



## RESEARCH INFORMATION NETWORK

### TRANSITION DYNAMICS OF IMPROVING ACCESS TO RESEARCH

---

May 2010

#### Issues Paper

Submitted by:

**CEPA LLP**

In association with

**Mark Ware Consulting Ltd**

---



## CONTENTS

<b>1. Introduction .....</b>	<b>1</b>
1.1. Purpose and structure.....	1
1.2. Definition of the scholarly communication process.....	1
<b>2. Principles informing scenario development.....</b>	<b>2</b>
2.1. Introduction.....	2
2.2. High-level principles.....	2
<b>3. Initial views on scenarios .....</b>	<b>7</b>
3.1. Introduction.....	7
3.2. Green OA Scenario .....	9
3.3. Gold OA .....	13
3.4. Licensing .....	16
3.5. Transactional.....	18
3.6. Summary table.....	20
<b>4. Common issues .....</b>	<b>22</b>
<b>Annex A: Definitions .....</b>	<b>24</b>
<b>Annex B: Other portfolio projects .....</b>	<b>27</b>

## **1. INTRODUCTION**

### **1.1. Purpose and structure**

This paper is the first deliverable of the Dynamics of Improving Access project ('the project') commissioned by the Research Information network (RIN). The work is being undertaken by Cambridge Economic Policy Associates (CEPA) in association with Mark Ware Consulting, between April 2010 and February 2011.

The paper has been developed to begin to explore the key issues and concepts associated with defining scenarios for increased access. It seeks to set out our initial thoughts on the possible dimensions of each of the four routes to access taking account of bilateral kick-off meetings with RIN and with members of the Funding Group<sup>1</sup>.

In particular, the paper is structured as follows:

- Section 2 sets out 'high-level' principles that will inform our selection of scenarios;
- Section 3 begins to define the drivers, scope and dimensions of the scenarios (as indicated in our proposal). Our initial assumptions are summarised for each scenario.
- Section 4 identifies some common issues across the scenarios that will need to be addressed.

We are grateful to the Funder Group for feedback on the issues and assumptions presented in an earlier version of this paper at the meeting held on 14 May.. However, what follows are the initial views of the authors alone.

### **1.2. Definition of the scholarly communication process**

It is important that this study uses a consistent and recognised set of definitions and terminology.

In order to do this, Annex A provide a list of some of the main definitions used in this paper. It is not exhaustive and we expect to develop the list as the project progresses. However, our general approach is to follow the approaches used in two papers where CEPA / Mark Ware consulting have played lead roles:

- "The STM Report: An overview of scientific and scholarly journals publishing", STM / Ware (2009); and
- "Activities, costs and funding flows in the scholarly communications system in the UK", RIN / CEPA (2008).

---

<sup>1</sup> Funding Group members are: the Joint Information Systems Committee (JISC); the Wellcome Trust; Research Libraries UK (RLUK); and the Publishing Research Consortium (PRC).

## **2. PRINCIPLES INFORMING SCENARIO DEVELOPMENT**

### **2.1. Introduction**

This study aims to assess a range of plausible five-year scenarios that will help the stakeholders involved in scholarly communication to “understand better the dynamics of the transitions needed to improve access to research papers in a variety of ways, and the costs, benefits, opportunities and risks that this entails.”<sup>2</sup>

As indicated in the terms of reference we think that it is reasonable to consider four general approaches of increasing access to scholarly communication (hereafter known as ‘routes to access’):

- Open access journals via an ‘author-side’ payment model (‘Gold OA’);
- Open access to articles through self-archiving in conjunction with a subscription (or licensing) model (‘Green OA’);
- An extension to ‘licensing’, such as increasing access via public or academic libraries, access for SMEs and national site licences; and
- transactional solutions, such as micropayments (e.g. an iTunes type model).

We consider that as the scenarios for this study only cover a five-year time horizon, none of these approaches in themselves constitute a ‘scenario’ for analysis. Rather the scenarios will likely be a combination of these approaches, but sufficiently different to one another to allow meaningful comparisons.

### **2.2. High-level principles**

#### **2.2.1. Common core publishing activities**

As we noted in our proposal, journals have typically been understood as fulfilling the following key functions:

- registration (establishing precedence);
- certification (validation e.g. through peer review and other quality-control measures); and
- archiving (preservation for future use and citation).
- dissemination (distribution);

In order to ensure consistency in our scenario analysis, we need to define a common set of activities that will be modelled in each scenario. We set out below those publishing activities that we expect to be ‘core’ (i.e. present in all scenarios) and those activities that might be referred to as ‘value-add’ – i.e. which provide added value, but are not

---

<sup>2</sup> As set out in the terms of reference.

necessarily part of the a basic [irreducible] minimum. The value-added activities may or may not be part of all scenarios.

*'Publication' – registration, certification and archiving,*

In our view where the line is drawn for the publication element is perhaps less important than consistency. This is because, in general, we do not expect that the significant impacts of the scenarios will relate to removal of activity. It seems more likely that impacts will relate to who performs the activity and how it is remunerated. This is not to suggest that there is not room to reduce unit costs or improve efficiency. Rather it relates to the focus and scope of the work - which place a greater emphasis on the whole system as opposed to detailed analysis of the unit costs of specific activities. However, where there is an evidence-based contention that a scenario does result in a change in underlying unit costs, we would expect to illustrate this using sensitivity analysis.

