articles at a total of €12,777,728, with a mean value of €1,273 and a median of €1,262. Although these numbers do not represent the whole picture (eg they do not contain information on Hybrid OA) the information can be valuable for libraries when negotiating offsetting deals. The uncertainty of the numbers derives from the fact that the Open APC data does not cover all APC expenditures since the funding can come from a variety of sources that are not necessarily captured by the Open APC project. However, despite those uncertainties it is evident that INTACT has the potential to aggregate and normalise publication and cost data on an international level supporting large scale transition to Open Access.<sup>8</sup>

*“The main difference between UK and Germany is that the UK is more engaged with Hybrid OA.”*

Dirk Pieper

## Footnotes

8 <http://oa2020.org>

## France

The new law for “a digital republic” in France (7 October 2016) ensures that the French researchers retain the right to make their articles available online after a maximum embargo of six months for scientific, technical, and medical (STM) literature and 12 months for humanities and social sciences (HSS). Although there is no national mandate there is a clear preference for Green OA in France. The Gold route is only used in a few disciplines (eg health) but is generally not very popular. No policy for Gold OA exists and there is no APC cost management system in place at the institutional level. As such it is hard to obtain information about the Gold OA publishing practice in France. However, in 2016 Couperin retrieved data about Gold OA from various sources, eg Web of Science, DOAJ, and publisher websites. This data shows an increase of 84% in the number of OA articles in France over the period 2010-2014, reaching a high of 12% of the total number of scholarly publications (approx. 75,000) in 2014.

Very little data has yet been collected in France. Sandrine Malotaux of Couperin emphasised the difficulties of obtaining data on APC expenditures in France due to a complex combination of centralised and decentralised organisational structures. Through five pilots in 2016 Couperin has surveyed institutional expenditure on APC's at four universities and at CNRS. Further studies are planned for 2017 using accounting data to understand APC expenditure. This is a difficult and laborious approach which involves manually checking invoices at a variety of institutions. However, this must be done because the Web of Science and publisher data is only useful for checking the publications but not for APC expenditures.

Jean Francois Lutz from the University of Lorraine illustrated the French difficulties by showcasing how the University of Lorraine has been collecting information on APC expenditures from 2012-2015. The data has been collected from bibliographic databases (Web of Science), through publisher and aggregator data, and by using accounting software. The amount of data collected has been relatively small but testing the methodology for collecting data has been useful. What is perhaps most striking about this test are the significantly different results that Lutz has obtained using the different methods. This shows how complex it actually is to collect the data, and how significant the variations can be depending on the method chosen. It also highlights the necessity for international alignment on monitoring methods in order to make meaningful comparisons between countries.

Using the three different methods also demands different levels of effort. Lutz found that using bibliographic databases is the easiest way of working and good for comparison between institutions, but does not yield such accurate data on expenditure. Collecting data through publisher platforms requires more work but is helpful for obtaining a complete overview (including for Hybrid OA), while not being so good at establishing accurate expenditure information. Finally, using accounting software is the most laborious approach but offers the most accurate expenditure knowledge. It is less appropriate for comparing institutions.

At the University of Lorraine the next steps are to initiate monthly monitoring, disseminate guidelines for researchers and in general to promote transparent and fair Gold OA.

*“The new law for “a digital republic” in France (7 October 2016) ensures that the French researchers retain the right to make their articles available online.”*

Sandrine Malotaux

*“Monitoring results vary greatly depending on the method used for collecting the data.”*

Jean Francois Lutz

**Figure 8: At the University of Lorraine they want to promote fair Gold OA to avoid this situation**



## Denmark

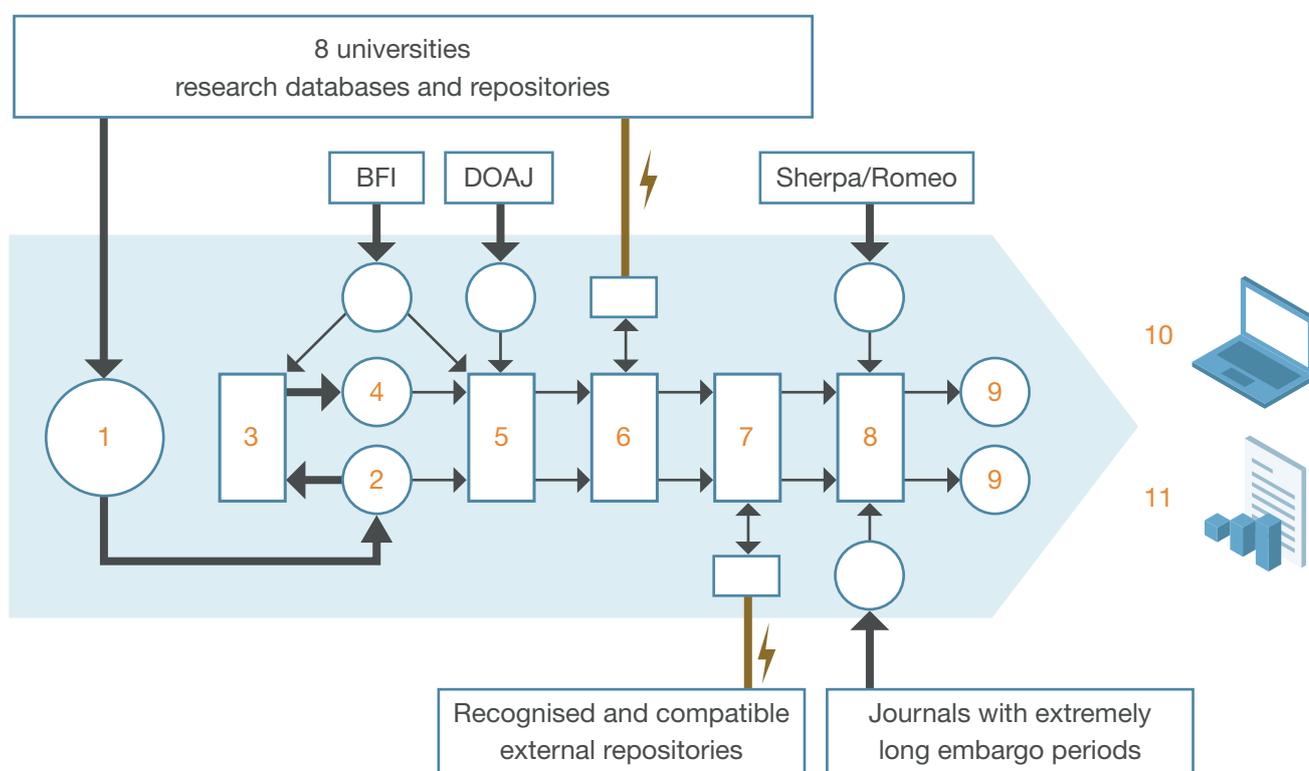
Mogens Sandfær from the Technical University of Denmark presented the Danish National OA Indicator. The indicator is a service of the Danish Agency for Science and Higher Education and the development was funded by DEFF. Since June 2014 its mission has been to monitor the fulfilment of the Danish national OA strategy. This strategy is Green, however the indicator monitors both Gold and Green OA peer reviewed articles and conference proceedings, but not Hybrid ones.

