Exploring the Effects of a Transition to Open Access: Insights from a Simulation Study

Steffen Bernius
Institute for Information Systems
Goethe-University Frankfurt
bernius@wiwi.uni-frankfurt.de

Matthias Hanauske
Frankfurt Institute for Advanced Studies
Goethe-University Frankfurt
hanauske@fias.uni-frankfurt.de

Berndt Dugall
University Library
Goethe-University Frankfurt
b.dugall@ub.uni-frankfurt.de

Wolfgang König
Institute for Information Systems
Goethe-University Frankfurt
wkoenig@wiwi.uni-frankfurt.de

Exploring the Effects of a Transition to Open Access: Insights from a Simulation Study

Abstract

The Open Access (OA) movement, which postulates gratis and unrestricted online access to publicly funded research findings, has significantly gained momentum in recent years. The two ways of achieving OA are self-archiving of scientific work by the authors (Green OA) and publishing in OA journals (Gold OA). But there is still no consensus which model should be supported in particular. The aim of this simulation study is to discover mechanisms and predict developments that may lead to specific outcomes of possible market transformation scenarios. It contributes to theories related to OA by substantiating the argument of a citation advantage of OA articles and by visualizing the mechanisms of a journal system collapsing in the long-term due to the continuation of the serials crisis. The practical contribution of this research stems from the integration of all market players: Decisions regarding potential financial support of OA models can be aligned with our findings – as well as the decision of a publisher to migrate his journals to Gold OA. Our results indicate that for scholarly communication in general, a transition to Green OA combined with a certain level of subscription-based publishing and a migration of few top journals is the most beneficial development.

1 Introduction and research motivation

The Open Access (OA) movement, which postulates gratis and unrestricted online access to publicly funded research findings for anyone, has significantly gained momentum in recent years (Oppenheim, 2008). Besides several hybrid models that are pushed primarily by publishers, two main ways of achieving "true" OA have emerged (for classifications see Bernius et al., 2009; Jubb, 2011). On the one hand, scientific authors have the possibility to deposit their work in institutional or subject-based repositories. This self-archiving is often called "the green road to OA" or simply "Green OA" (Harnad et al., 2004). On the other hand, OA journals are established – either through creation of new titles or through transformation of traditional subscription-based journals. This approach, which is often termed "golden road to OA" ("Gold OA"), shifts the financial model of a journal. Instead of generating revenues through subscription fees, the author (to be more precise: an institution the author belongs to) has to pay a fee if his/her article is accepted for publication.

In some disciplines OA models are already established to a certain degree. For instance, physicists and mathematicians use the arXiv server as central repository for communicating research results (http://www.arxiv.org). In these disciplines pre-prints as well as published versions of articles are normally self-archived in arXiv, allowing any interested reader to download the work via the internet for free. An example for a successful implementation of Gold OA is the journal Living Reviews in Relativity (http://relativity.livingreviews.org/). In the 2009 Journal Citation Reports this OA journal received its first impact factor (10.600) and ranked second in the category Particles and Fields Physics.

However, in the vast majority of disciplines the traditional publishing model is still predominant (Björk et al., 2010).

Nonetheless, the OA idea is strongly supported by all participants of the scientific publishing market – of course, except traditional publishers themselves. First, librarians see OA as way out of the serials crisis: Shrinking or stagnating library budgets with steadily increasing subscription prices have lead and still lead to a great number of title cancellations (Creaser & White, 2008; McGuigan & Russell, 2008). Second, scientists realize that they can benefit from OA in terms of a rapid and wide dissemination of scientific knowledge (e.g., Mann et al., 2009). Unfortunately, they still lack incentives to actually adopt new practices (Hanauske, Bernius & Dugall, 2007; Kim, 2010; Park, 2009). Third, research institutions and foundations advise their researchers to publish their work in OA and often consent to finance necessary steps for a transition of the actual system.

However, there is still no consensus which publication model (or combination of models) is to be preferred *"best"*. On the one hand, some funding organizations pay publication fees for the researchers whom they finance and therefore they support the OA journals model. On the other hand, more and more universities build institutional repositories as platforms for presenting their research output. In this context, the green model is also fostered through mandates that research institutions adopt in order to prompt scientists belonging to these institutions to select OA – especially by depositing papers in the institutional repositories (for an elaborate overview see http://openaccess.eprints.org). Supporting all OA models with rather untargeted subsidies (and perhaps completely disregarding the role subscription-based publishing may still play in future scenarios) carries the risk of investing money and manpower in order to establish models that in the long run turn out to be ineffective.