Core activities:

- registration of the article to establish claims of precedence;
- peer review.<sup>3</sup> This is the OSTP roundtable's first 'shared principle'. They note that 'peer review must continue as a critical role in maintaining high quality and editorial integrity';
- long-term preservation or archiving (which includes the process of updating / amending where necessary); and

Value-add activities:

- Copy-editing. There is not the same degree of consensus on this compared with peer review. On the one hand it is possible to take the view that copy editing remains important basic part of the quality control and validation process undertaken in the scholarly communication process. On the other hand it is clear that some researchers are happy to user un-copyedited articles some of the time.

Article selection and journal management. This activity is clearly an important part of the publishing activity that contributes to the filtration function. The reason for referring to it as a value-add activity is that it may not be present, to the same extent, in all scenarios – resulting in both reduced cost and benefit.

*Distribution*

The important, irreducible element is (by definition) that that the articles need to be made available to users. However, the activities here clearly depend on article format (print, e-only, or both) and will therefore vary by scenario.

**Principle 1: The core publishing activities – which we will assume to be part of all scenarios – includes registration, dissemination, peer review and archiving / preservation.**

---

<sup>3</sup> We note that there are different ways of undertaking peer review and these will be considered as part of the scenario development.

## *Access*

As with the Distribution the nature of activities under Access depend significantly on both the article format, and also the route to access (or business model)

### **2.2.2. Plausible extent of change**

By constraining the time horizon for scenarios to five-years and placing an emphasis on transition, the term of reference purposefully requires the scenarios to be **plausible** given where the scholarly communications process currently is in terms of access. However, at the same time, in order to be interesting we anticipate that the scenarios need to show a reasonable degree of change compared to the current situation.

In our view this means that we are looking at ‘moderate’ as opposed to ‘extreme’ shifts in such things as the business models (i.e. author-pays, toll access) and distribution formats (print, print and electronic and e-only). Where ‘extreme’ shifts are changes in characteristics of the access that are both:

- orders of magnitude different to the current scholarly communication system in terms of particular parameters (e.g. moving from 10% to 90% gold OA access); and
- universal access across journal types or subjects groupings;

Such scenarios are illustrated in the CEPA / RIN (2008)<sup>4</sup> analysis and in Houghton *et al* (2009)<sup>5</sup> and in Swan (2010)<sup>6</sup>.

Rather, the scenarios will represent moderate or incremental changes to the scholarly communication system with an expectation of a continued combination of the four ‘routes to access’.

In addition, given the current economic climate, i.e. restrained or reduced spending by Government and private funders, the scenarios will also need to be financially plausible.

**Principle 2: The scenarios will represent plausible changes in the level of access through some combination of the four broad approaches identified in the terms of reference for this project.**

---

<sup>4</sup> Cambridge Economic Policy Associates (CEPA) (2008). *Activities, Costs and Funding Flows in the Scholarly Communications system in the UK*. Research Information Network.

<sup>5</sup> Houghton J, Rasmussen B and Sheehan S; with Oppenheim C, Morris A, Creaser C, Greenwood H, Summers M and Gourlay A (2009). *Economic implications of alternative scholarly publishing models: Exploring the costs and benefits*. Joint Information Systems Committee.

<sup>6</sup> Swan, Alma. Key Perspectives (2010). *Modelling scholarly communication options: Costs and benefits for universities*. Joint Information Systems Committee.

### 2.2.3. Credible ‘drivers’ of and ‘barriers’ to change

Each scenario must also come with what might be referred to as a ‘credible’ narrative in terms of how and why the system is expected to evolve in this direction.

We do not see this work as a ‘futures’ exercise, and do not therefore expect to overemphasise or seek to be overly precise in defining drivers of change. In particular, we will not make any judgement on new technology that may become available within the next five years, rather the drivers/ barriers will be limited to incremental development of the existing technology (as opposed to disruptive innovations). However, the scenarios need to be recognised by the scholarly communication community and the Funding Group as being broadly credible. For example:

- a scenario focusing on Gold OA is unlikely to involve unilateral change on a country scale. We think this is a particular potential critique of some aspects of the Houghton / Swan analysis – although it is not clear whether it impacts on the conclusions.

it is clear that the introduction of policy mandates by funding groups and institutions requiring the self-archiving of peer-reviewed articles is a plausible driver for Green OA scenarios. Funding groups and institutions (e.g. University of Southampton) already require grant recipients/authors to self archive.)

**Principle 3: The scenarios will have clear, but credible, explanations as to how and why there is a change.**

### 2.2.4. Amenable to modelling and quantification

A requirement for this project is to estimate, through modelling, the baseline figures for the current state of provision of access and then the costs-benefits that will accrue to all users through the transition to the scenarios.

Our proposed approach to the modelling aspects of this project are, in summary, to build on the CEPA/ RIN funding flows model for costs and funding flows and to use the modified Solow-Swan growth model (as per Houghton et al) to consider wider economic impact. We remain of the view that this is an appropriate approach.

In order to support the modelling work the scenarios need to be amenable to quantification and modelling. This is in part a technical requirement (i.e. in terms of what the existing modelling tools can deal with). But it also relates to ensuring that:

- scenarios are sufficiently different from (i) each other (heterogeneous scenarios); (ii) the existing scholarly communication system; and
- there is clear sensitivity analysis available to illustrate the impact of changes to particular assumptions.

**Principle 4: The scenarios need to be amenable to modelling / quantification taking account of the technical limits of the model and the need to distinguish between the scenarios**

### **2.2.5. Balanced portfolio**

We recognise that there is considerable amount of disagreement on these issues, both within the community in general and within the Funders group particularly. Given this, an important principle in selecting scenarios will be to consider a balanced portfolio of scenarios, which illustrate the range of possible approaches to increasing access as well as the different views about the impact of changes on the participants of the scholarly communication process.

