

Creació d'un indicador d'accés obert a la producció científica de Catalunya

Informe realitzat per encàrrec del Consorci de Serveis Universitaris de Catalunya (CSUC). Àrea de Ciència Oberta.

Ángel Borrego
Universitat de Barcelona
Facultat d'Informació i Mitjans Audiovisuals
borrego@ub.edu

Juny de 2021

Sumari

Resum	2
1. Introducció	2
2. Accés obert a la bibliografia científica: evolució i situació actual	5
3. Eines per determinar la disponibilitat en accés obert.....	8
3.1 <i>Motors de cerca</i>	8
3.2 <i>Unpaywall</i>	10
4. Indicadors nacionals d'accés obert	12
4.1 <i>Iniciatives internacionals</i>	13
4.2 <i>Alemanya</i>	14
4.3 <i>Dinamarca</i>	16
4.4 <i>França</i>	17
4.5 <i>Països Baixos</i>	17
4.6 <i>Catalunya</i>	18
5. Conclusions	20
5.1 <i>Col·lecció documental per al càlcul de l'indicador</i>	21
5.2 <i>Eina per identificar la producció en accés obert</i>	22
5.3 <i>Definició i categories d'accés obert</i>	22
5.4 <i>Fonts addicionals d'informació</i>	23
5.5 <i>Cronologia</i>	23
5.6 <i>Reproductibilitat</i>	23
6. Recomanacions finals	24
Referències	25

Resum

El *Pacte Nacional per a la Societat del Coneixement* de la Generalitat de Catalunya preveu el càlcul d'un indicador d'accés obert a la producció científica de Catalunya. Es tracta d'una mesura factible, que no requereix de grans recursos i està en sintonia amb múltiples iniciatives al si de la Unió Europea per avaluar l'acompliment de les estratègies de ciència oberta. Un repàs de la bibliografia publicada en l'última dècada sobre l'evolució de l'accés obert a la bibliografia científica i l'anàlisi d'alguns indicadors d'accés obert desenvolupats arreu posa de manifest la necessitat de prendre un seguit de decisions metodològiques sobre la col·lecció documental sobre la qual calcular l'indicador, l'eina utilitzada per determinar la disponibilitat en accés obert, la definició d'accés obert emprada i el procediment de recollida de dades. En aquest informe, s'aconsella consensuar aquestes decisions entre tots els agents implicats i es recomana utilitzar el Portal de la Recerca de Catalunya com a font per identificar la producció, emprar Unpaywall per determinar la disponibilitat en accés obert, fer servir les seves categories d'accés obert i garantir la reproductibilitat de l'indicador publicant la informació metodològica i les dades que permetin replicar el seu càlcul.

1. Introducció

El *Pacte Nacional per a la Societat del Coneixement* (Generalitat de Catalunya, 2020) dedica el capítol 4 a l'impuls del sistema de recerca i les seves infraestructures. El novè acord en aquest àmbit és «Fomentar una estratègia catalana de ciència oberta» (p. 44) que abasti, entre d'altres, l'accés obert a les publicacions científiques. Per tal de verificar el grau de compliment de l'acord, l'annex A recull, entre altres indicadors de seguiment, el càlcul del «Percentatge de publicacions en accés obert respecte al total de publicacions» (p. 67). Així mateix, dins de l'àmbit de l'educació superior, s'inclou el càlcul del «Nombre de publicacions en accés obert de les universitats públiques catalanes» (p. 72), el «Percentatge de publicacions de les universitats públiques catalanes en accés obert respecte al total de publicacions» i el «Nombre de publicacions en accés obert consultables des del Portal de la Recerca de Catalunya» (p. 73).

L'accés obert és un conjunt de principis i pràctiques que han capgirat el sistema de comunicació científica durant les dues últimes dècades. En resposta als constants increments en els preus de subscripció de les revistes científiques —que creixen per sobre del cost de la vida—, el moviment per l'accés obert defensa que les publicacions científiques s'han de distribuir en línia, sense barreres econòmiques ni limitacions en els drets de còpia o reutilització.

En essència, hi ha dues estratègies per facilitar l'accés obert a la bibliografia científica: el dipòsit dels articles en repositoris institucionals o temàtics (*green open access*) i la publicació en revistes d'accés obert (*gold open access*). Dins d'aquestes últimes, s'acostumen a distingir les revistes que, per ésser viables, cobren als autors unes taxes d'edició (*article processing charges*, APC) de les finançades per institucions i societats científiques (conegudes com revistes *diamond* o *platinum*). D'altra banda, les revistes híbrides (*hybrid open access*) són revistes de subscripció que permeten als autors que ho desitgen publicar en accés obert els seus articles —mentre la resta de continguts continua sent d'accés restringit— previ pagament d'unes APC. Finalment, l'accés obert bronze (*bronze open access*) fa referència a articles que es poden consultar gratuïtament al web de l'editorial però que no disposen d'una llicència que garanteixi el reaprofitament dels continguts, de manera que tampoc no està clar si aquests articles estan disponibles en obert per a la seva lectura de forma temporal o permanent. En conjunt, aquest model de lliure distribució de la informació científica en obert garanteix l'accés a la bibliografia per part dels investigadors —tant de l'acadèmia com de la indústria— i al gran públic, accelerant l'empresa científica.

El mesurament de la disponibilitat en accés obert de les publicacions ha atret l'interès dels diversos agents implicats en la recerca científica. Durant l'última dècada, s'han dut a terme diversos estudis per mesurar l'evolució del percentatge de publicacions disponibles en accés obert. Al mateix temps, algunes institucions de recerca publiquen indicadors d'accés obert a la seva producció per demostrar el seu alineament amb les polítiques científiques d'obertura del coneixement i determinar el grau de compliment de les fites que s'han marcat.

La determinació del volum de publicacions en accés obert és, possiblement, l'indicador més pragmàtic per mesurar la concreció de les polítiques de ciència oberta i està present en la majoria de plans en aquest àmbit que habitualment es marquen com a objectiu arribar a un 100% de publicacions en accés obert en un determinat horitzó temporal. Per exemple, les edicions anuals del *Baromètre français de la Science Ouverte*¹ se centren de forma gairebé exclusiva en l'anàlisi de la disponibilitat en accés obert de les publicacions. Igualment, l'única dada quantitativa inclosa a la introducció del segon *Plan national pour la science ouverte*² fa referència a l'increment en el volum de publicacions en accés obert.

¹ L'última edició, publicada el gener de 2021 amb dades de 2020, està disponible a [https://cache.media.enseignementsup-](https://cache.media.enseignementsup-recherche.gouv.fr/file/2021/57/9/NF2021_01_Barometre_Science_Ouverte_1370579.pdf)

[recherche.gouv.fr/file/2021/57/9/NF2021_01_Barometre_Science_Ouverte_1370579.pdf](https://cache.media.enseignementsup-recherche.gouv.fr/file/2021/57/9/NF2021_01_Barometre_Science_Ouverte_1370579.pdf)

² *Deuxième Plan national pour la science ouverte: généraliser la science ouverte en France 2021-2024.*

[https://cache.media.enseignementsup-](https://cache.media.enseignementsup-recherche.gouv.fr/file/science_ouverte/20/9/MEN_brochure_PNSO_web_1415209.pdf)

[recherche.gouv.fr/file/science_ouverte/20/9/MEN_brochure_PNSO_web_1415209.pdf](https://cache.media.enseignementsup-recherche.gouv.fr/file/science_ouverte/20/9/MEN_brochure_PNSO_web_1415209.pdf)

Alguns rànquings universitaris, com el CWTS Leiden Ranking³, han introduït indicadors d'accés obert en paral·lel als de producció, impacte o col·laboració per classificar ordenadament els centres d'ensenyament superior. Com que l'accés obert varia entre disciplines, l'especialització de les universitats pot donar lloc a inconsistències en la classificació. Per aquesta raó hi ha hagut intents de normalització disciplinària que permetin una comparació més objectiva entre centres (Maddi, 2020; Maddi, Lardreau i Sapinho, 2021).