For instance, the institutional repositories that have spread massively in the last months are in the majority of cases more or less "empty". As Xia and Sun (2007) point out, the archiving of articles is mainly done by librarians or administrative staff. Researchers in most disciplines are far from adopting the self-archiving approach thoroughly (Björk et al., 2010; Creaser et al., 2010; Kim, 2010). On the other hand, if they do so, one can argue that the green model has the potential to undermine the business model on which it relies (Jubb, 2011). Assuming that the necessary (legal and infrastructural) prerequisites for a broad adoption of self-archiving be established, openly accessible articles would lead to a decrease in the publishers' subscription-based revenue. To cover the distribution margin, the prices for the remaining journal licenses then would probably be raised significantly. The danger of sliding into such a vicious circle, which leads to an intensification of the serials crisis, had already been described ten years ago by King and Tenopir (1999). A scenario of "dying" journals is surely not desirable for researchers, who are – at least in terms of quality assurance and reputation indication – used to rely on a journal-based system. In fact, skepticism about quality assurance through peer review and the pressure to publish in well-established high-quality outlets are two main reasons for not adopting OA (Kim 2010, Mann et al., 2009; Park, 2009). However, since it only shifts monetary flows, the OA journals model can be criticized, too. Some experts are afraid that, if adopted broadly, Gold OA could be more expensive than the traditional subscription-based system (Houghton & Oppenheim, 2009). To put it bluntly, who will deter publishers from raising publication fees for single articles in the same way journal prices are raised in the current system (Bernius et al., 2009).

Over the past years the number of articles that consider the impact of OA on scholarly communication has grown exponentially. However, many publications are rather opinion papers than theoretically and methodically sound analyses. Of course, there are exceptions. For instance, the above mentioned studies that analyze scientists' behavior towards OA by applying empirically quantitative data to established theories like the theory of planned behavior (e.g., Mann et al., 2009; Park, 2009). In addition, many studies that examine the influence of OA on citations (overviews on literature on this topic are provided by Swan, 2010; Wagner, 2010; Hitchcock 2011) are also based on quantitative data sets. Not all of these papers report a positive correlation between the accessibility of an article and the citations it receives (Swan, 2010). In particular, the question which mechanisms lead to the proposed citation advantage of OA articles is still heavily discussed (e.g., Craig et al. 2007, Moed 2007). A third important branch of studies related to OA tries to evaluate the cost implications of alternative access-

provisioning strategies. Especially the studies of Houghton and colleagues (Houghton et al., 2009; Houghton & Oppenheim 2010), who conducted a cost-benefit analysis of different models, have attracted attention. Their approach has been broadly appreciated for providing a comprehensive evaluation of the costs of scholarly communication; but it has also been criticized for several shortcomings. Some experts argue that by restricting the analysis to the identification of cost-based advantages of OA, the Houghton report disregards epistemic implications of OA, which derive from the emerging possibility to restructure the whole knowledge economy from scratch (M. Hall, 2010). Bernius (2010a, 2010b) addresses those issues by complementing the cost-based view of Houghton with a theoretical analysis of the impacts of OA on the creation and transfer of scientific knowledge. Such an approach takes into account that "intangible" factors cannot be measured in numbers. The same line of argument is followed by authors who criticize Houghton's approach for being built on empirically unsupported estimates and assumptions (S. Hall, 2010). What these research studies have in common is a rather static view on the influences of OA. Comparing disjunctive worlds of non-OA versus complete-OA neglects the dynamics effects resulting from the interferences between the market players during the transition phase.

This study differs methodically from those described above in that it uses a simulation approach to evaluate the effectiveness of alternative access-provisioning models. This procedural method has several advantages compared to static analyses, which have to rely on "hard" numbers that are often difficult to gather empirically. On the one hand, simulation is an accepted method for predicting future outcomes of dynamic processes like the ongoing transformation of the publishing market. On the other hand, a computational simulation enables exploration of effects deriving from the interplay between different publication models and differences in the behavior of market players – simply by varying input parameters. Based on the model of key processes and mechanisms that determine variances of important market output variables, a comparison of different "publication scenarios" is enabled. To put it in a nutshell, we address the following general research problem: *Does scientific communication benefit from a switch to OA, and if so, which access-provisioning model (or mix of models) should be supported in particular?* This research problem is refined in the next section; answers are suggested from a public perspective (researchers, libraries, funders) as well as from the viewpoint of commercial publishers.

The remainder of this paper is structured as follows: In section 2 we describe four specific research questions derived from the general research problem. Since the approach used in this study is quite complex, we give a detailed specification of the method we used (section 3). Subsequently, the theoretical logic underlying the developed simulation model is rooted in literature (section 4). The model itself, as well as its validation, is presented in sections 5 and 6. After discussing the results (section 7) we conclude with a summary of practical and theoretical contributions of this paper and a brief outlook on further research (section 8).

2 Specific research problems

In the following we specify four research questions, which we take as a basis for investigating the aforementioned general research problem. Since our approach is of an explorative nature, we refrain from formulating concrete hypotheses.

RQ1: What are the long term consequences if the traditional publishing system continues under current conditions? This research question addresses one basic argument of librarians and OA advocates when claiming the need of fundamentally restructuring the scholarly publishing market. All else being equal, a continuing serials crisis (steadily rising journal prices in combination with shrinking or stagnating library budgets) will amount to a collapse of the publishing system. For instance, King and Tenopir (1999) describe the consequences of spiraling journal prices as a vicious circle, in which all system participants lose in the end. We try to shed a light on the effects of the serials crisis on the accessibility and affordability of scientific information – especially by taking into account the mechanisms behind the predicted progress. Furthermore, simulation of this special case serves as basis for model validation: With the boundary conditions remaining unchanged, the final result of this progress

inevitably *has* to be a collapse of the (simulated) traditional publishing system. If the simulation outcome would not reflect this breakdown, a further experimentation with the inclusion of alternative models would be obsolete.