We consider that the scenarios can be further enhanced through the use of sensitivity analysis. This will provide further insight into the cost-benefit of transition if the assumptions around some of the dimensions are changed.

**Principle 5: The scenarios will be selected to create a balanced portfolio that illustrates (i) a range of possible approaches to increasing access; and (ii) the different views about the impact on scholarly communication system participants**

### 3. INITIAL VIEWS ON SCENARIOS

#### 3.1. Introduction

This section sets out our initial views of the dimensions of four scenarios. These scenarios each place emphasis on one of the ‘routes to access’ in the terms of reference. Although we are not wedded to this, we believe that there is value in this approach given the need to differentiate the scenarios. In any event, we propose to undertake various sensitivities around the dimensions of each route within a given scenario.

#### *Dimensions of scenarios / structure of discussion*

In what follows, for each scenario / route to access we structure our discussion around the following common set of headings (or dimensions):

- Driver (including technical, policy or commercial drivers)?
- Scope of scenario (e.g. disciplines covered)?
- Access to what?
- Access when?
- Access by whom?
- Access where?
- Access how (e.g. who deposits / submits)?

In addition to these questions we also set out some initial thoughts that relate to information requirements about access and expected impact. Some of these observations (particularly where they relate to disciplinary differences) are discussed as they relate to the dimension.

**For the avoidance of doubt, the proposed dimensions for each scenario are not definitive. Rather they are presented as the basis for a discussion with the Funders Group to inform our consultation, and will be informed by the literature review to establish the baseline.**

Another caveat to note is that we are at an early stage in considering how in practice to implement our modelling approach. This means that (in addition to comments received from the funder group and as part of our consultation), the following is subject to change.

#### *Who is accessing?*

A core part of our analysis of scenarios will relate to quantifying **changes in access** and assessing the impact of this in terms of both costs and benefits. In order to do this we expect to consider the impact of access on different sub-groups of user. These sub-groups include:

- institution-based researchers

- independent researchers
- health service
- other professional services (e.g. engineers, lawyers etc)
- corporate (both large and SME)
- central and local government (and other public sector)
- voluntary community sector.

This list is not necessarily exhaustive and it is also possible that these user groups will ‘collapse’ somewhat based on our literature review and available data. But the important point to note is that each scenario will need to clearly articulate marginal access by who; and distinguish as far as possible between the impact of changes in access for different groups. A key aim of the project will be to present metrics of increase in access (as an intermediate output), and the costs and benefits associated with that access.

We note that these groups may change based on the *Gaps in Access* project and, if time permits, we will take the findings of the study into account, as and when they become available.

## 3.2. Green OA Scenario

### 3.2.1. Introduction and summary

Green OA is seen by its proponents as a simple and cost-effective way of increasing access. The big unknown is the extent to which increased access through repositories will undermine subscription revenues required to cover both core and ‘value added’ activities currently undertaken by commercial and society publishers – leaving what has been referred to as an ‘unfunded mandate’. A roughly equivalent formulation presents this in terms of the length of embargo that maximises access utility while minimising impact on subscriptions.

Table 3.1 provides an overview of the shape of a possible base Green OA scenario. The points are discussed in more detail below. Given the potential impact on subscriptions and the fact that the size (or even existence) of this effect is highly contested, we expect as a minimum to consider two variants of this scenario – with a high and low (or no) impact on subscription revenues.

Other base (or variant scenarios) might involve a greater emphasis on institutional mandates and repositories and consider a wider scope (i.e. more subjects).

*Table 3.1: Green OA Scenario Summary*

Dimension	Initial view
Driver?	<b>Tightening of funder/ institutional mandates</b> in terms of monitoring and enforcement – resulting in significant increases in deposit rate compared with baseline.
Scope?	Scenario restricted to a <b>defined subject grouping</b> on a global or UK only basis, as a credible ‘pilot’ in the five-year horizon. Sensitivity analysis will be conducted on the access based on different increases in institutional repositories.
Access to what?	Focus on <b>‘the accepted manuscript’</b> (i.e. peer reviewed) that are deposited. Initial presumption that access is <b>no re-use</b> and it provides <b>low functionality</b> . Sensitivity on access to the version of record and high functionality.
Access when?	<b>Access after a six month embargo</b> . Sensitivity analysis (or alternative scenario) on access to the accepted manuscript or version of record on an immediate basis.
Access by whom?	Access to <b>all user groups</b> – although extent of change in access will depend on scope of scenario as well as baseline access in disciplines.
Access where?	Given the ‘funder mandate’ driver and subject grouping, we assume that primary change in access is through <b>subject-based repositories</b> . But institutional repositories expected to continue and develop.
Access how?	Who deposits is not a key determinant of scenario. Requires assumption of publisher / author mix and appropriate costs.

### 3.2.2. Dimensions

### *Drivers*

Existing drivers for Green OA include:

- (i) funders' policy mandates – driven by a desire to get the greatest possible returns to research;
- (ii) institutional mandates – driver by objectives relating to 'showcasing of work' and management of university research output and IPR; and
- (iii) the availability of (open source) institutional repository software.

Our view is that a green OA scenario will be driven primarily by a **tightening of funder/ institutional mandates in terms of monitoring and enforcement** – reflected in significant increases in deposit (see below for discussion of deposit of 'what'). There is also the possibility that researchers become more attracted to green OA because of benefits they see, e.g. more direct data usage, more demonstrable impact etc

### *Scope of scenario*

Given the Scenario drivers, the scope will be a reflection of changes in funding institutions or groups of UK institution. Ensuring that the scenario is credible and tractable, we believe that there is merit in restricting to a subject group or groups (e.g. STM, economics). It might be possible to consider the costs and impacts of 'piloting' over the period in different subject groupings.