D'entrada, sembla que mesurar la disponibilitat en accés obert d'un conjunt de publicacions ha de ser un procés relativament senzill. No sempre ho és. Alguns treballs han posat de manifest les discrepàncies entre les diverses fonts susceptibles de ser consultades. Recentment, el National Open Research Forum (NORF, 2021), creat a Irlanda el 2017 per orientar l'agenda nacional envers la ciència oberta, alertava de la manca de dades per mesurar la disponibilitat de publicacions en accés obert:

«We have no authoritative data source on the level of OA nationally. Particularly problematic is the measurement of OA in the humanities and social sciences. [...] Better data and much more analysis is needed to address progress against this and other NORF recommendations and to be able to project the trajectory of OA growth in order to set an appropriate target for 100% OA across all disciplines.»

En comparar els indicadors d'accés obert subministrats per diferents fonts, és fàcil sorprendre's per la disparitat de xifres aportades. El juny de 2020, dues bibliotecàries de la Universitat de Leiden (Woutersen-Windhouwer i Hettne, 2020) constataren que l'última edició del rànquing CWTS Leiden indicava que el 70,5% de les publicacions de la seva institució estaven en accés obert, una xifra sensiblement superior al 60% que preveïen aconseguir a l'indicador d'accés obert de l'associació de universitats neerlandeses (VSNU) que s'havia de publicar en aquelles mateixes dates.

Discrepàncies similars s'observen al nostre entorn. Per exemple, segons l'Observatori de l'Accés Obert⁴, hostejat a la Universitat Politècnica de Catalunya, el 75,2% de la producció de la Universitat Pompeu Fabra en 2018 estaria en accés obert. Per contra, el Portal de la Recerca de Catalunya⁵ indica que només el 30% de la producció de la mateixa Universitat el mateix any estaria en accés obert.

³ <https://www.leidenranking.com>

⁴ <https://biblioteca.upc.edu/observatori>

⁵ <https://portalrecerca.csuc.cat/estadistiques/publicacions/accesobert>

Aquestes inconsistències són atribuïbles a diversos factors:

- a) la col·lecció documental sobre la qual es realitza el càlcul;
- b) l'eina utilitzada per determinar la disponibilitat en accés obert;
- c) la definició d'accés obert emprada; i
- d) el moment de recollida de les dades.

El present informe descriu la metodologia utilitzada per realitzar estudis i construir indicadors que mesuren la disponibilitat en accés obert de publicacions científiques. Es pretén posar de manifest com les diferències en els mètodes emprats poden donar lloc a discrepàncies en els resultats obtinguts. A partir d'aquesta anàlisi es faran algunes recomanacions sobre la metodologia per construir un indicador d'accés obert a la producció científica de Catalunya.

L'informe s'estructura en quatre parts. L'apartat 2 revisa la bibliografia publicada en l'última dècada sobre l'evolució de l'accés obert a la bibliografia científica. L'anàlisi se centra en els estudis a gran escala que mostren les principals tendències en l'evolució de l'accés obert i les perspectives de futur. L'apartat 3 aprofundeix en la descripció de les eines emprades en els estudis identificats a l'apartat anterior: motors de cerca de bibliografia científica i Unpaywall⁶, principal instrument utilitzat en treballs d'aquest tipus. L'apartat 4 analitza les característiques d'un seguit d'indicadors d'accés obert desenvolupats a la Unió Europea, Alemanya, Dinamarca, França, Països Baixos i Catalunya. L'apartat 5 sintetitza els principals resultats obtinguts i presenta les conclusions per a la creació d'un indicador que mesuri la disponibilitat en accés obert de la producció científica de Catalunya. L'informe finalitza amb un seguit de recomanacions.

2. Accés obert a la bibliografia científica: evolució i situació actual

Durant l'última dècada, diversos estudis han mesurat el grau de disponibilitat en accés obert de la bibliografia científica publicada arreu. En aquest apartat revisarem la metodologia i les conclusions d'alguns treballs que adopten una perspectiva global, obviant estudis limitats a una disciplina, país o tipologia d'accés obert.

La primera anàlisi a gran escala va ser duta a terme per Archambault et al. (2014) que, en el marc d'un contracte amb la Unió Europea, van crear un motor de cerca per rastrejar la disponibilitat gratuïta d'una mostra de documents publicats entre 1996 i 2013 indexats a Scopus. Els autors van mesurar el que Suber (2008) va batejar com a *gratis open access* —la possibilitat de llegir els articles—, per

⁶ Unpaywall (<https://unpaywall.org>) indexa continguts publicats a revistes d'accés obert i disponibles en repositoris per oferir versions en accés obert de documents científics. A l'apartat 3 es descriuen amb més detall les seves prestacions.

contraposició al *libre open access* —que inclou la possibilitat de reutilitzar-los. Prenent en consideració tot tipus de disponibilitat —és a dir, incloent-hi còpies d'articles en xarxes socials o en pàgines web personals—van trobar un nivell global d'accés obert del 47%.

Piwowar et al. (2018) van utilitzar Unpaywall per mesurar la disponibilitat en accés obert de tres mostres de 100.000 articles —articles amb DOI⁷ introduïts a Crossref⁸; articles «recents» indexats a Web of Science; i articles visualitzats per usuaris d'Unpaywall— i van estimar que almenys el 28% de la bibliografia científica estava disponible en accés obert, si bé aquesta xifra s'elevava fins al 45% per als articles publicats en 2015, últim any analitzat. En aquest treball s'introduïa el concepte d'accés obert *bronze* per fer referència als articles «made free-to-read on the publisher website, without an explicitly open license.»

Poc després, Bosman i Kramer (2018) van fer servir les dades subministrades per Unpaywall a través de Web of Science —que, a l'igual que Dimensions⁹ o Scopus, incorpora un filtre de documents en accés obert entre els resultats basat en Unpaywall (Else, 2018a)— per analitzar la disponibilitat de 12,3 milions d'articles i revisions publicats entre 2010 i 2017. La disponibilitat en obert arribava a gairebé el 30% dels documents publicats en 2015 i 2016, si bé les xifres de 2017, últim any analitzat, es reduïen. Les raons d'aquesta caiguda per al període més recent tenien probablement a veure amb el fet que les editorials que posen continguts en obert per la via bronze ho fan almenys un o dos anys després de la publicació, un fenomen similar al que es pot produir amb el dipòsit en repositoris institucionals i temàtics.

Un altre estudi de Science-Metrix (2018) va mesurar la disponibilitat en accés obert de documents indexats a Web of Science i Scopus fent servir la base de dades 1science¹⁰. L'equip de treball era el de l'estudi d'Archambault et al. de 2014 i va fer servir la mateixa definició de *gratis open access*, és a dir, la disponibilitat d'una còpia gratuïta d'un article científic en el web que es pugui consultar sense necessitat d'autenticació, independentment de l'existència o no d'una llicència de reutilització. Els resultats mostraven que, en el tercer trimestre de 2016, el 50% dels articles

⁷ Un DOI (*digital object identifier*) és un enllaç en forma de codi alfanumèric que identifica de manera unívoca un contingut digital (un article de revista o una monografia, per exemple). Un DOI és un identificador persistent que roman associat a l'objecte independentment que canviï l'adreça web on es troba allotjat. Més informació: https://en.wikipedia.org/wiki/Digital_object_identifier.

⁸ Crossref (<https://www.crossref.org>) és una agència d'assignació de DOIs. En ella estan integrades editorials que comparteixen les metadades dels seus articles a través d'una base de dades de consulta pública que permet la interrelació dels articles a través de les citacions.

