Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 249 (2024) 179–185

Procedia Computer Science

www.elsevier.com/locate/procedia

16th International Conference on Current Research Information Systems (CRIS 2024)

Improving CRIS features to support new Open Access implementation workflows at institutions

Pablo de Castro[a,b]

[a] euroCRIS, Heyendaalseweg 141, 6525 AJ Nijmegen, The Netherlands
[b] University of Strathclyde, 101 St James Road, Glasgow G4 0NS, Scotland, United Kingdom

Abstract

Institutional Current Research Information Systems (CRIS) are already able to gather a wide range of metadata for multiple entities related to the research activity conducted at a university. However, as the Open Science landscape becomes ever more complex, new opportunities arise for these systems to hold even more data in the characteristic interlinked fashion of the CERIF data model. This contribution explores possible mechanisms to use institutional CRIS to tag bibliographic records for research publications to mark the application of new Open Science-related workflows. Emphasis is made on how to identify publications to which Open Access Rights Retention policies have been applied.

Keywords: current research information systems; open access; rights retention; data availability statements; open science; scholarly communications

*Corresponding author
E-mail address: pablo.de-castro@strath.ac.uk
Pablo de Castro, http://orcid.org/0000-0001-6300-1033

1. Introduction

The so-called author-pays business model for research publishing is relatively recent. Payment to publish papers Gold Open Access via Article Processing Charges (APCs) became mainstream in the 2010s. As the budgets invested in these payments grew ever larger, the need to capture the figures for actual payments somewhere among the bibliographic metadata for the publications became more pressing.

At the euroCRIS Membership Meeting in Paris in May 2015 this author delivered a presentation suggesting to enhance the data model for CRIS systems in order to capture these APC payments as part of the bibliographic information for research publications [1]. This MM2015 presentation in Paris was just one of the many inputs vendors received in this regard – this was also being simultaneously discussed within specific user groups. As a result of this

10.1016/j.procs.2024.11.062

joint input from research-performing organisations, an updated data model emerged that allowed these figures to be captured as part of the bibliographic description for a publication (see figure 1 below). This is a much more accurate way to account for the Open Access expenditure than using nominal APCs provided by journals and publishers, because it allows to reflect the actual fee that has been paid regardless of whether there may have been discounts or waivers [2].