This scenario might reasonably be considered on either a global basis (i.e. covering major US / European funders within the defined subject area) or possibly on a UK only basis. A sensitivity might consider different uptake rates across the different territories e.g. a much higher uptake level in the UK compared to the rest of the world.

### *Access to what?*

The primary Green OA scenario should focus on changes in numbers of **'the accepted manuscript' (i.e. peer-reviewed)** that are deposited. This is the norm for most Green deposit and mandates, but has not yet become common practice. Our presumption too is that access would be to **low functionality** and would **not allow re-use**.

As indicated in Section 2, we not propose to structure any scenario around 'pre-print' (un-peer reviewed) articles.

### *Access when?*

An article – preprint, accepted, version of record – can be deposited in a repository with or without an embargo. We recognise that an important sensitivity in terms of costs, benefit and transition dynamics might be to place greater emphasis on embargoed deposit of the 'version of record' (peer reviewed, registered and archived). We consider that a **six month embargo** is an appropriate starting point for the scenario.

### *Access by whom*

Green OA provides **all user groups** with access to an electronic version of an article. The change in access is expected to vary by:

- discipline – taking into account of existing levels of access. For examples, current practice in some disciplines (e.g. economics, physics) in assessing changes in access in all of our scenarios.
- the availability and quality of discovery tools – i.e. search engines, open biblio databases or paid-for biblio databases.

### *Access where?*

Jones et al (2008) defined three possible repositories types where articles could be deposited – institutional, subject and funder (see Annex A for definitions). For this project, we propose to treat funder and subject repositories as the same.

At present key differences between subject repositories and institutional repositories include:

- subject repositories are able to achieve economies of scale;
- in general articles in subject repositories have higher quality metadata meaning that search and discovery tools are more effective. However, with full article indexing metadata's importance as a discovery tool may be lessening;
- articles also typically have better consistency of content (e.g. data formats); and
- long-term preservation is probably easier to do in a consistent way (although this is a matter of debate).

Taken together this supports the view that – at present at least – subject repositories (where they exist) are likely to be more important platform for access by users than institutional repositories. We therefore see the primary route to change in access in scenario coming through subject repositories over the five-year time horizon.

However, given that over this time period it also seems credible to expect the comparative advantages of a subject repository are likely to reduce (e.g. as a result of development of the quality of metadata on articles institutional repositories; and links between institutional repositories could improve), we do not see a reduction in the number and quality of institutional repositories (rather, within the five-year time horizon the development of the Research Excellent Framework may drive greater usage of institutional repositories).

*Access how (e.g. who deposits / submits)?*

Experience has shown that, even when there is a mandate, authors may not deposit their article into a repository (e.g. current compliance rate with NIH mandate is about 55-60%<sup>7</sup>, and even then this may be carried out by the publisher).

It may be relatively time consuming for an author to enter the metadata and deposit the article (or at any rate be perceived as such by the author). Deposit by library or the publisher may be much more efficient (e.g. through bulk uploads, or because metadata is attached at an earlier stage of the process).

We do not see that as a key determinant of the scenario. Rather we would expect to model the costs of alternative approaches to depositing articles – and then to make an assumption about the mix of author / publisher deposit.

---

<sup>7</sup> <http://creativecommons.org/weblog/entry/13408>

### 3.3. Gold OA Scenario

#### 3.3.1. Introduction and summary

Gold OA as a business model is perceived to have a number of advantages over the current subscription based models:

- lower overall costs (as per RIN / CEPA (2008);
- better scalability with the growing number of research outputs;
- fairer (avoids exclusions, as access to information is increasingly seen as something like a human right); and
- macroeconomic benefits from freely available research.

The issues for this study relate to: (i) the extent of such advantages – some of which are contested; (ii) the mechanics and costs associated with transition to such a scenario.

Due to the difficulties of implementing a full shift to Gold OA (shifting funding flows etc.), we consider that over a five-year time horizon an appropriate scenario will be for pockets of researchers (e.g. by disciplines) to switch to either a full Gold OA or hybrid OA (i.e. where authors pay for their articles to be open within an otherwise subscription-access journal) approach.

A key aspect relating to Gold OA within the scenarios will be the transition of institutions from paying for full subscriptions to author side payments. Over the scenarios' time horizon there is a risk that institutions may, to some degree, pay for both concurrently – unless this dealt with explicitly.

Table 3.2 provides an overview of the shape of a possible base Gold OA scenario. Again, the points are discussed in more detail below.

Table 3.2: Gold OA Scenario Summary

Dimension	Initial view
Driver?	Primary driver to Gold OA is <b>coordination</b> within communities/ funder groups to move to author side payment models.
Scope?	Two scenario variants: (i) the scenario is restricted to ' <b>pockets</b> ' of high Gold OA take-up; (ii) subscription discounts <b>reflecting actual</b> author side payments (i.e. building on 'hybrid' journals)
Access to what?	Focus on ' <b>the version of record</b> '. Presumption that access for <b>re-use</b> of the article (except for hybrid OA) and <b>high functionality</b> provided by the publisher.
Access when?	By definition Gold OA provides <b>immediate access</b> to the version of record at the point of publication.
Access by whom?	Access to <b>all user groups</b> .
Access where?	Access to a Gold OA articles will be via the <b>publisher's site</b> and, secondarily, via a <b>subject or institutional repository</b> .
Access how?	Except for the need to process author-side payments, the submission

### 3.3.2. Dimensions

#### *Drivers*

Despite the perceived advantages, Gold OA (or hybrid) articles to date represent only a small fraction of total journal articles.