⁹ Dimensions (<https://www.dimensions.ai>) és un cercador d'informació científica de l'empresa Digital Science

¹⁰ 1science és una companyia canadenca que, inicialment, se centrava en la localització de bibliografia en accés obert. Va ser creada per Eric Archambault, responsable de l'estudi ja citat realitzat per a la Unió Europea l'any 2014. L'any 2018 la companyia va ser adquirida per Elsevier i, actualment, ofereix un cercador de bibliografia científica que es pot consultar des de: <https://1findr.1science.com>.

publicats en 2006 indexats a Web of Science estaven disponibles en accés obert. El percentatge s'incrementava fins al 60% per documents publicats en 2011 i, a partir de llavors, dequia. Els resultats mostraven diferències per països i disciplines i, dins d'aquestes, en les modalitats d'accés obert més habituals.

Fent servir una aproximació diferent, Martín-Martín et al. (2018) van emprar Google Scholar per localitzar «còpies gratuïtes» de gairebé 2,3 milions d'articles i revisions amb DOI indexats entre 2009 i 2014 a Web of Science. Les dades es van recollir l'estiu de 2016 i mostraven una taxa de disponibilitat del 54,6%. Els documents estaven disponibles en els webs de les editorials (23,1% que combinava accés obert daurant, híbrid i bronze), en repositoris (17,6%) i en altres fonts com ResearchGate¹¹ (40,6%), amb solapaments entre les diverses fonts.

El mateix any 2018, Severin et al. van dur a terme una metanàlisi que sintetitzava els resultats d'onze estudis sobre la disponibilitat de la bibliografia científica en accés obert. Els resultats posaven de manifest la diversitat de xifres obtingudes en aquests treballs. L'article conclouïa (p. 17) que «estimated OA levels for publication years after 2010 varied between 29.4% [la xifra oferta a l'estudi de Bosman i Kramer per a l'any 2016] and 66% [una xifra aportada per Larivière i Sugimoto (2018) a un treball sobre la disponibilitat en accés obert de les publicacions finançades entre 2009 i 2017 per dotze agències amb mandats de publicació en accés obert], with most studies reporting OA levels to lie somewhere between 50% and 60%.» Aquestes discrepàncies obeïen al fet que alguns dels estudis que mostraven les taxes més elevades de disponibilitat en accés obert incloïen còpies disponibles en xarxes socials i webs personals sota la categoria d'accés obert verd.

Un estudi posterior (Piwowar, Priem i Orr, 2019) va combinar dades sobre la disponibilitat en accés obert de 70 milions de documents publicats entre 1950 i 2019 amb dades de 2,8 milions de consultes de l'extensió de navegador d'Unpaywall durant el mes de juliol de 2019. Enlloc de mesurar la disponibilitat en accés obert dels documents en funció de la seva data de publicació —el que fa la major part de la bibliografia— els autors d'aquest treball optaven per intentar determinar la disponibilitat en accés obert de la bibliografia en una data d'observació. Els resultats mostraven que, en 2019, el 31% de la bibliografia científica analitzada estava disponible en accés obert. Aquests articles atreïen el 52% de les visualitzacions. A partir d'una anàlisi de tendències, els autors estimaven —en un model que qualificaven de conservador— que, l'any 2025, el 44% dels articles estarà disponible en accés obert, acaparant el 70% de les visualitzacions.

Finalment, Robinson-Garcia et al. (2020) van analitzar la disponibilitat en accés obert de 4,6 milions de documents indexats a Web of Science procedents de 963

¹¹ ResearchGate (<https://www.researchgate.net>) és una xarxa social per investigadors on, entre d'altres funcionalitats, es poden dipositar articles.

universitats. Fent servir Unpaywall, van observar que el 77% dels documents estaven disponibles en accés obert verd, el 33% en daurat, el 20% en bronze i el 16% en híbrid, tot i que amb un elevat grau de solapament entre el primer model i els altres tres.

En conclusió, tal com anunciava la revisió sistemàtica de Severin et al. (2018), s'observa una elevada disparitat en les xifres de prevalença de l'accés obert ofertes per diferents estudis. Aquestes oscil·lacions poden atribuir-se a diferents factors. En primer lloc, pel que fa a les col·leccions documentals analitzades, s'observa una elevada coincidència en la utilització de Scopus i Web of Science per identificar la producció. No obstant, sovint hi ha diferències en la finestra temporal entre la data de publicació dels documents i el moment de recollida de les dades sobre disponibilitat. La major part d'estudis inclouen una anàlisi longitudinal per determinar el grau de consolidació de l'accés obert. Malgrat que la tendència és a l'increment progressiu de la disponibilitat, per períodes molt recents el percentatge de documents en accés obert es redueix. La raó és que hi ha articles que es publiquen en tancat però que, en superar un període d'embargament, són alliberats per les editorials o dipositats en repositoris. Un altre element que varia d'un estudi a un altre és l'eina utilitzada per determinar la disponibilitat en accés obert. En línies generals, s'observen dues estratègies que descriurem amb més detall en el proper apartat: utilitzar un motor de cerca per localitzar còpies dels articles o fer servir Unpaywall. Aquesta tria incideix sobre un tercer element: la definició d'accés obert emprada en els diversos estudis. Els motors de cerca localitzen còpies gratuïtes disponibles en pàgines web personals o xarxes socials. En ocasions, els autors que difonen aquestes còpies no disposen dels drets per fer públics aquests documents, que no inclouen una llicència de reutilització, per la qual cosa no compleixen tots els requisits d'una definició més estricta d'accés obert com la que pot fer servir Unpaywall, que no indexa còpies en pàgines personals o xarxes socials.

3. Eines per determinar la disponibilitat en accés obert

Com hem vist a l'apartat anterior, la major part de treballs que mesuren la prevalença de la bibliografia científica en accés obert fa servir o bé un motor de cerca de bibliografia científica o bé Unpaywall. Descriurem ara amb més detall aquestes dues estratègies.

3.1 Motors de cerca

En el cas dels motors de cerca, alguns estudis fan servir un motor dissenyat *ex profeso* (Archambault et al., 2014; Science-Metrix, 2018), mentre que d'altres utilitzen Google Scholar (Martín-Martín et al., 2018).

En el primer cas, la feina desenvolupada en aquests treballs ha donat lloc a la creació del motor de cerca 1findr, que permet localitzar tota mena de bibliografia científica i filtrar els resultats en funció de la disponibilitat dels documents (figura 1).

Figura 1. Interfície de cerca a 1findr (1findr.1science.com)

Per la seva banda, Google Scholar és un motor de cerca que indexa la bibliografia científica que rastreja a la web, incloent-hi continguts que requereixen una subscripció, publicacions en accés obert i còpies en repositoris, pàgines personals o xarxes socials. La figura 2 mostra un exemple de registre a Google Scholar que identifica tant la versió d'un article publicada al web de la revista com una còpia disponible en la pàgina web de l'autor. Com es pot observar, la identificació de documents en accés obert a través de Google Scholar fa referència a l'existència d'una còpia que es pot consultar gratuïtament.

Figura 2. Registre a Google Scholar (scholar.google.com)

Al marge de la dificultat de classificar els documents recuperats a través de Google Scholar com d'accés obert segons la llicència de què disposen, la principal dificultat per la utilització del cercador en aquest tipus d'estudis ve donada per la impossibilitat de replicar els resultats obtinguts degut a la laboriositat del procés, tal com descriuen Martín-Martín et al. (2018, p. 835): «perhaps one of the most important limitations of this study is that it is not easily replicable because of the limitations on data extraction imposed by Google Scholar. Extracting a large amount of data from this source is still only possible if one is willing to commit an inordinate amount of time to the task.» Google Scholar ofereix escasses prestacions d'exportació massiva de resultats. No disposa d'una API que permeti executar consultes a gran escala o descarregar grans volums de dades.