The data is collected according to a well-structured and transparent methodology explained by Sandfær, beginning with the publication metadata which is collected automatically from universities. A subset corresponding to the definition of the OA Indicator's target field is then isolated and forms the "target field with duplicates", as publications with authors from several universities will

be collected more than once. This version of the target field is used as basis for calculations dealing with individual universities. Deduplication of these records is then processed using data from BFI (the Danish Bibliometric Research Indicator). This is used as the basis for calculations at a national level and by main research area. Then publication details are checked to establish whether the article is published in a dedicated and scientific Gold OA journal. Here BFI-data as well as data from Directory of Open Access Journals (DOAJ) are used. There are also checks on whether the article may be downloaded from a Danish university's research database or a trusted OA repository. Finally it is established whether the article is published in a journal with Green OA potential using data from the Sherpa/RoMEO-database. In addition, a list of journals with extremely long embargo periods is consulted in order not to claim Green OA potential for those publications that impose an embargo period of more than 12 months.

The resulting data can be displayed as: 1) Statistics by university, 2) Statistics by subject category, and 3) National statistics (using target field without duplicates"). The result is communicated via the Danish National Research Database website and via spreadsheets, which may be downloaded from the Danish National Research Database. Additionally, the underlying publication data are open and may be downloaded as spreadsheets.

Figure 9: Methodology for collecting data for the Danish OA Indicator



The OA Indicator is updated once annually in January and the current number (December 2016) shows that the proportion of OA publications in 2014 was 18% of the national total. Plans are in place to improve the indicator. For instance it would be more useful with monthly updates, with differentiation between Gold OA with and without APC expenditures, and with reference to global bibliographic data of higher quality and consistency than those currently available. This may also simplify the processing pipeline.

Mogens Sandfær concluded his presentation by reflecting upon the notion of **truth** when monitoring. It is crucial to remember that the answer to “How much

Open Access?” depends on definitions, data sources and calculation methods. Therefore it is important to be transparent about definitions and methods, and to share the data and documentation so that others may check the validity of the results and finally refrain from claiming to hold **The Truth**.

*“Be transparent about methods and definitions! Share the data! Refrain from claiming to hold the Truth!”*

Mogens Sandfær

# Monitoring Open Access publications

The second day of the workshop was devoted to group work for continued discussion and possibly development of efficient and sustainable standards and workflows for **measuring OA publications** and **collecting related publication cost data**.

Four key areas for these discussions had been chosen by Knowledge Exchange prior to the workshop making four working groups: 1) Data, 2) Workflows, 3) Standards, and 4) Policy. Each group had a chairperson and a minute-keeper appointed.

The morning breakout session was devoted to measuring OA publications. For the afternoon breakout sessions the participants stayed in the same groups to discuss the monitoring of cost data. Prior to each breakout session a keynote presentation was given.

To kick off the first breakout session Knowledge Exchange had invited Rachael Lamme from Crossref to give a keynote on monitoring OA publications.

## Keynote by Rachael Lamme: Crossref and OA Metadata

Since the creation of Crossref in June 2000 the growth in the use of DOIs has been overwhelming. More than 80 million Crossref links have been assigned to all sorts of scholarly content items generating persistent citation links to journal articles, preprints, books, chapters, and datasets all over the world. More than 5,500 publishers are members of Publishers International Linking Association (PILA) which is the non-profit, independent organization that operates Crossref.

Figure 10: Keynote speaker Rachael Lamme from Crossref



In addition to the basic cross-referencing service Crossref also runs other services like Crosscheck, Cited-by Linking, and CrossMark. Since 2012 they have also run a funding data search service (formerly known as Fundref) which is a registry of funders that can be attached to publications in the Crossref database and then queried back by funders and other interested parties. All the metadata is of the Cataloguing Cultural Objects standard (CCO), thus in the public domain. There has been a tremendous growth in the number of funders registered for this service, from 4,000 at the beginning to more than 13,500 international funders by November 2016.

The funder data is deposited in Crossref by the publishers, who get the data from the authors during the submission phase. Therefore, if the publisher actually enters the funder data it then sits in the publication metadata, which then makes it easy for funders to extract information about publications that they have funded. However, the data is only as good as the quality of the publishers' deposits, as Rachael Lammey pointed out.

When the publication metadata gets enriched with funding data, licence information (including embargo and archiving information), and ORCIDs, then funders, institutions, publishers and others can truly benefit from the Crossref API's to get advanced and detailed information about any given publication.