RQ2: How do individual authors benefit from a change in their publication behavior towards OA? Modeling researchers as agents enables the impacts of OA to be analyzed at an individual level. In particular, we investigate the influence of OA on the reputation of authors measured by citations to their work. In doing so, the focus is set on the OA citation advantage that is proclaimed by many OA advocates. Again, it is not our aim to measure such effects in definite numbers, but to explore underlying processes, which are characteristic for scholarly communication networks and which may lead to the expected outcome. Since we have already published a paper that deals with this research question in detail (Bernius & Hanauske 2009), we will only give a brief presentation of the results. Nonetheless, since the effects of OA on citations represent an important issue concerning incentives to encourage scientists to actually publish in OA, the answer to this question is important when the evaluation of the general research problem expressed in the previous section is made.

RQ3: Which access-provisioning model (or combination of models) is the most effective concerning the accessibility of scientific information? The simulation of the emergence of an article network is the central element of our model. In doing so, prediction of how different transition scenarios contribute to the degree of information accessibility is enabled. Furthermore, in this context, we investigate the development of the quality of information available in the network. The latter, for example, becomes an important factor when differentiating potential OA journals with respect to their reputation.

RQ4: With regard to affordability, which access-provisioning model (or combination of models) is economically efficient? Green OA was first and foremost pushed with the aim to solve the accessibility problem, i.e. providing a researcher with access to scientific articles published in journals that his/her library does not subscribe to. In contrast, advocates of Gold OA claimed that fundamental restructuring the revenue stream for scholarly publishing via a broad adoption of the OA journals model would be the answer to both the accessibility and the affordability problem. By focusing on the trade-off between increasing access and simultaneously increasing expenditures we try to explore the results of a transition to OA in terms of cost effectiveness. Predicting financial effects of possible scenarios is especially interesting against the background of diverging target functions of the different market players.

3 Method

Robert Axelrod (1997) is right when he indicates the difficulties to provide a complete description of a social science simulation model briefly. On the other hand, Davis, Eisenhardt and Bingham (2007) argue that *"a high-quality method includes justification for using simulation for the research question at hand and using a simulation approach that fits the research"* (p. 496). Thus, before the presentation of the simulation model, we want to explain why it is useful for our study to choose simulation as our research method and to draw on agent-based modeling as a simulation approach. Furthermore we characterize our approach in detail by explaining the techniques we use for data gathering, experimentation, and validation.

3.1 Why use a simulation approach?

As outlined in the introduction, we want to analyze how a transition to OA impacts the scientific publishing system. Thus, the object of study is the market of scientific publications, which consists of three main groups of market participants. Scientists are the producers and consumers of information. Publishers act as intermediaries and thus collect, bundle and sell articles in the form of journals mostly to academic libraries. The libraries in turn provide the information to the scientists (Odlyzko, 1997; Fladung, 2007). Central market mechanisms are the circulation and citation of papers. Other important variables are the reputation of authors and journals, the price of a journal, and journal usage (Bernius et al. 2009). The micro-level interplay of these components is too complex to predict market outcomes

on a macro-level by simply drawing on systems of mathematical equations (Gilbert & Troitzsch, 2005). In comparison, with the use of a simulation approach emerging properties of a system can be clearly revealed – especially when the theoretical focus is nonlinear and longitudinal as in our case (Repenning, 2002; Davis, Eisenhardt & Bingham, 2007).

By some authors simulation is seen as a third way of doing science – besides the two standard methods induction and deduction (Axelrod, 2007). It can be defined as a method for using computer software to model the operation of "real-world" processes, systems, or events (Law & Kelton, 1991). Simulation models are often denoted as experiments of thought (Axelrod, 2007) or virtual experiments (Carley, 2001), which require the creation of a computational representation of the underlying theoretical logic that connects the involved constructs together within a simplified world. Extensive experimentation is then enabled simply by changing the software code (Davis, Eisenhardt & Bingham, 2007). Furthermore, simulations are particularly useful for examining complex models: If derivations cannot be carried out because of analytical intractability, with a simulation the consequences of formal models can be tested computationally (Harrison et al., 2007).

3.2 Specification of the method

Dooley (2002) differentiates between three main *schools of simulation* practice: discrete-event simulation, system dynamics, and agent-based simulation. We selected the last for our approach. We sort the market participants (scientists, publishers/journals, libraries) into groups of agents that act under specific rules. As Axtell (2000) explains, agent-based models are especially useful for modeling social processes where social networks matter. Whereas these are to account for mathematically only in highly stylized ways, agent-based models allow for agent interactions mediated by networks. In this context, another typical characteristic of agent-based simulations is the fact that they have more than one level. Interactions of the agents on the lowest level result in the behavior of the system (the market) on a higher level (Marks, 2007). Thus, the behavior of the social system is not modeled directly; rather, it emerges from non-linear interactions among agents with specific rules of engagement (Chaturvedi et al., 2005; Harrison et al., 2007). For instance, the evolving citation network, which is a central component of our model (see next section), could – in its basic form – be modeled with few lines of code. But, embedded in a market simulation with different groups of agents the network *must* emerge from the interactions between the agent groups – that implies much more complex programming. On the other hand, comparing the simulated network structure with real-world structures of citation networks then provides a great possibility for operational validation, i.e. determining whether the simulation model's output behavior has the accuracy required for the model's intended purpose (Sargent, 2004). The problem of modeling emerging complex networks without putting too much "direct causality" in the program code is, for instance, discussed in a paper by Pujol et al. (2005).