It is possible (or indeed probable) that there will be new Gold OA Journals launched by publishers over the five-year horizon. However, the scale of this is unlikely to be sufficient to make material difference.

Our initial view therefore is that a Gold OA scenario needs to involve some degree of coordination to achieve material change. This co-ordination needs to involve publisher(s), funders and research communities. Two approaches to this coordination are discussed under ‘scope’ below:

- the first involving coordination of this kind by community; and
- the second builds on hybrid journals and involves increased coordination amongst subscribing institutions to negotiate direct links between the levels of subscription and author-side payments.

#### *Scope of scenario*

Our initial view is that with a five-year time horizon Gold OA will have **pockets** of high take-up (e.g. bioinformatics etc.), but there would be no significant move to Gold OA across all disciplines. We therefore propose to define a subset of journals / subjects where a move to Gold OA is possible /credible. We might look at this scenario on a UK unilateral basis or as part of a global ‘subject community’ shift (e.g.SCOAP3).

An alternative (or variant) scenario here might involve an extension of the hybrid journal concept. The extension would involve negotiated discounts for groups of subscription holders (e.g. RLUK, or more widely) which reflect actual author-side payments. This approach as the benefit of directly linking author-side payments with reductions in subscription revenue. Such agreements could in principle be negotiated at either an institutional level or through an aggregation mechanism (e.g. NESLI, JISC)

#### *Access to what?*

Gold OA, by definition, provides access to the version of record. We propose to assume that access is for **re-use** (except for hybrid OA) and to **high functionality** provided by the publisher.

We note that a key aspect of the analysis of this scenario (e.g. compared with Green OA scenarios) is the extent to which there are greater benefits associated with ‘full’ access with reuse.

### *Access when?*

By definition Gold OA provides **immediate access** to the version of record at the point of publication.

However, pre-print self-archived versions of the papers may be available earlier through a repository. An issue for the analysis is the extent to which access benefits can be attributed to Gold where it coexists with Green. This will vary by discipline.

### *Access by whom?*

Gold OA provides **all user groups** with electronic access to the version of record.

As with Green OA, the change in access can be expected to vary according to baseline disciplinary differences. However, there may be differences between Green and Gold in terms of cost and ease of search / discovery (e.g. publishers have tended to be better at making articles available for indexing by search engines).

### *Access where?*

Primary access to a Gold OA articles will be via the publisher's site – as part of specific Gold OA journals.

Secondary access will be through subject or institutional repositories. We therefore see the costs of institutional / subject repositories remaining – probably without additional benefit.

### *Access how?*

The current submission process to an OA journal is not notably different from subscription journals, except for the need to process author side payments. We therefore do not see this changing materially over the five year time horizon of the scenarios.

Currently authors typically pay an article processing charge (APC) if the article is published. One option to consider is a two-tiered payment system. For example, a author may pay a submission fee to submit his article; and if the article is published then the author would pay the full APC. This might be part of the scenario (e.g. to support sustainability) but we are not yet clear that we will model it explicitly.

### 3.4. Licensing Scenario

#### 3.4.1. Introduction and summary

A licensing extension is a relatively easy (if not cost free) way of increasing access for user groups, where unmet demand exists.<sup>8</sup> Extending current licensing arrangements could allow publishers to satisfy unmet economic demand. It is also a way of retaining subscriptions while providing access to groups of users.

A possible issue for consideration here will be the extent of any cannibalisation of subscriptions and/ or pay-per-view revenues.

Table 3.3 below provides an overview of a possible licensing scenario.

Table 3.3: Licensing Scenario Summary

Dimension	Initial view
Driver?	The primary drivers are likely to lie with <b>publishers</b> , but depending on how access is achieved may rely on collaboration with libraries or other mediating institutions
Scope?	Disciplines will be determined by defined groups where there is an access gap (see Access by Whom?)
Access to what?	Access to the <b>full version of record, no re-use</b> and <b>high functionality</b> .
Access when?	The presumption for this scenario is that access would be to the licence content at the point at <b>the point of publication</b> . Possible sensitivity for embargoed access
Access by whom?	Scenario likely to involve <b>targeting specific perceived access gaps</b> , such as SMEs or well-defined groups (e.g. professional bodies) – a comprehensive national licence (i.e. covering all UK residents) is conceivable but thought to be implausibly expensive
Access where?	Online, via the publishers' websites. One unresolved question is how access control would be managed cost-effectively for non-institutional users
Access how?	National or other bodies (e.g. JISC Collections, professional associations, British Library, etc.) to negotiate licence extensions.

#### 3.4.2. Dimensions

##### *Drivers*

The drivers for a licensing extensions route to increasing access are likely to lie more with the publishers than users. This might in part be motivated by publishers concern about the sustainability of the other routes to access.

---

<sup>8</sup> There is a potential relationship between licensing and green OA, that is terms may be negotiated that include deposit into a repository. This may affect price.

### *Scope of scenario*

The preferred approach for the scenario here is to target specific categories of users (see Access by whom?). Given this the scope will be determined by the categories of users. By definition this is UK access.

### *Access to what?*

The starting point would be licensed access to the **version of record, no re-use and high functionality**.

There could in principle there could be licenses to access a “lite” version of the article, with reduced or restricted version of the article (e.g. DeepDyve). However, our judgement is that these should be more the focus of a ‘transactional scenario’.