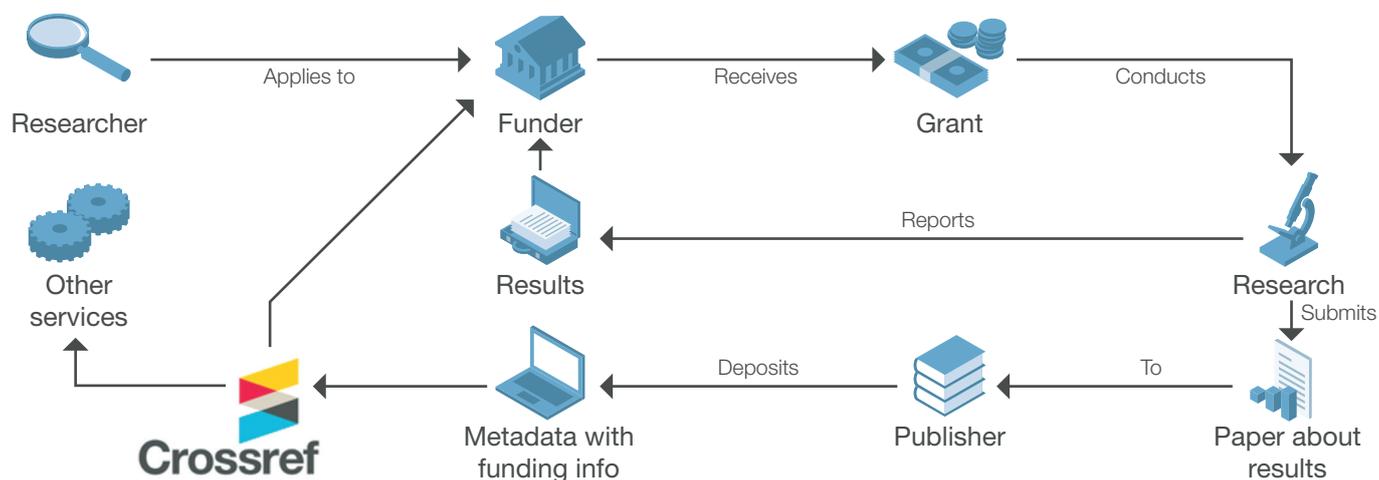
This ability to monitor publications is highly relevant for funders and institutions that want to check compliance with their funding mandates and policies. So far Crossref holds around 1.6 million pieces of content with funder information, of which 216,000 items have a Creative Commons Attribution (CC BY) licence attached. Crossref

encourages publishers to attach such information but they cannot require publishers to do so, nor can they impose certain standards for the information. It is up to the publishers to agree on the type of licences to be used, as Rachael Lammey clarified. Towards the end of her presentation she revealed that Crossref is considering how to include organisational identifiers in their metadata in order to improve their services for monitoring purposes.

Several of the participants applauded the important role that Crossref plays in this monitoring landscape. Later in the day the recommendations from the different working groups confirmed this perception.

*“Crossref was central to the development of ORCID and now we are working on an organisational identifier which we hope will become very useful, too.”*  
Rachael Lammey

Figure 11: How the funding data search works with Crossref



## Break out session 1: Monitoring OA publications – Recommendations from the working groups

Fuelled by the presentation from Rachael Lammey the working groups discussed four key aspects of OA publication monitoring: Data collection, workflows, standards and policy. These topics were chosen to address the topic of monitoring OA publications from different angles thus ensuring as comprehensive a picture as possible. With these four groups in place the workshop was expected to look at:

- ▶ The quality of data collected from available sources
- ▶ Efficiency in monitoring workflows
- ▶ Aggregation of OA publications and derived costs via standards
- ▶ Incorporating governance into monitoring processes for Open Access and related costs
- ▶ How policies work and if there is alignment across the policy landscape

### 1. Working group on Data

One size doesn't fit all, the group concluded, when it comes to the kind of data needed for monitoring Green, Gold or Hybrid OA.

For monitoring Green OA the group recommended the following for data collection:

1. Introduce metadata standards to improve the integration between CRIS's and repositories
2. Add to the metadata schema an identification of the corresponding author of an article - at least e-mail address and affiliation, ideally also ORCID. Similar identifiability of all (co)authors would be preferable

The metadata of journal articles (often from data sources such as Web of Science, Scopus, PubMed, but sometimes also from bibliographies made by national libraries in the domain of social science and humanities) and the deposited full texts in the CRIS's should be checked with Sherpa RoMEO to identify potential Green OA. These are journal articles for which the publisher allows deposit of the full text in an institutional repository, which the author(s) have failed to do. The authors are then contacted requesting that they deposit the text. Another possibility are policies and agreements that ask or require publishers to deposit metadata and articles on their own, eg to PubMed Central (PMC). That is already part of the author services program of some publishers (eg within the scope of deposit agreements with the National Institutes of Health (NIH) or through co-operation with Jisc Publications Router<sup>9</sup> or the DFG-funded project DeepGreen<sup>10</sup>).

The group identified an important hurdle: the rather low data quality of metadata in the repository, which makes synchronisation between the CRIS's and the repositories problematic. The group recommends aiming for better integration between CRIS's and repositories mainly relating to metadata standards, but also through two-way metadata linkage between CRIS's and repositories and metadata enrichment/ import mechanisms eg via Crossref. Specifically, the metadata for repositories should register the distinction between Green, Hybrid and Gold in the repository.

---

### Footnotes

<sup>9</sup> <https://pubrouter.jisc.ac.uk>

<sup>10</sup> <https://oa-deepgreen.kobv.de>

---

In order to facilitate communication with authors, it is important to know the corresponding author. Thus, this is an important recommendation for the metadata scheme of the CRIS and of the repository. This latter recommendation is also crucial for pure Gold and Hybrid Gold monitoring (see below).