Regarding the *characteristics of the modeling target* our approach can be described as case-based. Boero and Squazzoni (2005) differentiate case-based models (the target is a specific empirical phenomenon with a circumscribed space-time nature) from typifications (the target is a specific class of empirical phenomena that share some idealized properties) and theoretical abstractions (the target is a wide range of general phenomena with no direct reference to reality). A feature of case-based models is the theoretically thick representation: They are often built upon pieces of theoretical evidence and well known theories (Boero & Squazzoni, 2005). For instance, an important component of our simulation model is the well-investigated mechanism of preferential attachment, which plays a critical role at citation network emergence (see section 4.3). Moreover, simulating the scientific publication market is not a typical case, simply because of the size of the target domain. Hence, the argument of Ragin (1987) that case-based models aim at "appreciating complexity" rather than at "achieving generality" is particularly true for our case.

Besides other *goals of simulation* like training, education, entertainment and performance, the value of a simulation as a scientific methodology lies in prediction, discovery, explanation/understanding, and proof (e.g., Gilbert & Troitzsch, 2005; Axelrod, 2007). The aim of this present study is to analyze the economic effects of OA in different experimental publishing scenarios. Furthermore, in doing so, we

expect to discover relationships between the behavior of the market participants and market outcomes. Thus, the main goals can be characterized as prediction and discovery.

When conducting a simulation study, special emphasis must be laid on *model validation*. According to Sargent (2004) there are four basic approaches of validating a simulation model: validation by the development team itself, by a third party, by the user(s) of a simulation, or with scoring models seldom used in practice. In our study we resort to the first two alternatives. During the development phase of the simulation we continuously tested the program and the validity of the outcomes by ourselves. On the other hand we were able to draw on independent and anonymous reviews of interim reports we had to deliver to the project funders. In these reports we tried to give detailed descriptions of the theoretical grounding and the conceptual simulation model, so that the evaluators could make helpful comments on our approach.

In order to calibrate the computational representation and to validate the findings, social science agent-based models need to be embedded into *empirical data*. Referring to a classification of Boero and Squazzoni (2005), strategies to gather those data can be direct (taken directly from the target) and indirect (empirical data already available in the field). Conducting our research we mixed up both types of strategies. On the one hand, we used second hand quantitative empirical data to determine realistic ranges of input variables and ratios amongst different parameters. On the other hand, we spoke with involved stakeholders (e.g., librarians, scientists) in order to make use of their knowledge about agents and their rules of behavior. With the objective of testing the computational accuracy of the simulation (*internal validity*) we made use of several techniques appropriate for this purpose: verification of the software coding, sensitivity analysis and comparing simulation results with propositions of underlying theory (see Davis, Eisenhardt & Bingham, 2007). Verification at this stage is primarily concerned with determining that the computer model has been programmed correctly and the simulation functions properly (Sargent, 2004). For determining *external validity* (i.e., comparing empirical data with simulation model's output behavior) we mainly resorted to the most commonly used approach of graphical comparison (Sargent, 2004). Because of insufficient quantity of system data for most input and output measures, statistical results derived from hypothesis tests or confidence intervals would not be relevant. We address issues of verification and validation in more detail in section 6.

As stated above, one main reason why we use a simulation approach is that it enables experimentation. Davis, Eisenhardt and Bingham (2007) identify four – to some degree overlapping – approaches to *experimentation* from which we make use of two, namely varying the value of constructs that initially were held constant, and adding new features to the computational representation. For example, allowing the agent type "authors" to make their articles openly accessible is one additional feature that is not "enabled" in the baseline scenario of traditional publishing. The different experimentation scenarios are focused in section 7.1.

The last characteristic of our approach we want to address is the *programming language* we used. Because the simulated system is rather complex, we chose the general purpose programming language Java. While the use of a special-purpose simulation-software product reduces programming time and eases the verification of computational accuracy, a higher level programming language has the advantages of greater program control and flexibility (Law, 2003; Sargent, 2004). In addition, Java supports object-oriented design, which is commonly used in the construction of agent-based models (Epstein & Axtell, 1996). Table 1 summarizes the specification of our approach.

Table 1: Characterization of research approach.