3.2 Unpaywall

Unpaywall és una eina creada per l'organització sense ànim de lucre Our Research, prèviament anomenada ImpactStory. Indexa continguts publicats en revistes d'accés obert i disponibles en repositoris¹². Cal remarcar que únicament recopila documents que tinguin un DOI i que estiguin disponibles en fonts «legítimes», obviant còpies en ResearchGate o Sci-Hub (un web pirata de publicacions científiques). Segons la informació proporcionada al seu web, el maig de 2021 tenia indexats més de 29 milions de documents en accés obert. El sistema es pot implementar com una extensió del navegador que indica a l'usuari si hi ha disponible una còpia en accés obert del document que està consultant en pantalla. Així mateix, disposa d'una API pública i un *Data Feed* que requereix una subscripció per descarregar la base de dades actualitzada.

Unpaywall classifica els articles en alguna de les següents cinc categories: «tancat», «verd» (disponible en un repositori), «daurat» (publicat a una revista en accés obert), «híbrid» o «bronze»¹³. La distinció entre «híbrid» i «bronze» radica en què els articles classificats en la primera categoria compten amb una «licència oberta» (*open license*) que els creadors d'Unpaywall defineixen de forma laxa. Com que l'eina informa sobre el tipus de llicència de cada article, si l'usuari ho desitja pot establir criteris de classificació més estrictes: «we tend to err in the direction of greatest inclusiveness - in other words, our cutoff for “is this license open” is very low. We include even relatively restrictive licenses (e.g. CC-BY-NC-ND), and leave it to downstream users to filter these out if they want to.» Com que un mateix document pot estar disponible en accés obert a través de diferents vies, Unpaywall

¹² La llista de fonts indexades es pot consultar a <https://unpaywall.org/sources>

¹³ Més informació a: Unpaywall FAQ,

<https://support.unpaywall.org/support/solutions/folders/44000384007>

identifica una d'elles com la «millor» (*best_oa_location*). Per exemple, la versió disponible al web de l'editorial és preferida a la dipositada a un repositori.

Alguns treballs han ofert dades sobre la cobertura i fiabilitat d'Unpaywall. Piwowar et al. (2018) van seleccionar una mostra de 500 articles d'entre la població objecte d'estudi i els van buscar manualment a Internet. Els resultats mostraven una precisió de l'eina del 96,6%. Es a dir, pràcticament sempre que Unpaywall determinava que un article estava en accés obert, efectivament estava disponible. L'exhaustivitat era inferior: Unpaywall era capaç de trobar el 77% dels articles que estaven disponibles en accés obert.

Posteriorment, Akbaritabar i Stahlschmidt (2019) van analitzar la disponibilitat en accés obert dels documents indexats en la base de dades del Kompetenzzentrum Bibliometrie¹⁴ (Competence Centre for Bibliometrics), una entitat alemanya creada l'any 2008 per dur a terme tasques d'avaluació i monitoratge. Els autors van combinar els DOIs de les publicacions d'autors alemanys a la seva base de dades amb Crossref i Unpaywall per recopilar informació sobre les llicències i disponibilitat dels articles. L'estudi evidenciava problemes en la correcta identificació de DOIs en els registres de Web of Science i Scopus —que es van emprar com a punt de partida— i en la diversitat de llicències assignades a un mateix article a Crossref. Com a resultat, els autors van alertar de discrepàncies en la determinació de la disponibilitat en accés obert del 13% dels articles, majoritàriament articles classificats com tancats a Crossref, però com disponibles en accés obert a Unpaywall.

Les dues principals bases de dades bibliogràfiques de caràcter multidisciplinari a nivell internacional, Web of Science¹⁵ i Scopus¹⁶, incorporen filtres sobre la disponibilitat en accés obert dels documents a partir de la informació subministrada per Unpaywall (figura 3). El treball de Bosman i Kramer (2018) descrit a l'apartat anterior feia servir directament les dades subministrades per Web of Science per mesurar la disponibilitat en accés obert sense consultar Unpaywall.

¹⁴ <https://www.bibliometrie.info>

¹⁵ <https://clarivate.com/webofsciencegroup/solutions/open-access/>

¹⁶ <https://blog.scopus.com/posts/scopus-makes-millions-of-open-access-articles-easily-discoverable>

Figura 3. Filtres de documents en accés obert a Web of Science (dalt) i Scopus (baix) basats en Unpaywall

The top screenshot shows the Web of Science interface. On the left, a sidebar titled 'Open Access' is highlighted with a red box. It contains the following options:

- All Open Access (4,719)
- DOAJ Gold (1,822)
- Other Gold (544)
- Bronze (1,896)
- Green Published (3,483)
- Green Accepted (105)

Below these options is a 'Refine' button and a link to 'Learn more about Open Access versioning in Web of Science'. The main content area shows search results for COVID-19 related papers, with 'Free Full Text from Publisher' buttons highlighted in yellow.

The bottom screenshot shows the Scopus interface. On the left, a sidebar titled 'Refine results' is highlighted with a red box. It contains the following options:

- All Open Access (6,267)
- Gold (2,385)
- Hybrid Gold (550)
- Bronze (2,809)
- Green (3,675)

Below these options is a 'Learn more' link and a 'Year' filter showing 2021 with 3,221 results. The main content area shows a table of search results with columns for Document title, Authors, Year, and Source.

Document title	Authors	Year	Source
1 The advisory committee on immunization practices' interim recommendation for use of pfizer-BioNTech COVID-19 Vaccine — United States, December 2020 <i>Open Access</i>	Oliver, S.E., Gargano, J.W., Marin, M., (...), Bell, B.P., Doolling, K.	2021	Morbidity and Mortality Weekly Report 69(50), pp. 1922-1924
2 As vaccines emerge, a global waiting game begins	Cohen, J., Kupferschmidt, K.	2021	Science 370(6523), pp. 1385-1387

4. Indicadors nacionals d'accés obert

En aquest apartat s'analitzen, des d'un punt de vista eminentment metodològic, les característiques de diversos indicadors nacionals sobre la disponibilitat de bibliografia científica en accés obert. Començarem amb la descripció de dos instruments amb cobertura internacional i, a continuació, analitzarem els indicadors disponibles a Alemanya, Dinamarca, França i Països Baixos, per finalitzar amb les eines disponibles a Catalunya.

4.1 *Iniciatives internacionals*

Unió Europea, *Open Science Monitor*

https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en

A nivell europeu, l'Open Science Monitor fa un seguiment de diferents indicadors científics, entre d'altres la disponibilitat en accés obert de les publicacions d'autors afiliats a institucions dels països membres. L'indicador pren com a punt de partida les publicacions indexades a la base de dades Scopus i utilitza Unpaywall per determinar la disponibilitat en accés obert. El document que descriu la metodologia emprada (Open Science Monitor, 2019) inclou un apèndix elaborat per Thed van Leeuwen i Rodrigo Costas, del CWTS de la Universitat de Leiden. A més de recalcar que l'indicador limita el seu abast als documents que tenen assignat un DOI a Scopus, els autors mencionen les dificultats que planteja la categoria d'accés obert «híbrid» a Unpaywall en la que van trobar «some form of mixing up of Gold with true Hybrid (which means APC-based publishing in an otherwise tool [sic] access journal).» Una altra categoria problemàtica és la d'accés obert «bronze». Malgrat que els autors afirmen que no la consideraran com a una categoria independent, el cert és que sí es pot visualitzar per separat a la pàgina web de l'indicador: «A fourth OA category in Unpaywall is 'Bronze', which is basically a form of openly available publishing initiated by the publishers, in which the copyright status is not clear. As in our criteria this is not a sustainable form of OA we opt for not considering as a separate OA category, although it will be included in the overall consideration of OA publications.»

Austràlia, *Curtin Open Knowledge Initiative*

<https://openknowledge.community/dashboards/coki-open-access-dashboard/>

A Austràlia, la Curtin Open Knowledge Initiative (COKI) ofereix un *Open Access Dashboard* que reporta dades d'accés obert per països. Segons la metodologia¹⁷, les dades provenen de Crossref i es completen amb cerques a Microsoft Academic. S'ignora el futur d'aquest projecte per quant Microsoft Academic va anunciar a inicis de maig de 2021 que deixaria d'estar operatiu el desembre del mateix any¹⁸.