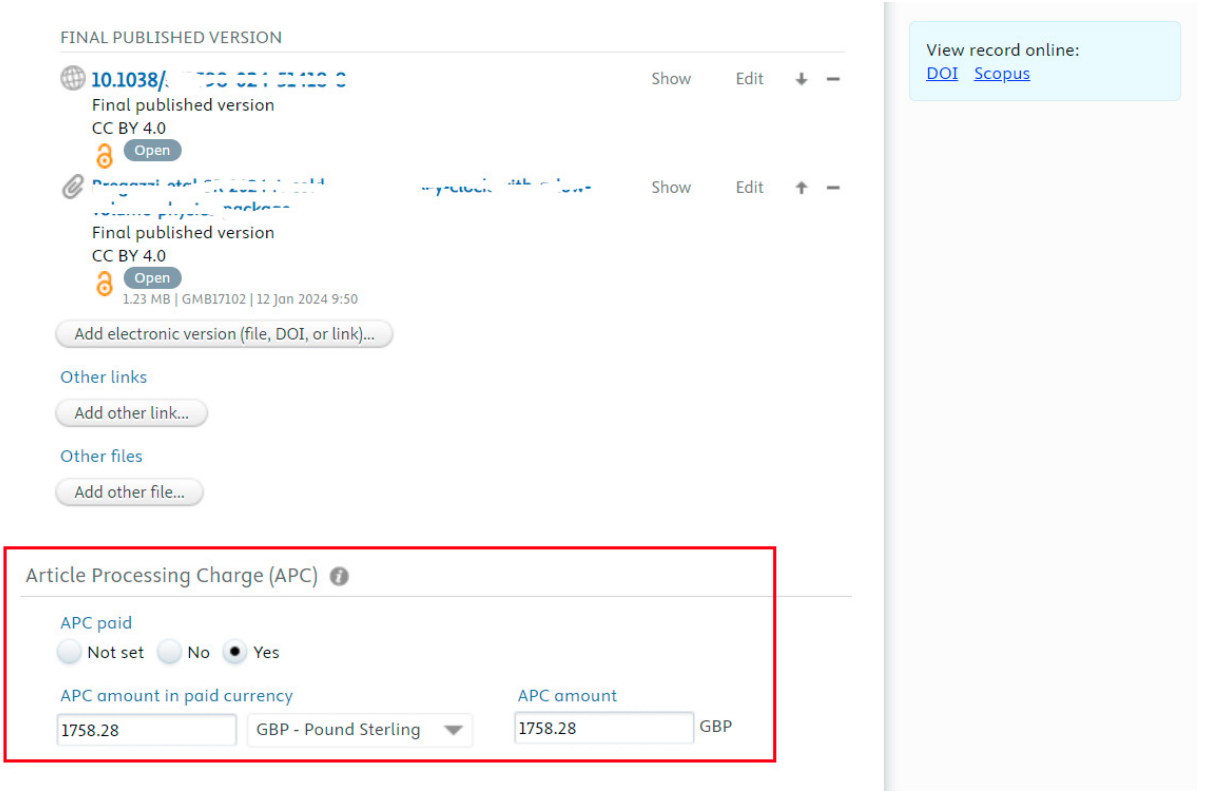


Fig. 1. APC section in the bibliographic description for a specific publication in Pure

The APC fee field that was added to the metadata set for research outputs is regularly used by many institutions these days both for internal and for external reporting. This has in turn strengthened the input that institutions and their consortia are able to deliver into collections of aggregated APC payments at a national or an international level such as OpenAPC. The analysis of these stats for payments allows the trends in Open Access publishing to be identified and acted upon from a policy perspective. It's just one example for the remarkable increase in data science-intensive workflows that have consolidated in the past few years in this domain.

The Open Access landscape has become more complex ever since APC payments started to get coded into bibliographic records for publications. We now have rights retention policies issued by research funders whose application we would like to capture without having any evident means to do so. Research outputs often include data availability statements – usually as a consequence of research funder mandates – that would also need to be captured in a more nuanced way than just adding a link to the research datasets available elsewhere. This increasing complexity is a staple of the Open Science domain, and it will only deepen as the trend towards defining alternative indicators for reformed research assessment processes consolidates. It is subsequently worth exploring mechanisms for the Open Science community to be able to use the already available e-infrastructure to capture these additional levels of research information. This can be done in collaboration with vendors – as was the case for the development around APC payments above – or by the Open Science community producing its own mechanisms and tools that may afterwards be shared across institutions and countries.

2. New Open Access workflows: Plan S and its rights retention policy

Plan S was released in Sep 2018 by an alliance of 12 international research funders called cOAlition S as a policy instrument to achieve full, immediate Open Access. Plan S proposes several coexisting routes for this objective to be achieved, including the APC payments addressed in the previous section. Given how onerous these APC payments may quickly become for research-intensive organisations and – more generally – the inequity built into the author-pays business model, the cOAlition S also introduced the so-called rights retention policies as an additional policy mechanism [4].

A rights retention approach to immediate Open Access involves the accepted author manuscript (AAM, also known as postprint) being made openly available under a Creative Commons licence upon first online release of the publication. This so-called embargo-free Green Open Access does away with the traditional requirement to apply embargo periods to these accepted manuscripts – which have typically been an average of 12 months for STEM disciplines and 18-24 months for the Social Sciences and Humanities. Critically, this brand of immediate Open Access comes at no Open Access publishing cost for the authors or their institution, thus becoming a far more equitable approach to Open Access both across the funded/unfunded researcher and the Global North/Global South divides [5].

The Plan S version of rights retention policies only applies to their funded publications (i.e. those research outputs that include in their funding acknowledgements any references to projects funded by cOAlition S-member funders) and only when no specifically approved means of making the publication Gold Open Access is available. This is enforced by the requirement to include a 2-line "rights retention statement" in such funding acknowledgements on the submitted manuscript whereby the author retains the right to make a copy openly available under a CC licence of any accepted manuscript arising from the submission.

These policies introduce a significant amount of additional complexity into the workflows for authors and especially for institutional Open Access support services in charge of their application. However, such institutional services are well aware of the advantages ushered by this approach as a counterbalance to the very expensive alternative (Gold Open Access) routes. Moreover, certain institutions piggybacked on the cOAlition S-member funder rights retention policies to pass their institutional rights retention policies (IRRPs) where the applicability of this rights retention policy was extended to *all publications produced at the university*. There had been previous adoptions of institutional rights retention policies, notably by Harvard University in 2008, but it was the University of Edinburgh where the first 'expansion' of the Plan S-inspired rights retention policy was adopted by means of its Research Publications & Copyright Policy (2021) which came into force as of 1 Jan 2022 [6].