Criteria	Specification (*highlighted: characteristics of our approach*)			
Research method *(Axelrod, 2005)*	Deduction	Induction	***Simulation***	
Simulation school *(Dooley, 2002)*	Discrete event	System Dynamics	***Agent-based***	
Type of model *(Boero & Squazzoni, 2005)*	***Case-based***	Typification	Theoretical abstraction	
Goal of simulation *(Gilbert & Troitzsch, 2005; Axelrod, 2007)*	***Prediction***	***Discovery***	Proof	Explanation
Basic approaches of simulation model validation *(Sargent, 2004)*	***Validation by development team itself***	***Validation by third party***	Validation by user(s) of the simulation	Scoring model
Strategies to gather empirical data *(Boero & Squazzoni, 2005)*	Experiments	***Stakeholders approach***	Other direct qualitative or quantitative methods	***Second hand qualitative or quantitative data***
Verification of computational accuracy (internal validity) *(Davis et al., 2007)*	***Comparing simulation results with propositions of theory***	***Sensitivity analysis***	***Verification of software coding***	
Validation with empirical data (external validity) *(Sargent, 2004; Davis et al., 2007)*	***Graphical comparison of data***	Confidence intervals	Hypothesis tests	
Ways of experimentation *(Davis et al., 2007)*	***Varying the value of constructs***	Unpacking key theoretical constructs	Varying assumptions	***Adding new features***

4 Theoretical grounding

A main goal of academic research is the diffusion of new research results. This is achieved by interaction between scientists in terms of reading and citing other authors' work. By this, complex citation networks emerge, which are the central element of our simulation approach. In this section we will focus on general characteristics of citation networks, but we address theoretical and empirical work towards citation networks and associated phenomena, which we use as basis for creating the model's processes (see section 5) and subsequently for validating simulation results (see section 6).

4.1 Structure of citation networks

In a citation network the vertices represent the articles, whereas the edges between the vertices depict the citations from one article by another. Three characteristics are typical for citation networks and differentiate them from other networks like the WWW or co-authorship networks (Leicht et al., 2007).

First, the edges of the network are directed. Second, the network evolves over time, which means that with every new article published a new vertex (together with the corresponding link) is added to the network. Thus, citation networks underlie a constant growth. Finally, the network is acyclic. Because a new article can only cite previously existing articles/vertices, there are no loops (citations of forthcoming papers are rare and therefore negligible). Figure 1 schematizes the citation network with inclusion of the vertex types „journals“ and „authors“, which in turn are assigned to specific articles. Some journals and authors are displayed with a "closed edge" that starts and ends in the same vertex. These edges represent self-citations.

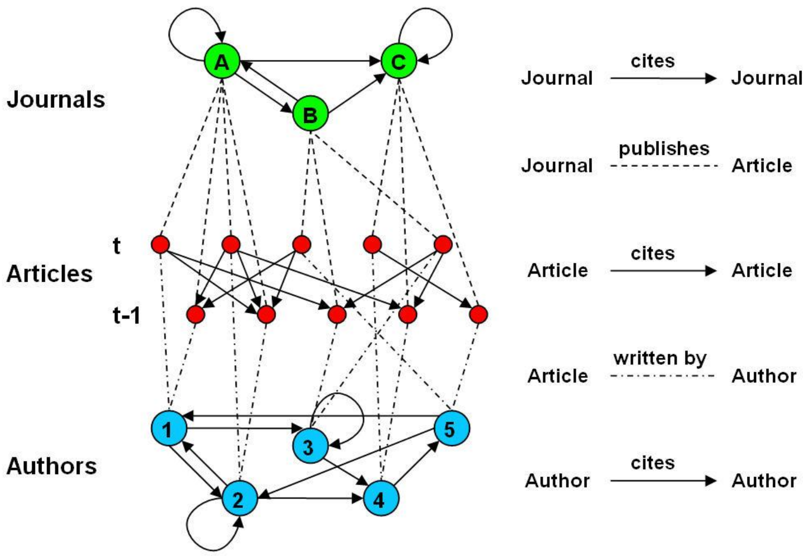

Figure 1: Schematic model of a citation network (Bernius & Hanauske, 2009).

A central measure determining the topology of the citation network is the distribution function of the ingoing citation degrees ($P(k^{in})$). Be $N(k^{in})$ the number of article nodes with k^{in} ingoing citations at time t, then $P(k^{in})$ can be defined as

$$P(k^{in}) = \frac{N(k^{in})}{\sum_{k^{in}=1}^{k^{in}_{max}} N(k^{in})}$$

In this case k^{in}_{max} is the highest realized degree of ingoing citations of an article node at time t. ($k^{in}_{max} \leq I * t$).

4.2 Empirical findings concerning citation networks

Within the scope of the analysis of real citation networks, several authors independently found a typical form of the citation distribution (Shockley, 1957; Redner, 1998; Laherre & Sonette, 1998; Lehmann, Lautrup & Jackson, 2003). Figure 2 depicts the findings of Redner (1998), who investigated two different citation networks, namely 783,339 papers listed in the ISI database in 1981 and cited between 1981 and 1997, and 24,296 papers which were published in volumes 11 through 50 in the journal Physical Review D. The citation distributions of the two networks (Δ= ISI database, o= Physical Review D) are visualized in double logarithmical representation, with the x-axis displaying the number of citations (x), and N(x) on the y-axis depicting the number of articles with x citations. What is remarkable is the asymptotic "long tail" of the citation distribution, which is caused by the fact that the majority of articles are cited only once or not at all, whereas only a few papers are cited frequently.