¹⁷ <https://openknowledge.community/wp-content/uploads/2020/11/Open-Access-Dashboard-Data-Details.pdf>

¹⁸ <https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/>

Tot i que no es tracta d'un indicador, en la línia metodològica de combinar múltiples fonts per determinar la disponibilitat en accés obert, cal ressenyar un estudi recent de White et al. (2021) per mesurar la disponibilitat de la producció dels investigadors de les vuit universitats de Nova Zelanda. Els autors van compilar una llista d'articles amb DOI i van interrogar diverses eines: Unpaywall —per conèixer la disponibilitat en accés obert—; Web of Science —per identificar l'autor de correspondència i les fonts de finançament—; DOAJ —per determinar l'existència i la quantia de les APC—; Sherpa Romeo¹⁹ —per conèixer les polítiques editorials respecte del dipòsit dels articles—; i Crossref —per quantificar les citacions rebudes. A nivell global, els resultats mostraven que, en 2019, només dos de cada cinc articles publicats per investigadors neozelandesos el 2017 estaven disponibles en accés obert.

4.2 Alemanya

Open Access Monitor

<https://open-access-monitor.de>

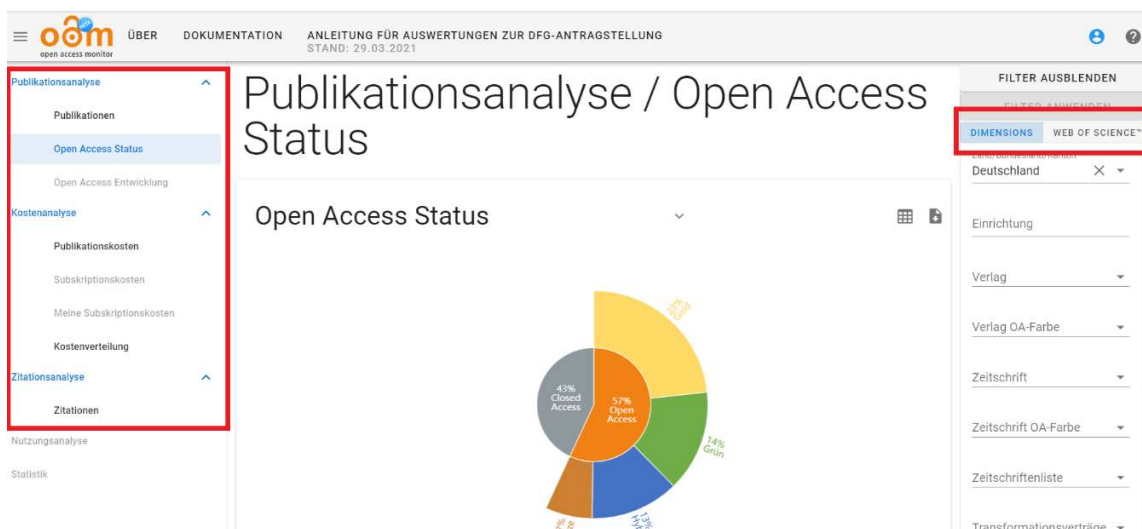
A Alemanya, l'Open Access Monitor és un projecte governamental anunciat l'any 2017 que es va adjudicar al centre d'investigació Forschungszentrum Jülich. Actualment, la seva implementació s'anuncia al web com en versió *beta*. La figura 4 permet observar els trets més destacats de l'indicador. En el menú de l'esquerra, es distingeixen tres apartats relatius a la disponibilitat de les publicacions en accés obert, l'anàlisi de costos (pagaments d'APC) i l'impacte en termes de citacions. Malgrat que la informació metodològica disponible en anglès és escassa,²⁰ el menú de la dreta suggereix que les dades relatives a les publicacions s'obtenen de dues fonts: Dimensions i Web of Science.

¹⁹ <https://v2.sherpa.ac.uk/romeo/>

²⁰ <https://www.fz->

[juelich.de/zb/EN/About_us/cooperation/syn_oa_monitoring/syn_oa_monitoring_node.html](https://www.fz-juelich.de/zb/EN/About_us/cooperation/syn_oa_monitoring/syn_oa_monitoring_node.html)

Figura 4. Interfície de l'Open Access Monitor (open-access-monitor.de)



El setembre de 2020, Clarivate Analytics, empresa propietària de Web of Science, va anunciar un acord per subministrar informació sobre finançament i publicacions de recerca a les autoritats alemanyes, austríaques i suïsses que sigui emprada en l'elaboració de l'Open Access Monitor.²¹

Un article recent (Hobert et al., 2021) sobre la disponibilitat de publicacions en accés obert a Alemanya no utilitzava la informació d'aquest indicador, sinó que duia a terme un procés original de recollida i anàlisi de dades combinant diverses fonts: la base de dades del Kompetenzzentrum Bibliometrie ja mencionada en parlar de l'estudi d'Akbaritabar i Stahlschmidt (2019); una llista de revistes en accés obert compilada per la Universitat de Bielefeld²²; Unpaywall; i OpenDOAR²³, un directori de repositoris d'accés obert. L'estudi conclouia que el 45% dels articles publicats entre 2010 y 2018 per autors afiliats a institucions de recerca alemanyes estava disponible en accés obert en el moment de recollida de dades.

²¹ <https://clarivate.com/news/clarivate-collaboration-with-open-access-monitor-germany-to-provide-web-of-science-data-across-dach-region/>

²² <https://doi.org/10.4119/unibi/2934907>

²³ <https://v2.sherpa.ac.uk/opensoar/>

4.3 Dinamarca

Open Access Indicator

<https://www.oaindikator.dk/en/>

El Danish Open Access Indicator té per finalitat fer el seguiment del grau de compliment de la *Danish Open Access Strategy 2018-2025*, que es marca com objectiu arribar al 100% d'accés obert l'any 2025 amb un màxim de dotze mesos d'embargament. L'indicador pren com a punt de partida les publicacions incloses en els sistemes d'informació de les vuit universitats daneses. Aquestes dades es complementen amb altres fonts d'informació: DOAJ, per conèixer les característiques de les revistes d'accés obert; Sherpa Romeo, per esbrinar les polítiques editorials; i sengles llistats de repositoris validats i revistes que imposen un període d'embargament superior als dotze mesos que marca l'estratègia danesa.

Pel que fa a les rutes d'accés obert, l'indicador contempla l'accés obert «daurat» (en revistes d'accés obert) i «verd» (en repositoris, diferenciant els repositoris de les universitats daneses d'altres repositoris externs validats), però no considera l'accés obert «híbrid», a no ser que els articles en qüestió estiguin també a un repositori.

L'indicador disposa d'una detallada descripció metodològica²⁴. En destacarem tres aspectes. En primer lloc, l'indicador es publica anualment, al mes d'abril, permetent l'embargament de dotze mesos previst a l'estratègia danesa. Existeix un calendari que detalla les etapes de construcció de l'indicador: recollida de dades, audiència a les universitats, ajustaments, etc. En segon lloc, l'indicador no només mesura l'accés obert assolit, sinó també l'accés obert «potencial», és a dir, el percentatge de documents que podrien estar disponibles en obert a través de repositoris complint les polítiques editorials de les revistes a les quals s'han publicat. En tercer lloc, per garantir la transparència i la reproductibilitat de l'indicador, és possible descarregar els *datasets* amb les dades individuals de tots els articles considerats per al càlcul. Fruit d'aquest minucios procediment metodològic, l'indicador ha estat assenyalat com un exemple de bona pràctica a l'informe de NORF (2021).