### *Access by whom?*

Although a full national license is theoretically possible, our judgement is that more practical (affordable) options are likely to involve targeting specific perceived access gaps, such as SMEs or well-defined groups (e.g. members of professional associations).

### *Access when?*

The presumption for this scenario is that access would be to the licence content at the point at the **point of publication**. However, it might be appropriate to consider a sensitivity in which access is provided after an embargo period – which would be cheaper.

### *Access where?*

Licensing extensions could allow an article to be accessed either through ‘location’ (walk-in) access or ‘online’ access. Given that walk-in access is only likely to benefit the general public (it is unlikely to be valuable to SMEs) this scenario is expected to focus on **online** access.

### *Access how?*

Negotiations are likely to require a body (such as JISC, British Library, professional associations) to agree appropriate rates/ coverage.

### 3.5. Transactional Scenario

#### 3.5.1. Introduction and summary

A transactional solution provides article access to users without access to a subscription/ licence. However, as Gold OA and to some extent Green OA increase demand for pay per view article will reduce.

Table 3.4 provides an overview of the shape of a possible base transactional scenario. The points are discussed in more detail below. Our initial view is that an appropriate scenario will have publishers targeting transactional solutions on extending access to specific user groups (e.g. SMEs, practitioners, etc.) and this will be coupled with a ‘background rise’ in Green OA/Gold OA.

Table 3.4: Transactional Scenario Summary

Dimension	Initial view
Driver?	Publishers’ desire to <b>meet demand</b> in user groups without access to a subscription/ licence.
Scope?	Focus on increasing access to <b>targeted</b> user groups e.g. SMEs, practitioners
Access to what?	Focus on the <b>version of record</b> . Presumption that access for <b>no re-use</b> of the article, and <b>low functionality</b> .
Access when?	Access will be at the <b>point of publication</b> .
Access by whom?	Focus on targeting user groups with <b>unmet demand</b> .
Access where?	Primary access to articles will be via a <b>aggregating site</b> , which might be an existing document delivery provider (e.g. British Library) or a new entrant (e.g. DeepDyve).
Access how?	Focus on a <b>‘pay per view’</b> model.

#### 3.5.2. Dimensions

##### *Drivers*

Our view is that primary driver of transactional solutions is likely to arise from publishers’ desire to meet unmet demand.

##### *Scope of scenario*

Given that the scenario drivers, the scope of the scenario will be restricted to increasing access to defined user groups without access a subscription/ licence e.g. SMEs. In keeping with the need for the scenario to be credible tractable, we consider that there is merit in restricting the user group access to a subject group (e.g. STM).

##### *Access to what?*

The access provided will be dependent on the transaction model i.e. a ‘rental’ model (e.g. DeepDyve) or an new pay per view model, access to **high functionality** or **low**

**functionality respectively.** Our initial view however, is that a transactional scenario will be based on providing **no re-use** to the **version of record**, but access to additional services available via a publisher's site will not be provided

*Access when?*

Again this is dependent on the model (i.e. there could be differential pricing depending on the age of the content). Our initial view however, is that a transactional scenario will provide full access to the version of record at the **point of publication**.

*Access by whom?*

A transactional route to access would **target** user groups without subscriptions/ licences or researchers needing access to peripheral content.

*Access where?*

Our initial view is that the primary access will be via an **aggregating site** i.e. an iTunes/ Amazon type site.

*Access how?*

We consider that there are four broad transactional models:

- Deepdive type “rental” – limited use i.e. no printing, saving, copying etc;
- Pay per view – but with lower prices, simpler price bands, aggregated content, easier to use etc. (e.g. iTunes model);
- Licence plus usage based charges – combined licence access and usage based charge which could extend access to institutional researchers; and
- low cost time limited access – user would have access to full version/ full functionality, but for a limited time e.g. 24hrs.

Our initial view is that a transactional scenario will focus on a **pay per view model**.

### 3.6. Summary table

Table 3.4: Scenario Summary

Dimension	Green OA	Gold OA	Licensing extension	Transactional
Driver?	<b>Tightening of funder/ institutional mandates</b> in terms of monitoring and enforcement – resulting in significant increases in deposit rate compared with baseline.	Primary driver to Gold OA is <b>coordination</b> within communities/ funder groups to move to author side payment models.	The primary drivers are likely to lie with <b>publishers</b> , but depending on how access is achieved may rely on collaboration with libraries or other mediating institutions	Publishers' desire to <b>meet demand</b> in user groups without access to a subscription/ licence.
Scope?	Scenario restricted to a <b>defined subject grouping</b> on a global or UK only basis, as a credible 'pilot' in the five-year horizon. Sensitivity analysis on the access based on increase in institutional repositories.	Two scenario variants: (i) the scenario is restricted to <b>'pockets'</b> of high Gold OA take-up; (ii) subscription discounts <b>reflecting actual</b> author side payments (i.e. building on 'hybrid' journals)	Disciplines will be determined by defined groups where there is an access gap (see Access by Whom?)	Focus on increasing access to <b>targeted</b> user groups e.g. SMEs, practitioners
Access to what?	Focus on <b>'the accepted manuscript'</b> (i.e. peer reviewed) that are deposited. Initial presumption that access is <b>no re-use</b> and it provides <b>low functionality</b> . Sensitivity on access to the version of record and high functionality.	Focus on <b>'the version of record'</b> . Presumption that access for <b>re-use</b> of the article (except for hybrid OA) and <b>high functionality</b> provided by the publisher.	Access to the <b>full version of record, no re-use</b> and <b>high functionality</b> .	Focus on the <b>version of record</b> . Presumption that access for <b>no re-use</b> of the article, and <b>low functionality</b> .
Access when?	<b>Access after a six month embargo</b> . Sensitivity (or alternative scenario) on access to the accepted manuscript or version of record on an	By definition Gold OA provides <b>immediate access</b> to the version of record at the point of publication.	The presumption for this scenario is that access would be to the licence content at the point at <b>the point of publication</b> . Possible sensitivity	Access will be at the <b>point of publication</b> .