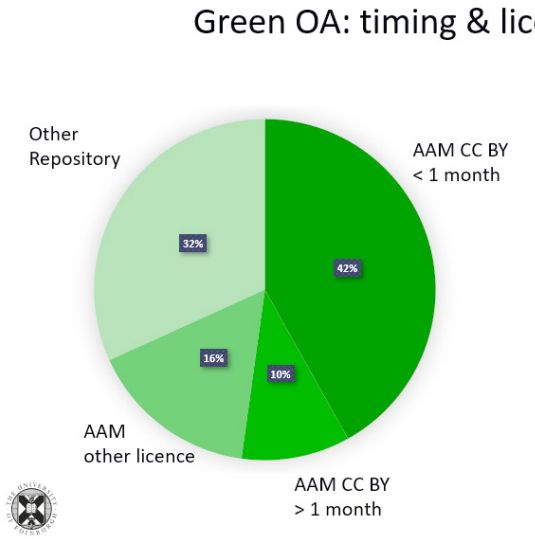
At the time of writing, close to thirty universities in the United Kingdom have now passed their own IRRP, most of them research-intensive ones [7]. Critically, the institutional rights retention policy does no longer require authors to include the 2-line rights retention statement in the funding acknowledgements section of their manuscripts. This makes its implementation notably simpler – if not simple at all – and relies instead on institutions notifying publishers about the passing of their IRRPs as a blanket enabler for their application.

However, this removal of the need to include 2-line rights retention statements introduces in turn issues around the monitoring of the implementation of these policies. The standard approach to *externally* monitoring the uptake of rights retention policies issued by cOAlition S-member funders involves the text-mining of full-text accepted manuscripts available in Open Access repositories to identify instances of rights retention statements in their funding acknowledgements sections [8]. This was no silver bullet in the first place, since many articles carrying rights retention statements – including the one marked as reference [5] in this paper, providing more detail on how rights retention policies work and what they aim to achieve – were published Gold Open Access anyway. If these 2-line rights retention statements are no longer present, the external monitoring becomes even more difficult: the sole strategy to accurately identify instances of publications made Open Access under a rights retention approach in content aggregations such as the EuropePMC or CORE databases is to search for full-text accepted manuscripts which: (i) are available under no embargo period, (ii) carry a Creative Commons licence and (iii) are not available Gold Open Access elsewhere.

This is of course not impossible, but it requires a level of granularity in the metadata that is mostly not available at present. The much simpler alternative is to resort to *internal* monitoring of rights retention instances at institutions. And this is where institutional CRIS come into play.

Institutional CRIS like the University of Edinburgh's are able to provide a comprehensive snapshot of the publications whose full-text accepted manuscript was made openly available embargo-free – including very granular

details such as how long it took for the manuscript to be deposited since it was accepted and what kind of Creative Commons (CC) licence the AAM was published under. This analysis is shown on figure 2 below for the first year of application (2022) of the University of Edinburgh IRRP. The data shown on the figure is critical to estimate the uptake of an institutional rights retention policy and is likely to be very relevant too for the next iteration of the Research Excellence Framework (REF) Open Access policy in the United Kingdom.



Type	#	%
AAM CC BY < 1 month	468	42
AAM CC BY > 1 month	115	10
AAM other licence	179	16
Other Repository	355	32
Total Green OA	1117	100

Fig. 2. Statistics for rights retention instances at the University of Edinburgh 12 months into their IRRP [9]

Data availability statements (DAS) represent another domain where some sort of monitoring may be required. The need for these typically emerges from funder mandates – the Wellcome Trust, the European Commission and the UK Research and Innovation all have such policies in place – but these requirements can be met in various different ways [9]. Some mechanism to identify which publications carry a DAS and what kind of data availability they offer is also needed. This means that the quest for additional research information exceeds the realm of Open Access and rather moves under the wider Open Science umbrella.