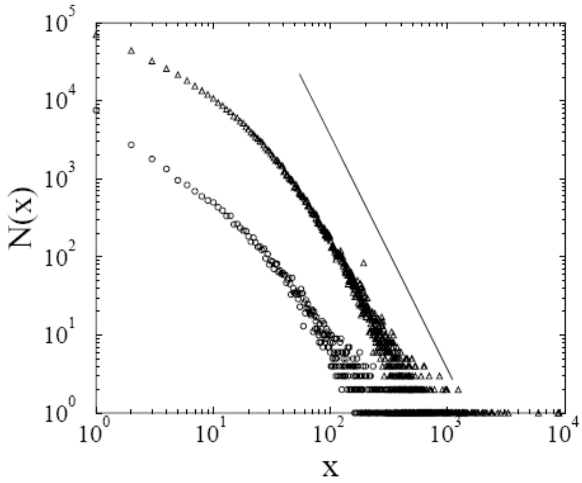

Figure 2: Distributions of real citation networks (Redner, 1998).

Although the distribution functions of different citation networks look almost the same, scholars slightly differ in interpreting the exact analytical expression of the function. However, they agree with each other on the point that a large part of the citation distribution (particularly the asymptotic tail) can be described by a power-law of the form $N(x) \sim x^{-\alpha}$, with the exponent α close to 3. Thus, for visual reference, a straight line of slope -3 is also displayed in Figure 2 (Redner, 1998). Those empirical results represent the most important data source for validating the simulation model. Hence, in section 6.2 we compare the citation distribution of real citation networks with the characteristics of the network emerging through the simulation.

4.3 The mechanism of preferential attachment

During the evolution of the network the mechanism of "preferential attachment" plays a critical role. This dynamic phenomenon was introduced for the first time under the name of "Matthew effect" by the sociologist Robert K. Merton (1968). Based on a quote in the biblical Gospel of Matthew ("For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath") Merton described the effect that an eminent scientist will usually get more credit for similar work than a comparatively unknown researcher. Applied to a network, the phenomenon that "the rich get richer and the poor get poorer" means that a new vertex will be linked with a higher probability to those existing vertices, which already have many ingoing edges (Hein, Schwind & König, 2006). In other words, an article which has, at a particular time t, more citations than a comparable article in the network, also has a higher probability to be cited again in the future. Cole and Cole (1973) denote this characteristic of literature networks as "cumulative advantage".

The emergence of power-law networks based on the mechanism of preferential attachment was primarily modeled by the physicists Barabasi & Albert (1999). They described how in a growing network a power-law degree distribution emerges, when new vertices/players that enter the network prefer to relate to those existing vertices, which already have a high degree of ingoing edges. As Pujol et al. (2005) point out, the model of Barabasi and Albert (and the wide range of subsequent models, which were inspired by their work) has mainly been formulated in the tradition of network theories based on graph theoretic concepts. They consider the individual agent as a mindless player ("node") in the game and disregard the explanation of the emergence of power-law structures as a consequence of plausible sociological micro-processes. In the simulation model presented below, preferential attachment is not explicitly modeled, but results indirectly from the behavior of the individual agents, which – in contrast to the models described above – do not have complete information about the structure of the whole network.

4.4 Citation probability of an article

The probability of an article getting cited is not only determined by its degree of ingoing edges (k), but also by its age (τ). In literature, approaches to model the citation probability commonly use a product separation between the degree-dependent and the time-dependent part:

$$\Pi(\mathrm{k}, \tau) = \mathrm{h}(\mathrm{k}) * \mathrm{T}(\tau).$$

The preferential attachment is modeled by the function h(k), which is usually assumed to be linear. Concerning the time-dependent term T(τ), Hajra & Sen (2004), who analyzed three different sets of citation data (the hep-th and cond-mat sections of the arXiv server and the journal Physical Review Letters), found the following function:

$$T(\tau) = \begin{cases} \tau^{-0.9} & \forall\, \tau \leq 10\ years \\ \tau^{-2.0} & \forall\, \tau > 10\ years \end{cases}$$

Their results provide several insights. First, they show that the choice of a power law is indeed reasonable for citation networks. Second, they observe a crossover in the value of the exponent at ten years, meaning that the majority of papers have a fair chance of citation within ten years after publication, while papers older than ten years are then relatively rarely cited (Hajra & Sen, 2004). In the simulation program we draw on these empirical results when modeling the citation phase. However, by initializing the preferential attachment mechanism through attaching a certain quality to every article entering the network we factor another important measure into the function (see section 5.1). Authors try to read those articles with a relatively high quality. Since only papers which have been read can be cited, the quality aspect is also a determinant of citation probability – just settled on the previous stages of producing and reading scientific work. By doing so we ensure furthermore that preferential attachment is not directly modeled.