For monitoring Gold OA the group recommended the following for data collection:

3. Match metadata of journal articles in the CRIS's with the DOAJ list in order to identify the pure Gold articles

For monitoring Hybrid OA the group recommended the following for data collection:

4. Include requirements for specific data delivery in offsetting agreements, eg the corresponding author, licence information, and the exact date of publication of the journal article. Such a description could be made generic and standardised (NISO) and made applicable for offsetting deals worldwide, for instance via Crossref
5. Publishers should use standardised data formats for author affiliation

For data on Hybrid OA, we remain dependent on publishers' data. Several publishers do provide this kind of data, but a number of workshop attendees reported rather important discrepancies in those data.

## 2. Working group on Workflows

An interesting question that was posed to this group was: Can we pool data from various countries into an aggregated picture for example by using common standards?

Reflecting on these questions, a number of challenges were immediately identified by the group members. For instance, solid infrastructures, like CRIS's, are fundamental to the workflow. Consequently, huge problems arise

without CRIS's in place. Secondly, and very importantly, persistent identifiers are needed to ensure reliable and comparable data and to achieve interoperability between monitoring systems. Additionally it was highlighted that persistent licensing information from publishers is a prerequisite for successful monitoring.

Other problems in the workflow were identified, eg getting information from publishers on the OA share of publications, about identifying corresponding authors, about monitoring Hybrid OA, and getting the right metadata sets for repositories.

Based on constructive discussions around these challenges and problems the group concluded its discussions with some recommendations:

1. Add to offsetting contracts (eg in terms and conditions) that publishers should include in Crossref a licence statement for each publication
2. CRIS's should integrate different categories of OA (this recommendation could be addressed to CRIS companies or EuroCRIS)
3. Monitor OA after embargo periods end
4. Use ORCID's in the workflow. In fact we believe that if ORCID's become widely adopted they can become the solution to workflow challenges, eg auto-updating ORCID profiles through Crossref

## 3. Working group on Standards

The group discussed the purpose of using standards when monitoring OA publications. Many articles miss metadata pointing to Open Access.

1. Use standards when depositing articles in the repository

This could be done through a library validation process which has been tested in Denmark. The group then also

raised the question whether information about non-OA also should be added into the registration.

For Hybrid OA there is a lack of standards. In an ideal world publishers would provide information about the OA status of a given article in a standardised format.

2. Publishers should deliver standardised information about whether Green, Gold or Hybrid OA containing at least licensing information – defined by the community – PIDs, not just DOIs, and metadata fields for rights information

This could be negotiated through library consortia.

3. Be very specific about what libraries need from publishers. Jisc for instance has developed a best practice guide for publishers available at: [http://ji.sc/publisher\\_guide](http://ji.sc/publisher_guide)
4. Commercial sources, such as those from publishers, should be CERIF compliant
5. Add a new field for APC/publication to the Open Archives Initiative (OAI) Protocol for Metadata Harvesting
6. For all stakeholders registering OA publications, we suggest: Let libraries help you!

There was general agreement in the group about using CRIS's for monitoring and reporting on OA publications. However, difficulties arising from different CRIS's and repositories in different countries should not hold back current efforts to monitor publications.

7. Explore the possibilities of using the persistent identifiers (PID) of CRIS's to connect systems as a way of cross-checking publications
8. Use CRIS's to report on the extent of OA

9. CRIS's should follow the OpenAIRE interoperability guidelines

Furthermore, the group found it important to agree whether data should be aggregated across countries and for what purpose. Is the purpose to report on the status quo of OA publications and related costs in Europe, or to be able to compare and analyse the variety in national situations? If the purpose is comparison, then appropriate definitions of Open Access are needed since the scope might be quite different according to the country. A need to define standards to identify Green, Gold, and Hybrid OA was also expressed.

#### 4. Working group on Policy

The group came up with a number of recommendations. There was agreement that monitoring should look beyond the current systems. This demands a certain flexibility since systems will most certainly change over time. There was also general agreement that the monitoring data should be open and transparent for others to use.

1. Monitoring should not only focus on the current system but should be flexible and adapt over time.
2. Policies should be clear about the rules (filters) applied to decide what data are eligible to be part of the monitoring policy and what data are not
3. Use monitoring not only to show compliance but also to indicate a desired direction
4. Collect as much data as possible and making it available without filtering
5. Use CRIS's as sources for collecting data on Green and Gold (by using identifiers/interoperability)
6. Measure more broadly by including more research outputs than simply articles (eg books)

7. Funders and all stakeholders should use clear definitions in their policies: i) The definitions should support filtering. ii) The definitions should help clarify whether the target goals are met, aligning the monitoring results across nations
8. Policies should state that the data connected to OA should be made openly available by providing an API

inclusion and auto-updating of ORCIDs. Finally, there seems to be agreement that libraries should play a very active role in this field.

The group also discussed issues around data and data gathering. There was a clear message to the publishers that the data should be open in standardized formats allowing for data mining. There was also a conversation about who should be collecting the data referring to best practices from The Netherlands (for Gold OA) and Denmark (for Green OA). There was also consensus that the alignment of monitoring mechanisms and definitions can have beneficial effects for comparing data.