3. Capturing these bespoke developments in institutional CRIS

Widespread discussions have been taking place for quite some time within user groups for different CRIS solutions and within the wider scholarly communications community on how to best capture these additional elements of research information – instances of rights retention and data accessibility statements – for internal and external reporting purposes. Same as the discussion on the coding of APC payments in bibliographic records for publications, the discussion on these other, newer elements is also happening at various levels. Some institutions would rather not wait for the vendors – whose to-do list is overcrowded, even more so with a new REF exercise on the horizon for the UK – to come up with a specific upgrade to their systems that allows this information to be captured. One such institution is the University of Manchester, who developed their Open Access Compliance Platform (OACP) as an open source add-on to their Pure institutional CRIS and presented it at the Pure International Conference 2022 in Portugal [10]. This OACP includes a series of boxes that can be ticked to indicate that a specific publication has followed the embargo-free Green Open Access route (i.e. the rights retention route) to immediate Open Access or that a specific manuscript includes a data availability statement or DAS.

The advantage of developing this in-house is that the functionality can be tailored to the specific institutional needs. Figure 3 below shows a screenshot of the University of Manchester OACP v2, which includes additional features such as a box to mark a REF exception. The REF Open Access policy entails a mandate to deposit full-text accepted

manuscripts in institutional systems within a given period of time – but it also defines a range of exceptions where this mandate may not be met without disqualifying the publication for the research assessment exercise. Because these exceptions typically evolve over REF exercises, it's very useful if the ability to update a list of applicable exceptions lies with the institution rather than with the vendor. More so if the development is based on open source code and the institution – as it is the case for Manchester University – states its willingness to make the code available to other institutions interested in implementing a similar platform.

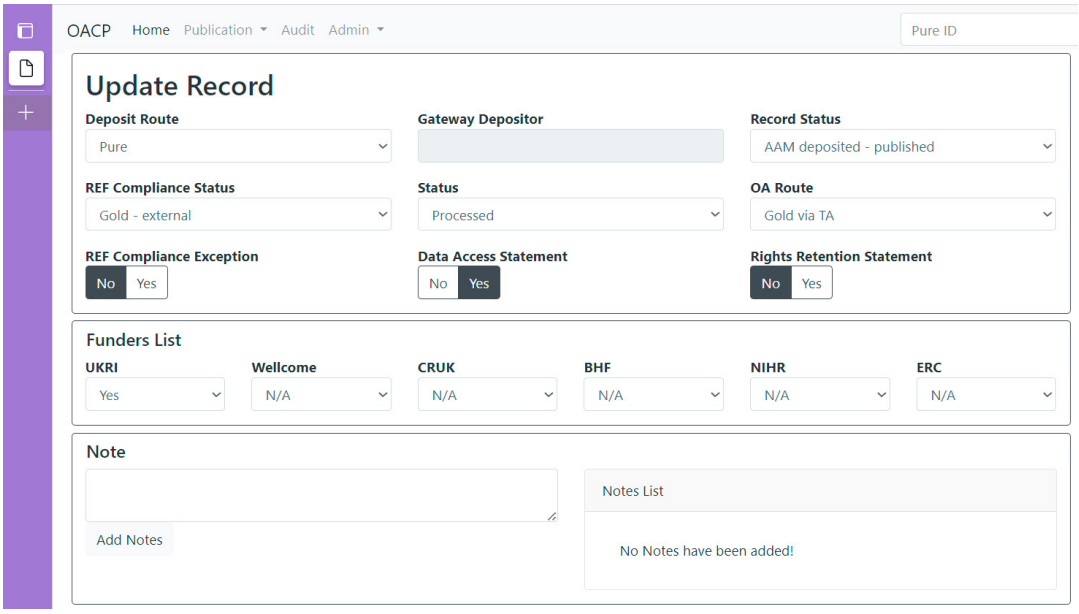


Fig 3. University of Manchester Open Access Compliance Platform (OACP)

There are plenty of universities however who lack the IT resources to implement their own platform, even if it were available as an open source-based development for them to reuse. These other institutions may have the opportunity to implement a much lighter mechanism to identify rights retention instances or data availability statements without necessarily having to embark in any in-house development. Institutional CRIS systems have reached a degree of autonomy that allows a relatively simple feature to be added without necessarily relying on the vendor for the purpose.