²⁴ https://www.oaindikator.dk/oa-docs/overview/Open_Access_Indicator_2018_Overview_en.pdf
https://www.oaindikator.dk/oa-docs/technical/Open_Access_Indicator_2019_Technical_en.pdf

4.4 França

Baromètre français de la Science Ouverte

<https://ministeresuprecherche.github.io/bsol/>

Malgrat el seu nom suggereix un ventall més ampli d'indicadors, hores d'ara el *Baromètre français de la Science Ouverte* presenta primordialment informació sobre la disponibilitat de les publicacions en accés obert. El document metodològic sobre la seva construcció (Jeangirard, 2019) destaca que, a diferència de l'indicador de la Unió Europea, que es basa en Scopus, en aquest cas es va optar per emprar únicament fonts públiques. El procés de càlcul de l'indicador s'estructura en tres etapes: s'inicia amb la identificació de les publicacions (documents amb DOI majoritàriament extrets d'Unpaywall) i l'anàlisi de la font original per detectar la presència d'algun autor amb afiliació a una institució d'investigació francesa; la segona etapa consisteix a enriquir les metadades: afegir l'ORCID dels autors o altres identificadors d'abast local com IdRef²⁵, classificar els documents per disciplines, etc.; finalment, la tercera etapa té per objectiu determinar si el document en qüestió està disponible en accés obert, tasca per a la qual s'utilitzen Unpaywall i HAL²⁶. Les dades recollides en 2018 mostraven que la taxa de disponibilitat en accés obert de publicacions d'autors afiliats a institucions franceses oscil·lava entre el 39% i el 42% per documents publicats entre 2013 i 2017 tot i que amb diferències substancials per tipologia documental, editorial i disciplina.

4.5 Països Baixos

Monitor

<https://www.openaccess.nl/en/in-the-netherlands/monitor>

Entre els anys 2016 i 2017, l'associació d'universitats neerlandeses (VSNU) va dur a terme un estudi sobre la disponibilitat en accés obert de la producció científica del país. Aquestes dades han estat actualitzades en successives ocasions —l'última el març de 2021— per oferir algunes estadístiques no gaire detallades, especialment d'accés obert híbrid.

També als Països Baixos, el portal NARCIS²⁷ ofereix estadístiques sobre el percentatge de documents en obert, tancats, d'accés restringit (*restricted*) i embargats (*embargoed*) disponibles des de 2000. Malauradament no hi ha gaire informació sobre l'abast de cada categoria ni sobre el seu càlcul (per exemple, no queda clara la distinció entre *restricted* i *embargoed*), tot i que una presentació de

²⁵ <https://www.idref.fr>

²⁶ HAL (<https://hal.archives-ouvertes.fr>) és un repositori multidisciplinari

²⁷ <https://www.narcis.nl>

2018 il·lustra algunes discrepàncies entre les dades d'accés obert de NARCIS i les ofertes per Unpaywall (Bode, 2018).

4.6 Catalunya

Observatori de l'Accés Obert

<https://bibliotecnica.upc.edu/observatori>

L'Observatori, hostejat al web de la Universitat Politècnica de Catalunya, té per objectiu «fer el seguiment de l'estat de l'Accés Obert a les universitats catalanes.» El fons documental sobre el qual es duu a terme aquest seguiment resulta de la combinació dels registres amb DOI extrets de Scopus i Web of Science. Tot i que a la metodologia de l'Observatori no s'explicita, sembla que s'utilitza Unpaywall (Rovira i Labastida, 2019) per determinar la disponibilitat en accés obert de les publicacions considerant quatre categories: daurat, híbrid, bronze i verd. Els resultats s'actualitzen dos cops l'any, els mesos d'abril i octubre, i s'anuncia que en el futur, l'Observatori «presentarà els costos dels investigadors i de les institucions a l'hora de publicar en obert.»

Portal de la Recerca de Catalunya

<https://portalrecerca.csuc.cat/estadistiques/publicacions/accesobert>

El Portal té com a finalitat visualitzar i difondre l'activitat investigadora duta a terme a Catalunya, facilitant l'accés obert a la producció científica. A llarg termini, també es pretén simplificar els processos de gestió i recolzar els d'avaluació. Per aconseguir-ho, s'aprofita la feina feta per les universitats i centres de recerca als seus sistemes de gestió de la recerca (CRIS²⁸) dels quals es nodreix el Portal. Al seu apartat d'estadístiques, el Portal disposa d'una secció sobre accés obert amb dades globals i per institució. Per determinar el percentatge de publicacions en accés obert, s'extreu l'URL de les publicacions que en disposen. Si aquest URL és un DOI, s'executa a Unpaywall. En la resta de casos, es determina si es tracta d'un domini validat com d'accés obert pel Grup de Treball de Suport a la Recerca.

A nivell estatal no existeix un indicador d'accés obert similar als descrits amb anterioritat. No obstant, la Red de Bibliotecas Universitarias Españolas va elaborar unes recomanacions sobre la monitorització de l'accés obert (REBIUN, 2019a) a partir de les quals va preparar un informe sobre la disponibilitat en accés obert de la producció de les universitats espanyoles i del CSIC (REBIUN, 2019b). L'estudi es

²⁸ CRIS és l'acrònim de Current Research Information System.

basa en la identificació de les publicacions amb DOI indexades a Scopus i Web of Science per comprovar, a través d'Unpaywall, la seva disponibilitat en obert.

Taula 1. Resum de les característiques dels indicadors d'accés obert

	Col·lecció documental	Disponibilitat en accés obert	Tipologies d'accés obert	Freqüència
Unió Europea, <i>Open Science Monitor</i>	Scopus	Unpaywall	<i>Gold, green, hybrid i bronze</i>	Anual
Austràlia, <i>Curtin Open Knowledge Initiative</i>	Crossref i Microsoft Academic	No s'indica	<i>Gold, green, hybrid i bronze</i>	Anual
Alemanya, <i>Open Access Monitor</i>	Dimensions i Web of Science	Unpaywall	<i>Gold, green, hybrid i bronze</i>	Anual
Dinamarca, <i>Open Access Indicator</i>	CRIS de les universitats	DOAJ i llistat de repositoris	<i>Gold i green</i>	Anual
França, <i>Baromètre français de la Science Ouverte</i>	Diverses fonts obertes	Unpaywall	<i>Gold i green</i>	Anual
<i>Observatori de l'Accés Obert</i>	Scopus i Web of Science	Unpaywall	<i>Gold, green, hybrid i bronze</i>	Biennal
<i>Portal de la Recerca de Catalunya</i>	CRIS de les universitats	Unpaywall i llistat de repositoris	No es distingeixen categories	Anual

5. Conclusions

Els apartats anteriors han posat de manifest el creixent interès d'agències de finançament i d'institucions de recerca per mesurar l'evolució de la disponibilitat de publicacions en accés obert. Fruit de la necessitat d'homogeneïtzar aquests processos per evitar les discrepàncies que també venim de descriure, Science Europe, una associació que aplega una trentena d'entitats que financen o duen a terme recerca científica, va publicar el maig de 2021 un conjunt de recomanacions sobre el mesurament de la disponibilitat de publicacions en accés obert (Philipp et al., 2021). En fer un exercici de monitoratge caldria diferenciar tres etapes: definir el propòsit de l'indicador —per què es vol mesurar el nivell d'accés obert?—; identificar el conjunt de publicacions objecte d'anàlisi; i fixar el procés de recollida, interpretació i presentació de les dades. Malgrat els objectius d'institucions diverses puguin diferir, habitualment hi ha un interès per comparar entitats que requereix d'una certa col·laboració i estandardització: «Although purpose, timeframe and other parameters might differ between organisations interested in monitoring Open Access, an aligned implementation and collaboration can improve the impact these efforts will have. This would not only save resources, but also contribute to sustainable practices overall, and facilitate the inevitable comparisons between results of Open Access monitoring exercises.»