5 Simulation model

In the previous sections we described the theoretical basis of our simulation approach. In this section we focus on the conceptual model and its computational representation. This involves the operationalization of theoretical constructs, building algorithms of agent behavior and specifying assumptions (Davis, Eisenhardt & Bingham, 2007). When creating the simulation we followed the building block approach that amounts to increasing complexity stepwise (Harrison et al., 2007). With the overall aim to build a simulation of the whole publishing system including interactions of the agent groups authors, journals/publishers and libraries, we started first of all with modeling processes that lead to the emergence of an article network. The central agents at this stage are the authors as producers of scientific articles. The next step included the creation of a citation network. For this, algorithms were implemented that enable the authors to cite a certain number of articles from of the network. To be as close as possible to reality, articles in the system have to be distinguishable in terms of accessibility (if an author has no access to a specific article he/she is not able to read and thus cite it) and quality (e.g., authors are more likely to cite articles published in high quality journals). At this point, journals and libraries were included as agent types. Those agents also play important roles with regard to the final step of modeling framework conditions like journal prices and system costs.

As Balci (1989) argues, for a complex system with stochastic elements, where many events take place simultaneously and influence each other, decomposing the system into subsystems (or blocks) helps to overcome complexity and thus facilitates model verification. Figure 3 gives an overview of the phases that are run through during the simulation. We will describe them in detail in the following subsections.

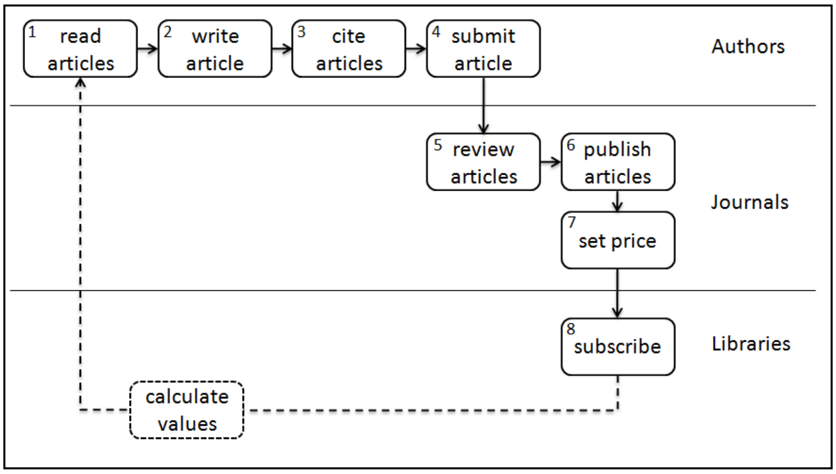

Figure 3: Simulation phases.

5.1 Modeling author's behavior in order to initiate network emergence

Authors are the central agent group in our model. The behavior of the individual authors and the interactions within this group has a strong influence on the characteristics of the emerging citation network. At the authors' level the production process of a new article starts with reading a specific number of papers (*phase 1*). In this context, we assume that authors can only read those articles that are published in a journal licensed by the author's library or that are archived in an OA repository. The latter alternative comes into play, when experimentation is conducted: we then split the author group into traditional authors and OA authors, who self-archive their papers simultaneously to the publication in a journal (see section 7.1). The number of articles an author reads in a simulation period (np_i) is restricted. In the simulation program the reading behavior of scientists is kept rather simple: When selecting articles to read, they orientate themselves towards the reputation of the journals (r_j). In this context, it is assumed that authors prefer to read articles of those journals that have a relatively higher reputation than other journals (i.e., which publish papers of higher quality). The probability that author i selects a paper from journal j is calculated as follows:

$$P_j^R = \frac{r_j}{\sum_{j=1}^{J} r_j}$$

For instance, in a system consisting of 100 authors and 3 journals with the reputation values of r_1=1, r_2=3, and r_3=6, on average 10 authors want to read journal 1, 30 authors prefer journal 2, and 60 authors select a paper from journal 3. After selecting a journal, the author tries to read one randomly chosen article. If his/her library has licensed the journal or the article is published by an author who offers OA, he/she reads the paper and proceeds with selecting another journal. The reading process ends, when the reading capacity of the author (np_i) is depleted.

It is assumed that every author writes exactly one article per simulation period (*phase 2*). Thus, in every period the number of articles in the network increases by the number of authors in the network (n=I). In consequence, the citation network also grows constantly over time. An important variable for the following simulation steps (especially regarding the submission phase; see below) is the quality of a new written article (q_a). Besides the fact that an article *must have* a certain quality, it is furthermore important that articles vary in quality. On the one hand, it would be unrealistic to assume that individual authors produce articles of a constant quality over time. On the other hand, article qualities vary between authors – only on rare occasions do different authors produce articles of exactly the same quality. In reality the article quality depends on many factors and can therefore not be approximated by a complete deterministic equation. In the simulation model we solve this problem by implementing a stochastic differential equation that includes a deterministic term as well as random elements. In

real-world applications this approach is often applied to systems that comprise stochastic, unpredictable factors (e.g., stock market fluctuations). The equation follows a geometric Brownian motion:

$$dq_{a,t} = \mu_i * q_{a,t} * dt + \sigma_i * q_{a,t} * dW_t$$

The percentage drift μ_i represents the average slope of the article quality of author i per time unit dt (deterministic part). The percentage volatility of author i (σ_i) specifies the magnitude of the quality fluctuations of the produced articles, which are modeled with the use of a Wiener process (dW_t) (stochastic part). Contrary to other systems that are subject to stochastic variations, in our case it is not possible to determine the free parameters of the equation (μ_i, σ_i) on the basis of historical averages: the quality of a scientific article is not an empirically ascertainable variable. Modeling the progress of article qualities over time on the basis of the described equation simply serves the purpose of implementing a diversity of article (and therefore author) quality in the simulation program.