3.1. The Library Keywords feature in Pure

One such low-hanging-fruit feature is the Library Keywords functionality in Pure. As its name suggests, this is primarily aimed at allowing the library team responsible for the creation of bibliographic records for institutional research publications in the CRIS to assign specific library keywords to a research output on top of the author keywords that feature on the manuscript. However, it is not difficult to expand this tagging mechanism to allow an additional range of tags to be used to characterise specific records.

Figure 4 below shows how this approach has been implemented in the University of Strathclyde Pure-based institutional CRIS in a way that allows bibliographic records to be tagged both for data availability statements and rights retention policy application. This tagging is as simple as it is valuable for monitoring purposes: while the tags are not displayed on the bibliographic information snapshot provided for the publication on the institutional research portal or in the repository, the feature allows the filtering of records that carry a specific tag. The list of publications can also be downloaded for analysis – for instance, to get stats by research department, funder or publisher.

As shown in the figure below, a careful design of the categories allows different tags to be available for cOAlition S-member funder rights retention (these tend to be applied to Wellcome- or UKRI-funded publications for which no Gold Open Access route was readily available) and for the Strathclyde institutional rights retention policy or IRRP. There is also a third category ("I") used to tag records for which the application of rights retention remains uncertain

at the time of checking ("I" stands for "Indeterminate"). Any record tagged "rights retention—I" will eventually become "IRRP-G" (or an instance for the general institutional rights retention policy) or "RRS-F", i.e. an instance for the application of the funders' rights retention strategy. If the research output is eventually published Gold Open Access, then no rights retention tag will remain.

This internal Library Keywords feature also includes a mechanism to tag publication records in a number of different categories with regard to their data availability statement. Different keywords have again been created to mark the different categories a publications may fall under. These are:

- DAS – Data availability statement present; it also includes external link to data
- DP – All data present and/or included in research publication
- DR – Data restricted
- DUR – Data available upon request, or variant of this statement
- I – Indeterminate: article cannot be accessed and it is unknown whether a statement is present or not
- NDAS – No data availability statement present

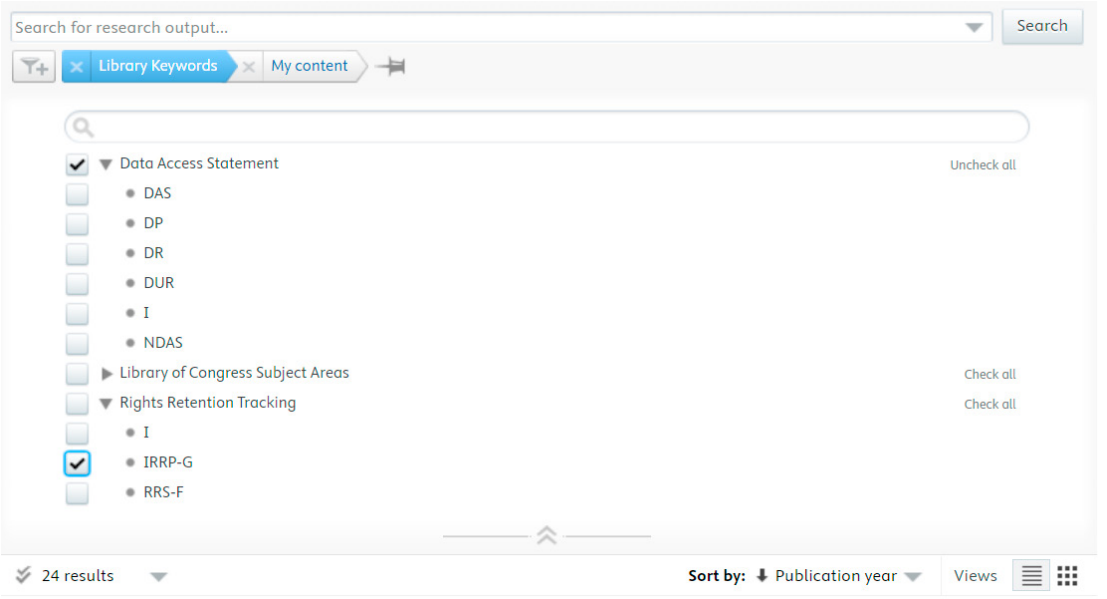


Fig 4. Use of the Library Keywords feature at the Pure institutional CRIS at the University of Strathclyde

While this simple development allows institutions to easily monitor and analyse the uptake of rights retention policies, the issue remains of how to collectively monitor this uptake for a whole range of institutions. Because institutional rights retention policies are largely in their infancy yet, the first steps for their monitoring involve the implementation of mechanisms to internally check their uptake at an institutional level. Some other mechanism should be found for the information to be exported and aggregated across institutions – the easiest way forward in this regard would be a harmonised use of an agreed metadata element in Dublin Core metadata schemas for the bibliographic description of research publications in repositories such as RIOXX or OpenAIRE.