A la introducció d'aquest informe mostràvem les discrepàncies en les xifres de disponibilitat en accés obert de publicacions de la Universitat Pompeu Fabra mostrades per l'Observatori de l'Accés Obert i pel Portal de la Recerca de Catalunya. En el primer cas, l'Observatori recopila les dades de producció a Scopus i Web of Science, selecciona els documents que tenen DOI i determina la seva disponibilitat en accés obert amb Unpaywall. El percentatge de disponibilitat es calcula sobre els documents amb DOI. En el cas del Portal, les dades de producció provenen dels sistemes d'informació científica de les universitats. En molts casos, els registres no contenen un DOI, bé perquè el document no en té cap d'assignat, bé perquè l'investigador que ha incorporat el registre a la base de dades no l'ha afegit. En mancar un DOI, Unpaywall és incapaç de determinar la disponibilitat en accés obert del document i el classifica com tancat, però l'indicador es calcula sobre el conjunt de la col·lecció, mostrant una xifra de disponibilitat sensiblement inferior a la de l'Observatori.

D'altra banda, la cobertura de la producció científica a Scopus i Web of Science difereix de la del Portal. Una cerca, a inicis de 2021, de la producció de la Universitat Pompeu Fabra l'any 2019 a les tres fonts oferia resultats quantitativament similars: 1954 documents a Web of Science, 2053 documents a Scopus i 1985 documents al Portal. No obstant, es tractava de documents diferents. A les dues primeres fonts trobàvem de manera gairebé exclusiva articles de revista amb un elevat grau de

solapament (el 84% de la producció de la Pompeu Fabra a Web of Science també estava indexada a Scopus i, a l'inrevés, el 77% de la producció a Scopus també estava a Web of Science). En el cas del Portal, però, els 1207 articles només representen el 61% dels 1985 documents de la institució, amb un volum important de documents indexats a Scopus i Web of Science absents. Entre el 39% restant hi havia un elevat nombre de monografies que no acostumen a estar en accés obert.

Aquest exemple il·lustra com, en crear un indicador d'accés obert, és necessari prendre decisions relatives a la col·lecció documental sobre la qual es realitzarà el càlcul, l'eina emprada per determinar la disponibilitat en accés obert, la definició i tipologies d'accés obert considerades, la possible consideració de fonts d'informació addicionals, els aspectes temporals i la publicació de la documentació necessària per garantir la reproductibilitat de l'indicador.

5.1 Col·lecció documental per al càlcul de l'indicador

El primer element a considerar és la col·lecció documental sobre la qual es durà a terme el càlcul de l'indicador. En el cas de l'indicador previst al *Pacte Nacional per a la Societat del Coneixement*, donat que previsiblement no estarà aïllat, sinó que formarà part d'una bateria més àmplia, caldria determinar si hi ha un indicador més bàsic adreçat a mesurar la producció científica de Catalunya. Si és així, sembla lògic fer convergir tots dos i utilitzar el primer com a denominador per al càlcul del segon.

En els treballs revisats hem observat que, majoritàriament, es fan servir les bases de dades Scopus i Web of Science per a la identificació de la bibliografia. Són eines reconegudes a nivell internacional i disposen de prestacions per la recuperació i exportació àgil dels registres. A més, donat que totes dues bases de dades tenen implementat un filtre de documents en accés obert basat en Unpaywall, seria possible fer una anàlisi de disponibilitat de manera molt senzilla. No obstant, també cal tenir presents les limitacions d'aquestes eines. Indexen primordialment articles de revista, deixant de banda altres formats documentals com les actes de congressos o les monografies. Al mateix temps, han estat criticades pel seu biaix anglosaxó amb una cobertura deficient de publicacions en llengües alienes a l'anglès. Per al càlcul del baròmetre francès, aquests productes han estat descartats pel seu caràcter comercial en benefici de fonts públiques.

En el cas de l'indicador català, donada l'existència d'un Portal de la Recerca que té per objectiu recollir la producció científica del país, sembla lògic utilitzar-lo com a punt de partida per al càlcul de l'indicador. Aquesta és l'estratègia que se segueix, per exemple, a Dinamarca. Això permetria defugir les limitacions de cobertura de bases de dades com Scopus o Web of Science i ajustar-se millor a la producció bibliogràfica dels investigadors catalans. El principal inconvenient radica en què, en

estar nodrit pels sistemes d'informació de les universitats i centres CERCA (fins a 51 entitats en l'actualitat), la qualitat en la descripció dels documents pot variar sensiblement entre institucions. L'absència de metadades com el DOI pot esdevenir una limitació indefugible en la identificació de les còpies en accés obert.

5.2 Eina per identificar la producció en accés obert

Un segon aspecte a valorar és la font utilitzada per determinar la disponibilitat en accés obert de les publicacions. Com hem observat, la major part de treballs opten o bé per un motor de cerca de bibliografia científica o bé per Unpaywall. La primera opció presenta serioses limitacions, tant pel que fa a la inclusió de còpies disponibles en pàgines personals o xarxes socials —que estrictament no compleixen els requisits per ser catalogades com d'accés obert— com a la laboriositat del procés de recollida de dades i la impossibilitat de replicar-lo. En conseqüència, la major part d'estudis identificats a la bibliografia utilitzen Unpaywall com a font d'informació fins al punt que, com asseguren Hobert et al. (2021), «the OA discovery service Unpaywall has become the de facto standard for OA bibliometric studies.»

5.3 Definició i categories d'accés obert

Un tercer element a valorar és la definició d'accés obert emprada. Aquesta qüestió està molt vinculada a la decisió prèvia sobre la font a utilitzar per determinar la disponibilitat. Els estudis que utilitzen motors de cerca fan servir definicions laxes d'accés obert, equiparant-lo a l'accés gratuït a les publicacions. Per exemple, el treball d'Archambault et al. (2014) utilitzava el que els autors qualificaven com «a simple definition of OA—freely available online to all (no money had to be paid, no registration to a service or website had to be made).» De manera similar, l'article de Martín-Martín et al. (2018) a partir de Google Scholar se centrava en la identificació de documents «freely available». La tria d'Unpaywall implica l'adopció d'una definició més estricta d'accés obert, ja que l'eina no pren en consideració les còpies d'articles disponibles en pàgines personals o xarxes socials.

Una altra qüestió a sospesar són les «rutes» de l'accés obert. Hi ha consens a considerar els continguts disponibles en revistes d'accés obert (*gold open access*) i en repositoris (*green open access*). Altres modalitats generen més dissensió. Hem vist que, en ocasions, Unpaywall té dificultats per distingir entre accés obert daurat i híbrid o que, en el cas de Dinamarca, aquesta última ruta no es contempla. De manera similar, hi ha discrepàncies sobre l'accés obert bronze que alguns indicadors com el de la Unió Europea no consideren sostenible en no permetre la reutilització.

Una qüestió que també pot plantejar-se és en quina mesura una determinada versió d'un article compleix amb la seva difusió en accés obert. Per exemple, un *preprint* sense revisar d'un article publicat posteriorment a una revista es considera una forma d'accés obert? Si els dos documents tenen DOIs diferents i s'utilitza Unpaywall per mesurar la disponibilitat, seran considerats documents diferents.

5.4 Fonts addicionals d'informació

Una característica de nombrosos estudis i indicadors és la combinació de múltiples fonts d'informació. A més de les dades de disponibilitat en accés obert d'Unpaywall, alguns treballs recopilen informació de Crossref per conèixer de primera mà el tipus de llicència que tenen els articles o el nombre de citacions que han rebut; DOAJ informa de les taxes d'edició que cobren les revistes, de manera que es poden fer estimacions del cost del trànsit a l'accés obert; Sherpa Romeo facilita informació sobre les polítiques editorials de les revistes que permet estimar el volum d'articles que podrien estar disponibles en accés obert a través de repositoris; etc. La integració d'aquestes dades permet oferir un indicador d'accés obert més ric, que vagi més enllà del càlcul de la disponibilitat de les publicacions.