### Some conclusions from the breakout session on monitoring OA publications

Standards and common definitions are crucial. These do already exist to a large extent, eg in CERIF and OAI-PMH so if new standards are needed they should be added to the existing protocols. However, some very basic and important definitions, like the concept of Open Access, are not yet in place. These are needed to make monitoring exercises comparable between countries. With the standards and definitions in place, policies and agreements can require publishers to deliver data in ways that make the workflows open and transparent. CRIS's can be used as sources for monitoring OA publications and ensuring that the monitoring data is open through open API's so that monitoring results can be validated thus ensuring transparency and reproducibility. An important tool for monitoring is the persistent identifier. DOIs have certain limitations mainly due to the fact that they are based on metadata entered by publishers. However, Crossref was still seen to play a vital role in the workflow for monitoring through the perceived

# Monitoring Open Access cost data

The second breakout session was devoted to monitoring cost data. As the country presentations showed this is a complicated field with a lot of experimentation going on. We saw that Denmark is currently not monitoring APC expenditures probably due to the fact that the national strategy is to promote Green OA.

In France experiments with collecting APC data are ongoing but prove very complicated. In Finland subscription cost data was recently revealed. VSNU promised that they would soon disclose their offsetting deals. In the UK Jisc presented their new monitor programme, and the OpenAPC project supported by the DFG in Germany showed impressive results collecting global APC data.

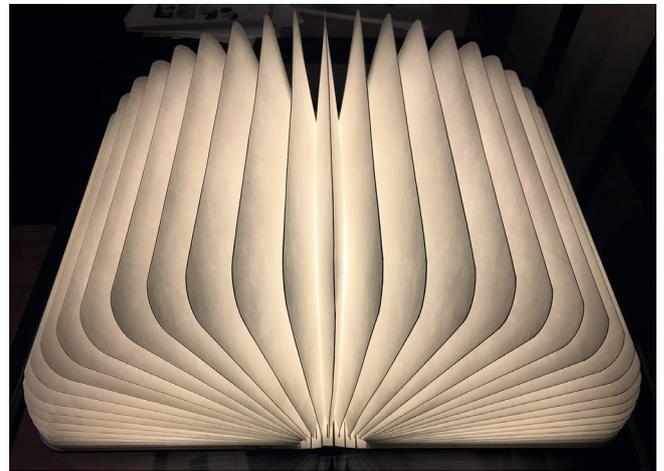
To kick off the second breakout session two keynotes were given. The first presentation was by Kai Karin Geschuhn from Max Planck Digital Library, who spoke about cost monitoring as part of the INTACT project. The second presentation was given by Graham Stone from Jisc, about collecting APC cost data and information.

## **Kai Geschuhn: From Offsetting to Pay-as-you-publish. Let's reintroduce transparency and competition to the scholarly publishing market!**

Kai Geschuhn introduced her presentation by stating that despite the many necessary and important efforts towards open access such as the green route, the several political mandates and open access advocacy strategies, the subscription system is still alive and kicking. The journal crisis has still not been overcome and as a result libraries are still facing substantial price increases year on year. In this regard, the offsetting part of OA negotiations, according to Geschuhn, for the first time really touch the subscription system. That's why offsetting deals seem to be a promising approach when it comes to finally achieving substantial progress towards Open Access. Offsetting may pave the way to a large-scale Open Access transition as outlined by the OA2020 initiative.

By shifting the business model of the existing corpus of scholarly journals we may soon reach the necessary tipping point of OA availability in order to make OA the default for scientific publishing, Geschuhn argued.

---



---

Following this track Geschuhn then showed the differences between the APC model and the subscription model. The emergence of a tool like Sci-Hub shows that the subscription system is broken. Subsequently, new business models are needed since it is much more disruptive for the business model if users move to illegal platforms. The APC model comes forward as such an alternative model.

The APC model is transparent as opposed to the subscription model which inherently relies on historical print subscriptions. Additionally, the APC model is

driven by the demand of the authors: publishers need to compete for authors. According to Geschuhn this is good for competition, fostering a positive market model. To get there she recommended a **pay-as-you-publish** model where research institutions cover the costs of their publishing output only (publications by corresponding authors affiliated with the research institution), where there are no upfront payments, no lump sums, no guaranteed amounts and no access-based cost components.

Geschuhn then demonstrated some of the weakness of the two different types of offsetting agreements. **Read and Publish** is a combination of a publishing fee (main component) and an additional reading fee, while basic **Offsetting** is where institutions receive a deduction on their licence costs depending on the APC paid in the previous year.

In Geschuhn's opinion it is essential that these two piloting models are transformed to a truly pay-as-you-publish model. This is a key prerequisite for the new business model to become sustainable. As an example Geschuhn showed results from monitoring the Springer Compact offsetting deal. The first results show that only two thirds of the journals in Springer Compact attract a significant number of publications by participating institutions. Thus, much is being paid for that is not being used (as with the big subscription deals).

Looking at the current offsetting deals, a clear strategic approach for a real switch of the business model from subscription to APC is not yet visible. Even where publishing charges set the basis for an agreement, the models still contain many elements equivalent to the subscription business such as price increases, fixed article contingents, arbitrary growth rates and guarantee amounts. However, despite the current shortcomings of offsetting, the opportunities compared to the former subscription model clearly prevail.

After her presentation the floor questioned Geschuhn's assumption that the pay-as-you-publish model would

*“Offsetting may pave the way to a large-scale Open Access transition as outlined by the OA 2020 initiative, however the deals should be based on pay-as-you-publish in order to facilitate a real switch of the business model from subscription to APC.”*

Kai Karin Geschuhn

make the market sound and functional. The questioner argued that the model would ultimately leave the author on his own negotiating with the big publishers. This, it was believed, would only improve the position of the publishers in the negotiations. Geschuhn responded that the libraries would not totally delegate the negotiations with the big publishers to the authors but would be likely to take part in some way.

### Graham Stone: Collecting APC cost data and information

Graham Stone (Jisc) began his presentation by quoting a pressing question from a new Jisc report<sup>11</sup> by Katie Shamash: Are we engaging in 'Bigger big deals'? ie are we just flipping the publisher' big subscription deals to big APC deals at the same or even higher prices and with the same opaqueness as the big subscription deals? The rapid increase in APC costs in the UK – which are monitored and also included in the OpenAPC project – and the big offsetting deals (like Springer Compact) could indicate this. Although these offsetting agreements bring some value for money, it may not be sufficient.