Dimension	Green OA	Gold OA	Licensing extension	Transactional
	immediate basis.		for embargoed access	
Access by whom?	Access to <b>all user groups</b> – although extent of change in access will depend on scope of scenario as well as baseline access in disciplines.	Access to <b>all user groups</b> .	Scenario likely to involve <b>targeting specific perceived access gaps</b> , such as SMEs or well-defined groups (e.g. professional bodies) – a comprehensive national licence (i.e. covering all UK residents) is conceivable but thought to be implausibly expensive	Focus on targeting user groups with <b>unmet demand</b> .
Access where?	Given the ‘funder mandate’ driver and subject grouping, we assume that primary change in access is through <b>subject-based repositories</b> . But institutional repositories expected to continue and develop.	Access to a Gold OA articles will be via the <b>publisher’s site</b> and, secondarily, via a <b>subject</b> or <b>institutional repository</b> .	Online, via the publishers’ websites. One unresolved question is how access control would be managed cost-effectively for non-institutional users	Primary access to articles will be via a <b>aggregating site</b> , which might be an existing document delivery provider (e.g. British Library) or a new entrant (e.g. DeepDyve).
Access how?	Who deposits is not a key determinant of scenario. Requires assumption of publisher / author mix and appropriate costs.	Except for the need to process author-side payments, the submission process <b>will not</b> change.	National or other bodies (e.g. JISC Collections, professional associations, British Library, etc.) to negotiate licence extensions.	Focus on a ‘ <b>pay per view</b> ’ model.

## 4. COMMON ISSUES

In this section we set out some more general issues that will require consideration. This list is not exhaustive, rather it highlights some of the key difficulties that will arise when developing and modelling the scenarios.

At this stage we have not sort to fully work through these points. Rather we note them here as being issues that need to be considered. We would of course welcome any comments / views.

- **E-only vs print.** It is likely that e-only will have replaced print for all sciences within five years. The extent to which e-only has replaced print is a dimension across all scenarios. Sensitivity analysis can be carried out. The scenario will also have to consider the payment of VAT on electronic journals
- **Article processing charge.** OA APCs are currently well below those reported across the STM industry. However, currently OA APCs are only made on a small proportion of the papers published, so it is hard to describe this as a market rate. A range for APC may need to be consider through sensitivity analysis.
- **Full use vs. Free use.** An economic benefit is likely to arise from full use over free use (e.g. reuse), and from the additional functionality provided with a publisher's version of record and website.
- **Multiple access points.** Green OA and Gold OA are not mutually exclusive, however in order to restrict the possible permutations of the scenarios we will consider that if an article is available via Gold OA it will also be self-archived and the costs and benefits for this will exist in the scenario. (Note; There is also a potential benefit from an preprint/ accepted manuscript being available earlier.)
- **'New' business models.** Some of the scenarios, e.g. licensing and transactional, may consider the development of business models for which limited data exists.
- **Enforcement of mandates.** An issue that has been raised in relation to a Green OA route to access is a lack of author 'motivation' to deposit. Is it likely/ feasible that institutions will enforce a mandate to deposit.
- **International dimensions.** A consideration across all the scenarios is the degree to which there are different rates of adoption of the routes to access between the UK and the rest of the world e.g. differential rates of adoption of Green OA as between UK and rest of world. These is also more general increases in research production e.g. China.
- **Diminishing marginal returns.** The Solow-Swan model does not incorporate diminishing marginal returns for article access.
- **Change in research practice.** This may result in a change (probably an increase) in the number of research outputs and a change in their nature (e.g. database publication, multimedia digital objects rather than conventional research papers, etc.).

- **Existing technologies impact.** The internet has had a significant impact on many business models (e.g. the music industry). The scenarios will need to consider how this technology is impacting on the existing business model and those proposed in the scenarios.

## ANNEX A: DEFINITIONS

### Routes to access

- **Full open access (Gold OA).** The final published paper is made available online immediately on publication using a business model in which publication is paid for rather than access. For the purpose of this study we are treating the following as variants of Gold OA:
  - *Immediate full OA:* the entire contents of the journal are made freely available immediately on publication. A well-known example is *PLoS Biology*.
  - *Optional (or hybrid) OA:* here only part of the journal content is made immediately available. The journal offers authors the option to make their article OA in an otherwise subscription-access journal in return for payment of a fee (e.g. Springer's Open Choice or OUP's Oxford Open schemes). This can be seen as a route to convert subscription journals to full open access; some publishers offering this have said that the subscription price will be reduced in line with uptake of the OA option.
- **Open access via self-archiving (Green OA).** The “green” route to open access is by self-archiving, which makes available a version of the article (usually either an author's original or accepted manuscript), either immediately or delayed. Self-archiving has no independent business model, in that it relies on the assumption that making a version of the article freely available will not compromise the sales of the version of record (i.e. journal subscriptions).

The author (or someone acting on their behalf) deposits the article in an open repository. This repository might be an *institutional* repository run by the author's institution (typically a university) or a central subject-based repository (such as arXiv in physics and PubMed Central in biomedicine).