5.5 Cronologia

La taxa de disponibilitat en accés obert de les publicacions no és una dada fixa, sinó que varia en funció de dues característiques cronològiques: el període de temps durant el qual es van publicar els articles considerats i el moment de recollida de dades. En el cas de Dinamarca, donat que l'estratègia d'assoliment de l'accés obert permet un embargament de fins a dotze mesos, es deixa passar un any des de la publicació dels articles fins al moment de càlcul de l'indicador, que també es publica amb freqüència anual.

5.6 Reproductibilitat

Per tal de dotar de transparència a l'indicador i garantir que sigui reproducible serà necessari publicar un document metodològic amb informació detallada del procediment de càlcul. En aquest sentit, caldria valorar la publicació del conjunt de dades en les que es sustenta l'indicador tal com es fa en el cas de Dinamarca.

6. Recomanacions finals

La revisió de les iniciatives internacionals de monitorització de l'accés obert mostra que la decisió d'establir un indicador d'accés obert a la producció científica de Catalunya és oportuna i útil. Es tracta d'una iniciativa viable, que no requereix de grans recursos, i que està en sintonia amb el que s'està fent a la Unió Europea per mesurar la concreció de les polítiques de ciència oberta.

L'anàlisi de diversos models ha posat de manifest la necessitat de prendre un seguit de decisions que s'haurien de consensuar entre les institucions responsables de la producció bibliogràfica i els agents que utilitzaran l'indicador.

Decisions metodològiques:

- El Portal de la Recerca de Catalunya seria la font més adient per identificar la producció científica de Catalunya, tot i que caldria garantir la compleció de les dades per part de les institucions participants.
- Unpaywall sembla la millor eina per determinar la disponibilitat en accés obert d'aquesta producció. En ser l'instrument emprat a la major part d'iniciatives internacionals, facilitaria la comparació de les dades.
- Per bé que alguns indicadors es completen amb dades d'altres fonts per determinar, per exemple, el cost del trànsit a l'accés obert, en aquesta primera etapa segurament seria més adient limitar l'abast de la iniciativa al mesurament de la disponibilitat de les publicacions en accés obert.
- L'indicador podria publicar-se amb freqüència anual, amb un calendari públic que facilités la revisió de les dades per part de les institucions participants i tingués en compte els possibles períodes d'embargament.
- Seria convenient publicar la informació metodològica i les dades recollides que permetessin reproduir el càlcul de l'indicador.

Referències

- Akbaritabar, A., & Stahlschmidt, S. (2019). *Applying Crossref and Unpaywall information to identify gold, hidden gold, hybrid and delayed Open Access publications in the KB publication corpus*.
<https://osf.io/sdzft/download>
- Archambault, É.; Amyot, D.; Deschamps, P.; Nicol, A.; Provencher, F.; Rebout, L.; Roberge, G. (2014). *Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013*.
https://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf
- Bode, E. (2018). Open Access monitoring using NARCIS: What if sources conflict? *Workshop Monitoring Open Access*.
https://dans.knaw.nl/nl/actueel/Workshop_monitoring_openaccess_EmilBode.pptx
- Bosman, J.; Kramer, B. (2018). Open access levels: a quantitative exploration using Web of Science and oaDOI data (No. e3520v1). *PeerJ Preprints*, 6:e3520v1.
<https://doi.org/10.7287/peerj.preprints.3520v1>
- Else, H. (2018a). How Unpaywall is transforming open science. *Nature*, 560, 291–292.
<https://doi.org/10.1038/d41586-018-05968-3>
- Else, H. (2018b). How I scraped data from Google Scholar. *Nature*.
<https://doi.org/10.1038/d41586-018-04190-5>
- Generalitat de Catalunya. (2020). *Pacte Nacional per a la Societat del Coneixement*.
http://empresa.gencat.cat/web/.content/actualitat/pacte-nacional-societat-coneixement/documents/document_final_multiidioma/Pacte-Nacional-per-a-la-Societat-del-Coneixement.pdf
- Hobert, A.; Jahn, N.; Mayr, P.; Schmidt, B.; Taubert, N. C. (2021). Open Access uptake in Germany 2010–18: Adoption in a diverse research landscape. *Scientometrics*, en premsa. <https://doi.org/10.1007/s11192-021-04002-0>
- Jeangirard, E. (2019). *Monitoring Open Access at a national level: French case study*. <https://hal.archives-ouvertes.fr/hal-02141819>
- Maddi, A. (2020) Measuring open access publications: a novel normalized open access indicator. *Scientometrics*, 124, 379–398.
<https://doi.org/10.1007/s11192-020-03470-0>
- Maddi, A.; Lardreau, E.; Sapinho, D. (2021). Open access in Europe: a national and regional comparison. *Scientometrics*, 126, 3131–3152.
<https://doi.org/10.1007/s11192-021-03887-1>

- Martín-Martín, A.; Costas, R.; van Leeuwen, T.; Delgado López-Cózar, E. (2018). Evidence of open access of scientific publications in Google Scholar: A large-scale analysis. *Journal of Informetrics*, 12 (3), 819-841. <https://doi.org/10.1016/j.joi.2018.06.012>
- NORF (2021). *National Open Research Landscape Report*. <http://bit.ly/NORFLandscape>
- Open Science Monitor (2019). *Updated methodological note*. https://ec.europa.eu/info/sites/default/files/research_and_innovation/knowledge_publications_tools_and_data/documents/open_science_monitor_methodological_note_april_2019.pdf
- Philipp, T.; Botz, G.; Kita, J-C.; Richards, P.; Sängner, A.; Siegert, O.; Reumaux, M. (2021). *Open Access Monitoring: Guidelines and Recommendations for Research Organisations and Funders*. <https://www.scienceurope.org/our-resources/open-access-monitoring/>
- Piwowar, H.; Priem, J.; Larivière, V.; Alperin, J. P.; Matthias, L.; Norlander, B.; Farley, A.; West, J.; Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>
- Piwowar, H.; Priem, J.; Orr, R. (2019). The Future of OA: A large-scale analysis projecting Open Access publication and readership. *BioRxiv*, 795310. <https://doi.org/10.1101/795310>
- REBIUN. (2019a). *Recomendaciones sobre la monitorización del acceso abierto*. <http://hdl.handle.net/20.500.11967/245>
- REBIUN. (2019b). *Medición del Acceso Abierto en las universidades españolas y el CSIC (2014-2018)*. <http://hdl.handle.net/20.500.11967/354>
- Robinson-Garcia, N.; Costas, R.; van Leeuwen, T.N. (2020). Open Access uptake by universities worldwide. *PeerJ*, 8:e9410. <http://doi.org/10.7717/peerj.9410>
- Rovira, A.; Labastida, I. (2019). The Open Access Observatory. *LIBER Quarterly*, 29 (1), 1–18. <http://doi.org/10.18352/lq.10295>
- Science-Metrix (2018). *Open access availability of scientific publications*. <https://perma.cc/XCN2-6UA9>
- Severin, A.; Egger, M.; Eve, M. P.; Hürlimann, D. (2018). Discipline-specific open access publishing practices and barriers to change: an evidence-based review. *F1000Research*, 7. <https://doi.org/10.12688/f1000research.17328.2>
- Suber, P. Gratis and libre open access. *SPARC Open Access Newsletter*, 124. https://dash.harvard.edu/bitstream/handle/1/4322580/suber_oagratis.html
- White R.K.A.; Angelo, A.; Fitchett D, Fraser M, Hayes L, Howie J, Richardson E, White B. 2021. Only two out of five articles by New Zealand researchers are free-to-access: a multiple API study of access, citations, cost of Article Processing Charges (APC), and the potential to increase the proportion of open access. *PeerJ* 9, e11417. <https://doi.org/10.7717/peerj.11417>

Woutersen-Windhower, S.; Hettne, K. (2020). OA Monitoring: why do we get different results? *Digital Scholarship@Leiden*.

<https://digitalscholarshipleiden.nl/articles/oa-monitoring-why-do-we-get-different-results>