Stone showed examples on the discounts obtained through offsetting agreements with a number of major

### Footnotes

<sup>11</sup> Katie Shamash: “Article processing charges (APC's) and subscriptions: monitoring open access costs”, Jisc, May 2016, [jisc.ac.uk/reports/apcs-and-subscriptions](https://www.jisc.ac.uk/reports/apcs-and-subscriptions)

publishers. The numbers showed discounts of only 3-5% for the bigger publishers (Taylor & Francis, Wiley and Sage). In addition, the offsetting agreements have generated extra administration costs which should be monitored and added to the total costs of publication. Stone called for a profound reflection on cost structures and the soundness of the offsetting approach. In the UK the Green road to OA could still be (re)considered remembering that UK policies support both Green and Gold OA. Stone gave detailed insights into the UK version of the Springer Compact deal. It has generated around 3,000 OA articles, which is a significant contribution to the 128% increase of UK OA articles published in 2015. Still, it is worrying that approximately 30% of eligible authors in the Springer Compact deal have opted out of the OA option when submitting their articles in Springer journals despite very clear information about this option being free of charge. It could be due to the CC BY licence requirement but according to Springer it has mainly to do with lack of knowledge about OA among authors or authors perhaps even being suspicious about OA.

Towards the end of his presentation Stone outlined some clear priorities, which should be remembered when negotiating offsetting agreements:

1. Contribution to the transition to Open Access
2. Affordability
3. Ease of administration
4. Transparency
5. Facilitating compliance with funder policies

Stone's experience with monitoring offsetting agreements has made him aware of a number of challenges, like ineffective workflows, too much human interaction (raising related administrative costs), poor communication to authors and OA managers, challenges with cost allocation across and within institutions, and possible tensions between efficiency, transparency and cost. As a conclusion to his talk about cost monitoring he questioned the use of research funding for Hybrid OA in order to stimulate

preference for Gold OA over Hybrid OA. He also proposed greater support for Green OA in policies, as well as developing a fuller range of quality indicators through support of society publishers close to the academic community. Finally he pointed to the need to explore innovative business models like, for instance, the Open Humanities Library model.

In response to Stone's presentation a participant asked whether these models actually work, and increase usage and impact. As basic as it may seem the question is very important. In fact the five negotiation priorities that Stone presented respond well to this question in the sense that one should always stay focused on what makes a positive difference and measuring that difference.

*“30% of eligible authors in the Springer Compact deal have opted out of the OA option when submitting their articles in Springer journals despite very clear information about this option being free of charge.”*

Graham Stone

## Break out session 2: Monitoring cost data – Recommendations from the working groups

With these two keynote presentations in mind the workshop participants regathered in their groups to discuss cost data monitoring. Once again the key areas for the discussion were data, workflows, standards and policy. As in the first breakout session, the groups were expected to look at:

- ▶ Quality of data collected from available sources
- ▶ Efficiency in monitoring workflows
- ▶ Aggregation of OA publications and derived costs via standards

- ▶ How governance is being built into the actions and process of monitoring Open Access and related costs
- ▶ How policies work
- ▶ Whether there is any alignment across the policy landscape

The group work was very lively and resulted in constructive recommendations. These recommendations are summarised below.

### 1. Working group on Data

The group applied a similar approach to this discussion about collecting cost data as it had done in the previous discussion about OA publications. The discussion, therefore, was divided into the collection of data on Green, Gold and Hybrid OA respectively.

The overall recommendation was to aim for fair comparisons between types of policies and ways of obtaining e-resources. To achieve this goal the group initially focused on the cost items related to the different kinds of OA:

For Green OA:

- ▶ Costs of the development and maintenance of CRIS's and institutional repositories
- ▶ Costs of the depositing process (labour costs by supporting staff and by the academic staff)
- ▶ Costs of the maintenance of subscriptions to journals (negotiating licences, maintaining access management software and systems)

For Gold OA:

- ▶ APC expenditures
- ▶ Composition of APC costs (APC calculation of the publisher, granted rebates and discounts, VAT)

- ▶ Financial and administrative handling costs of APC's (by authors themselves or by APC funds)
- ▶ Membership deals (negotiating membership deals and the fees for membership deals)

For Hybrid OA:

- ▶ APC's (paid outside of offsetting deals, mostly by individual authors)
- ▶ Composition of APC costs (APC calculation of the publisher in contrast to Gold OA, granted rebates and discounts, VAT)
- ▶ Offsetting deals (costs per OA article, negotiating licences)

Several of the group members shared their experiences with collecting this kind of data and reported many problems in doing so. Based on these experiences and further discussions the following recommendations were made by the group:

1. Conduct talks and projects with administrative staff and publishers to get more and, ideally, standardised data regarding total costs of publication, especially APC calculation procedures and administrative handling costs
2. Create a link between the accounting system of the University and its CRIS. Such a link could be created using DOI
3. Publishers should include the DOIs of the articles in their invoices for the APC's

An obstacle to achieving the third recommendation could be that different publishers send out the invoices at different points in the process. Some do this upon acceptance of the manuscript, at which point the final

DOI may not yet be known; still others upon publication. Crossref links DOIs for preprints with DOIs for any final publications, but this additional linkage adds an extra complication. Potentially a better idea but more difficult to deploy is to encourage publishers to standardise the point in the process when an invoice will be sent.

4. Seek standardisation of the point at which publishers send the APC invoice

Further ways of collecting cost data were discussed in the group. One idea was to implement an accounting code for APC's (or publication costs). Another more large scale idea was to standardise accounting systems and thereby simplify the 'bridge function' of the DOI between the accounting system and the CRIS.