### Article versions

We consider that it is appropriate to limit the scenarios to three article versions:

- **Preprint.** Author's un-refereed draft manuscript for consideration by a journal. (“author's original” or “submitted manuscript under review” using the NISO Versions preferred terms).
- **Accepted manuscript.** Author's final refereed manuscript accepted for publication by a journal and containing all change required as a result of peer review.
- **Version of record.** Final published citable article available from the journal's website.

### Article rights

Access rights describe the rights that the user has in relation to the article. It is determined by both copyright and contract law (in the form of the licence between publisher and library/ purchaser). We also propose to limit the scenarios to two primary options:

- **No re-use.** The typical conventional bundle of rights enjoyed by the user (e.g. to read it, print it out, quote extracts in other publications (“fair dealing”), etc.); and
- **Re-use.** Extensive right to re-use material without permission (*including* commercial re-use) that is typically granted by Open Access journals (typically either the Creative Commons “attribution” licence, CC-BY or one equivalent to it).

### Article functionality

This is a purely online (or at least electronic) distinction and refers to the presentation of the article and the additional embedded utility (for instance in terms of mark-up, linking, embedded semantic content, etc.). We propose to limit the scenarios to two options:

- **Low functionality.** the flat PDF version. You can print it and read it, save it locally, search it but little else. (If it’s not encrypted you can extract content but only with great difficulty, especially for technical content.) We can also consider very low functionality, e.g. where the use is deliberately crippled using DRM etc. to support a particular business model, e.g. DeepDyve’s “rental” model prevents saving, copying and printing.
- **High functionality.** the enhanced HTML version on the publisher’s website, incorporating e.g. two-way updated citation linking, semantic enhancement, rich media content, two-way linking to research datasets and structured databases, etc.

### User groups

A core part of our analysis of scenarios will relate to quantifying changes in access and assessing the impact of this in terms of both costs and benefits. In order to do this we expect to consider the impact of access on different sub-groups of user. These sub-groups include:

- institution-based researchers;
- independent researchers;
- health service;
- other professional services (e.g. engineers, lawyers etc);
- corporate (both large and SME);
- central and local government (and other public sector);
- and voluntary community sector.

This list is not necessarily exhaustive and it is also possible that these user groups will ‘collapse’ somewhat based on our literature review and available data.

### **Repository definitions**

The following repository definitions are taken from the *Report of the subject and Institutional Repositories Interactions Study*, by Catherine Jones, Robert Darby, Linda Gilbert and Simon Lambert (2008).

- ***Institutional repository.*** This is a collection of research outputs with a common link to a particular institution, usually by authorship. These repositories are likely to cover more than one research discipline, to have funders in many if not all the Research Councils and support communities who have different approaches to research dissemination. Whether deposit of content is mandatory is a decision that will be made by each institution. The institutions may have many requirements for the content of the repository, from open access dissemination, through metrics, marketing to strategic planning. It is likely that many of these processes in the past were undertaken through collection of bibliographic information.
- ***Subject repository.*** This is a collection of research outputs with a common link to a particular subject discipline. Subject repositories are likely to cover one broad-based discipline, with contributors from many different institutions supported by a variety of funders; the repositories themselves are likely to be funded from one or more sources within the subject community. Although for some subject repositories the funding may be fragile, if they are of enough importance to the community then funding crises are usually weathered. Deposit of content is voluntary. These repositories are usually concerned with dissemination; for example the emergence of the arXiv repository replacing the practice of circulating paper preprints in the particle physics community.
- ***Funder repository.*** This is a collection of research outputs with a common link to one (or more) funder(s). These are likely to cover the funder’s remit, which is usually subject-based but can become indistinct at subject boundaries, and will have authors from many institutions. Deposit of content is usually mandatory and can include project-related material, such as completion project reports. The funders will have requirements for the content of repository, from metrics through to strategic planning.

## **ANNEX B: OTHER PORTFOLIO PROJECTS**

RIN described four projects which comprise its current and planned portfolio. The Dynamics of improving access project which CEPA is undertaking is one of the four projects. We describe the other three below.

### **4.1.1. Gaps in and barriers to access**

This project has two principal objectives:

- To map barriers to (and gaps in) access to research;
- To identify the key groups who demand (need) access to research, but who are unable to (lack) access to research.

The project will seek to define and articulate the ‘access’ people require, rather than simply identifying gaps in provision. This project hasn’t started yet – an ITT has been sent put and RIN aims to contract a provider in the next couple of weeks. RIN want for the final project deliverables to be made available in October / November 2010.

### **4.1.2. The shift to e-only**

JISC is coordinating this ‘two component’ project, the ITT for which is due to be released shortly, with final deliverable to be made available in September 2010.

This first component of this project will identify and articulate the range of factors constraining a more comprehensive shift towards E-only journals – it will consider:

- how journals are being used;
- why (to date) there hasn’t been a more comprehensive shift towards the use of e-only journals;
- what activities different organizations are undertaking to support a shift to e-only access to research articles; and
- how a more comprehensive shift to e-only access can be achieved (by setting out a ‘roadmap’).

The second component of the JISC-coordinated project will look specifically at the use of e-archiving in the process of achieving a shift to e-only access. RIN emphasised how this study will need to consider possible conflict/contradiction between the views of publishers and libraries in relation to a transition to e-only access to articles.

### **4.1.3. Futures project**

The third major project in the portfolio will seek to identify and articulate possible ‘20-year’ scenarios (and drivers of possible scenarios) – i.e. over a longer-term horizon than the CEPA study, which will look at ‘5-year’ scenarios. The ITT for this study has not yet been drafted, but the intension is for this study to be delivered by the end of FY2010/11.