4. Conclusions

CRIS systems are very powerful instruments for data collection and may offer – when properly used – very valuable information to the institutions using them. In cases like the application of rights retention policies where it's very difficult to capture such instances from the outside, CRIS systems may provide the most suitable way forward for the stats to be collected, analysed and reported. This may also be the case for other features of the internal Open Science implementation workflows such as data availability statements.

Figuring out mechanisms to collect the information in these systems – particularly if able to avoid the need to rely on the vendor to implement them – is an important aspect in the operationalisation of policy instruments such as rights retention. It may be that CRIS systems eventually see a similar development in this domain to the one used to capture APC payments following the identification of this need about a decade ago, but in the meantime it's useful to have simple alternatives in place. This is particularly relevant at a time when discussions on the reform of research assessment are progressing and there are early reflections already on the possible relevance of CRIS systems to eventually capture the emerging indicators that will allow a more rounded evaluation of research activity [11].

There's still a need for some sort of cross-institutional interoperability in the workflows to monitor the uptake of rights retention policies. This should address the identification of a publication where one of the coauthoring institutions has already applied rights retention, as well as the aggregation of institutional rights retention instances so that it's possible to assess the impact of this policy area across the sector.

Acknowledgements

The support from the University of Strathclyde repository manager Dr Leigh Stork – who has actually implemented the Pure features shown on the screenshots – is gratefully acknowledged. Parallel conversations with Pure managers at institutions in The Netherlands were also very valuable.

References

[1] De Castro, P. (2015). Challenges in Coding Information on Article Processing Charges (APCs) into Institutional Systems. Spring 2015 euroCRIS Membership Meeting (Paris, 11-12 May 2015). http://hdl.handle.net/11366/370

[2] Pallares, C., Vélez Cuartas, G., Uribe-Tirado, A., Restrepo, D., Ochoa, J., and Suárez, M. (2022). Situación del acceso abierto y los pagos por APC en Colombia. Un modelo de análisis aplicable a Latinoamérica. Revista Española De Documentación Científica, 45(4), e342. https://doi.org/10.3989/redc.2022.4.1931

[3] Plan S (Wikipedia), https://en.wikipedia.org/wiki/Plan_S (accessed 7 July 2023)

[4] cOAlition S (2020). "Plan S: Rights Retention Strategy" [factsheet]. https://www.coalition-s.org/wp-content/uploads/2020/07/RightsRetentionStrategy.pdf

[5] Moore, S. (2023). "The Politics of Rights Retention". Publications 11(2): 28. https://doi.org/10.3390/publications11020028

[6] The University of Edinburgh Research Publications & Copyright Policy (2021). https://www.ed.ac.uk/files/atoms/files/research_publications_copyright_policy_2021_approved.pdf (accessed 7 July 2023)

[7] Englen, S. (2024). List of UK universities that have passed an institutional rights retention policy. Github. https://github.com/sje30/rrs/blob/main/ukinstitutions.csv (accessed 7 July 2024)

[8] Mounce, R. (2021). Research outputs including rights retention language. Github. https://github.com/rossmounce/rrs-language-including-outputs/blob/main/RRS-outputs-asof-2021-09-01.csv.

[9] Wellcome Trust (2021). Data availability statements – some examples of good practice. https://cms.wellcome.org/sites/default/files/2021-01/Data-availability-statement-guidance-2021-01.pdf

[10] Ganeshwaran N (2022). The University of Manchester Open Access Compliance Platform (OACP): An Extensible Tool for Managing & Monitoring the Submission & Compliance of Research Outputs. Pure International Conference 2022 (Vila Nova de Gaia, Porto, 2-3 Nov 2022). https://www.youtube.com/watch?v=DMXYx4FvjEI

[11] Schöpfel, J., Azeroual, O. (2022). What Does DORA Mean for CRIS?. euroCRIS Strategic Membership Meeting 2022 – Autumn (Nijmegen). http://hdl.handle.net/11366/2248