Although the group's focus, for good reason, was on how to gather cost data from accounting systems, the group's final recommendation was:

5. Financial considerations should not cloud the real value of Open Access for scientific research – even if OA costs more it's still worth it!

## 2. Working group on Workflows

Despite the initially different focus, this group ended up focusing on some of the same aspects as the Data group. As a result, the Workflow group also agreed that DOIs are crucial to include in invoices and that the timing of the APC payment should be standardised. The group highlighted that it is important to pay attention to business practices, publisher requirements and good reporting between universities/libraries and publishers, and that in fact publishers are also interested in improving workflows.

Another conclusion from the group was the need to look at the overall cost structure. What is the cost of writing an article and of peer review, including in-kind contributions and other invisible costs (eg administration

connected to journal publishing)? The group proposed looking at the eLife journal as a reference to better and more transparent publisher services, as their cost structure for Gold OA is fully transparent.

The group offered these recommendations:

1. Analyse the overall cost structure of publishing an article by dissecting it into its different parts from the writing phase to publishing
2. Use CrossMark as a possible container for OA metadata on versions and costs
3. Capture the cost and licence related metadata, making it visible in invoices: eg DOI, funder and licence information, author names and affiliation, funder identifiers and ORCIDs
4. Centralise OA funding offices regarding payments (libraries are obvious actors)
5. Investigate the complexity of many-to-many payments between universities and publishers within third party e-commerce solutions

## 3. Working group on Standards

Standards are hugely important owing to the complexity of APC payment processes. For instance it would be very useful to register the payer(s) of an APC for a multi-authored article in a standardized way. Moreover, it is not always the corresponding author who pays – sometimes the institutions split the APC. This data (along with data such as VAT, currency and 'extras' like colour page charges) should be stored in CRIS's and repositories in a standardized way across countries. During discussions it was proposed that linked data could be used to push data between accounting systems and CRIS's.

In the UK a funder driven initiative has produced a standardised template for such data but the group agreed

that this template perhaps was too elaborate and therefore would leave the dataset with too many blank cells, which would not be very useful. On the other hand the OpenAPC scheme seems more manageable to fill in. OpenAPC operates with 18 metadata fields of which only five are mandatory<sup>12</sup>. Their data gets enriched with data from sources such as Crossref, International Standard Serial Numbers (ISSN), DOAJ, Web of Science, and PubMed. Again, there was agreement that Crossref is a very useful and reliable source of information.

Based on these discussions the group came up with the following recommendations:

1. Increase communication between the people working with the data entry points. It is important that accounting staff, licensing staff, technical staff and metadata specialists discuss and agree on standards
2. Gain greater insight into the extras/"other costs". These costs ought to be part of the APC payment
3. Increase alignment among funders on their reporting requirements for institutions
4. Compare the different standard options (OpenAPC and others)
5. Deploy a standard which is easy to use and manageable, eg the OpenAPC scheme
6. Make it technically easy for the publishers to provide data to the repositories eg by using the SWORD protocol

#### 4. Working group on Policy

The policy group was particularly focused on the need for transparent models: to be clear about **who does what? how much does it cost?** and **who has to pay?** It was stated that in order to ensure a healthy publishing eco-system it is wise to maintain a division of responsibilities

between content, evaluation, and communication. If one actor holds all responsibilities the system will stagnate into an unhealthy monopolistic structure. Therefore it is very important to create innovative models that are transparent and sustainable, involving all actors. The Wellcome Open Research model was mentioned as an initiative supporting such a model.

The group proposed a number of recommendations relating to policies and funders:

1. Maintain a division of responsibilities between content, evaluation, and communication
2. The institution that has paid the APC should be informed by the author and/or publisher (eg via DOI) when the publication is published
3. Include funder information in the publication itself and in its metadata
4. Set up datasets in such a way that the data can be exchanged, with associated standards for collection and exchange
5. Policies should include clauses stating that both the DOI and cost data should be open and shareable, for instance in OpenAPC and/or MonitorUK
6. All costs related to publishing in any way should be transparent; non-disclosure agreements should be avoided. Institutions should have processes that make it clear what the total costs of publication are including APC, administrative costs, infrastructural costs, and other costs, eg page and colour charges

---

#### Footnotes

- <sup>12</sup> The OpenAPC project documentation etc. is available at <https://github.com/OpenAPC/openapc-de>
-

Finally the group discussed the impact of price caps on APC's. The APC cap can be useful to contain costs but the group speculated that the price cap could conversely motivate publishers to increase their APC's to reach it. However, in Germany where there is a price cap of €2,000 the average APC in Germany is €1,273.<sup>13</sup> This could indicate that there are mechanisms in the market that work.

### Some conclusions from the breakout session on monitoring of cost data

Accounting systems and CRIS's are central to the topic. These systems should be interoperable and aligned so that cost data at all levels can be easily retrieved. This requires the data to be open and shareable, which is an important conclusion that was arrived at by all groups. The DOI is a key tool for data transferal between systems. Therefore Crossref is expected to play a crucial role as a central hub for all the metadata that is being transmitted between publishers and CRIS's and accounting systems. It is, thus, very important that publishers are required to enter all funder data in the publication metadata as well as in the publications themselves. Such requirements should be settled in offsetting or licensing contracts with the publishers. Non-disclosure regulations in these contracts should be avoided at all times. Finally, the notion of **total costs of publication** (TCP) kept arising. It is really important to dissect the costs of publishing carefully. The APC does not necessarily cover all costs, eg administrative costs, infrastructural costs, special extra charges set by the publishers on a per publication basis and so on. Again, transparency and access to the cost data was recommended.

---

#### Footnotes

<sup>13</sup> According to data from OpenAPC ([intact-project.org/openapc](http://intact-project.org/openapc)) figures (captured 29-11-2016).

---

# Next steps – beyond Copenhagen

The concrete result of this workshop was a large number (48) of recommendations. These can be categorised and summarised and brought into further discussions and action points on monitoring of OA publications and cost data in different ways. Below is one such attempt to highlight some of those recommendations that were particularly relevant to certain crucial actors.

## 1. Current Research Information Systems (CRIS)

A key factor for successful monitoring of OA publications turned out to be the Current Research Information Systems (CRIS's). Several recommendations focused on the CRIS's:

- ▶ Improve the integration between CRIS's and institutional repositories
- ▶ Integrate different categories of Open Access in CRIS's
- ▶ Ensure that CRIS's follow the OpenAIRE interoperability guidelines
- ▶ Use CRIS's to collect data on Green and Gold OA
- ▶ Create links between accounting systems and CRIS's

## 2. Publishers

Publishers are essential actors in the monitoring process since they hold necessary information about the publication and author affiliation and about the costs. Therefore publishers were included in a number of recommendations.

- ▶ Publishers should use standardised data formats for author affiliation
- ▶ Publishers should include in Crossref a licence statement for each publication and indicate whether the publication is Green, Gold or Hybrid OA

Libraries and funders should be very specific about the kind of metadata that they require from the publishers and include these requirements in offsetting or licensing contracts. Jisc has developed a template, which could be used or further developed, available at [jisc-collections.ac.uk/Jisc-Monitor/APC-data-collection](https://jisc-collections.ac.uk/Jisc-Monitor/APC-data-collection)

- ▶ The point at which publishers send their APC invoice should be standardised
- ▶ Publishers should be encouraged to include in their invoices all cost and licence related metadata: DOI, funder and licence information, author names and affiliation, funder identifiers and ORCIDs

## 3. Libraries

It was evident from the conversations that libraries have a very important and central role to play when it comes to monitoring of publications and cost data.

- ▶ Libraries should be very specific about their requirements from publishers and use the contracts with publishers as instruments to obtain what is required. They should act in common and work in larger consortia contexts and cooperate across countries to present unified requirements to publishers (eg Efficiency and Standards for [Open Access] Article Charges or ESAC)

- ▶ Libraries should collect as much data as possible and ensure the data is open (via API) and findable, accessible, interoperable, and reusable (so-called 'FAIR' data)
- ▶ Libraries could be used as centralised OA funding offices regarding payments
- ▶ Libraries should require transparency in all costs related to everything that has to do with publishing and they should not enter into non-disclosure agreements
- ▶ Institutions should have processes that make it clear what the Total Costs of Publication (TCP) are, including APC, administrative costs, infrastructural costs and other costs like page and colour charges.

These central recommendations should be carried on from this constructive Knowledge Exchange workshop in Copenhagen as the basis for action that could improve the capabilities for monitoring OA publications and cost data and thus have an impact beyond the presentations and conversations in Copenhagen.

# Annex A: List of workshop participants

First Name	Surname	Organisation
Adrian	Price	Copenhagen University
Anette	Schneider	DTU
Angela	Holzer	DFG
Anna Mette	Morthorst	Aarhus University Library
Anna	Sigríður	Landspítali
Anne	Sandfaer	The Royal Danish Library
Arja	Tuuliniemi	Finnish National Library
Bas	Cordewener	Knowledge Exchange
Beate	Eelend	Kungliga Bibliotek, Sweden
Bernhard	Mittermaier	Forschungszentrum Jülich
Birgit	Schmidt	OpenAIRE
Britt	Ansbjerg Hansen	The Royal Danish Library
Camilla	Smith	SwePub, Kungliga Bibliotek, Sweden
Christian	Hagen Thomassen	DEFF/Knowledge Exchange
Christoph	Broschinski	Universitätsbibliothek Bielefeld
Clifford	Tatum	Leiden University
Danny	Lämmerhirt	Open Knowledge
Dirk	Pieper	Bielefeld University
Elly	Dijk	DANS
Frank	Manista	Jisc
Gernot	Deinzer	University Library of Regensburg
Gitte	Bruun Jensen	DTU
Graham	Stone	Jisc
Hannah	Hope	Wellcome
Hanne-Louise	Kirkegaard	Danish Agency for Science, Technology and Innovation,
Jean-François	Lutz	Lorraine University
Jens	Aasheim	CRISTin
Jeroen	Sondervan	Utrecht University
John	Doove	SURF

First Name	Surname	Organisation
Just	de Leeuwe	TU Delft Library
Jyrki	Ilva	University of Helsinki
Kai	Geschuhn	Max Planck Digital Library
Karen	Hytteballe	Roskilde University Library
Karin	van Grieken	SURF
Katharina	Rieck	Austrian Science Fund (FWF)
Katie	Shamash	Jisc
Lars	Bjørnshauge	SPARC Europe
Marco	Tullney	Technische Informationsbibliothek (TIB)
Marianne	Gauffriau	The Royal Danish Library & University of Copenhagen
Markus	Putnings	ENUL
Maurits	van der Graaf	Pleiade Management
Michael	Svendson	The Royal Danish Library
Mogens	Sandfær	DTU
Nina	Karlstöm	CRISTin
Pablo	de Castro	LIBER
Rachael	Lammey	CrossRef
Robert	van der Vooren	VSNU
Sandrine	Malotau	Couperin Organization
Sarah	James	Knowledge Exchange
Saskia	Woutersen	University of Amsterdam
Serge	Bauin	CNRS
Sólveig	Thorsteinsdóttir	Landspítali
Stuart	Lawson	Birkbeck
Uffe	Smed	The Royal Danish Library







Knowledge Exchange Office  
C/ O Jisc,  
One Castlepark,  
Tower Hill,  
Bristol, BS2 0JA

t: 0203 697 5804

e: [office@knowledge-exchange.info](mailto:office@knowledge-exchange.info)