Because the simulation proceeds in discrete time steps ($dt \rightarrow \Delta t = 1$), the equation can be simplified as follows:

$$\Delta q_{a,t} = q_{a,t+1} - q_{a,t} = \mu_i * q_{a,t} * \Delta t + \sigma_i * q_{a,t} * \Delta W_t, \qquad \Delta W_t = \varepsilon * \sqrt{\Delta t}$$

$$\Leftrightarrow q_{a,t+1} = q_{a,t} * (1 + \mu_i + \sigma_i * \varepsilon)$$

ε is a normally distributed random number. The quality of an article written at time period t+1 ($q_{a,t+1}$) results of the quality of the article the author has written in the previous period, weighted with factor ($1+ \mu_i + \sigma_i + \varepsilon$), which contains deterministic as well as stochastic elements. Figure 4 visualizes article qualities of different authors in the simulation program as they evolve over time. In this example, 30 authors were generated, the simulation ends at period t=50, and the parameter values were fixed at μ_i = 0.005 and σ_i = 0.05.

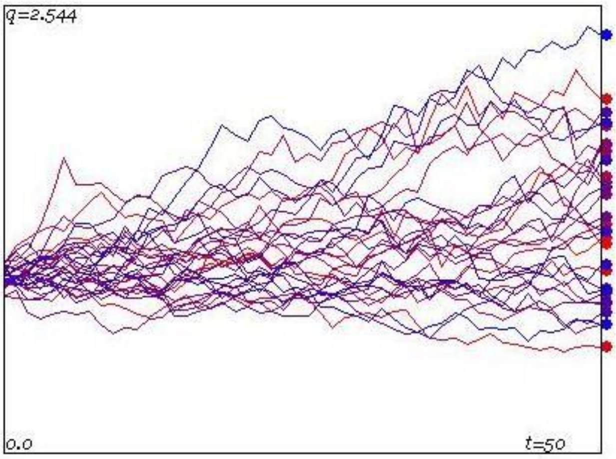

Figure 4: Typical behavior of article qualities from different authors during a simulation run.

During the elaboration of writing a paper, an author cites papers of other authors (*phase 3*). In this context, we assume that authors only cite those articles they actually have read (in the current or in previous periods). From the papers an author has read in the current period, he/she chooses nc^{new} articles dependent on their quality. Those with the highest quality are cited first. From the papers an author has read in the past, he/she chooses nc^{old} articles dependent on their citation probability (Π). Thus the total number of articles cited by an individual author in one simulation period is $nc = nc^{new} + nc^{old}$. As described in section 4.4, the probability of an article to be cited mainly depends on its age (older articles normally get fewer citations) and on the number of citations it has received in the past compared with other articles (preferential attachment). Hence, when modeling the citation probability of an

article we draw on the functional form suggested by Hajra and Sen (2004), which can be generally expressed as follows:

$$\Pi(k_a, \tau_a) = k_a * T(\tau_a) \qquad with\, T(\tau_a) = \begin{cases} \tau_a^{-\alpha_1} & \forall\, \tau_a \leq TC \\ \tau_a^{-\alpha_2} & \forall\, \tau_a > TC \end{cases}$$

The number of received citations of an article is denoted as k_a, τ_a is the age of the article, and $T(\tau_a)$ is a piecewise defined function that determines the curve behavior. TC represents the crossover in the value of the exponent. Papers of age ≥ TC are only rarely cited. However, parameterization revealed that we had to slightly vary the parameter values found by Hajra and Sen. This is due to the facts that the simulated article network is of smaller size than the examined real-world network, and that a simulation period cannot be equated with the time span of one year in real time.

5.2 The publishing decision

The last phase at the authors' level is the submission of the new paper to an appropriate journal (*phase 4*). We model the decision which journal to choose for submitting a new paper with a game-tree-motivated algorithm. Depending on the quality of his/her article, the review decision (accept vs. reject) of the author's previous article, and the reputation of the journal, we identify four cases:

- If in period t-1 the article of author i was accepted for publication and the quality of the article written in the current period is higher, then he/she submits it to the next higher ranked journal (case 1).
- If in period t-1 the article of author i was accepted, but the quality of the new article is lower, then he/she submits it to the same journal as in the previous period (case 2).
- If in period t-1 the article of author i was rejected and the quality of the article written in the current period is higher, then he/she submits it to the same journal as in the previous period (case 3).
- If in period t-1 the article of author i was rejected and the quality of the new article is lower, then he/she submits it to the next lower ranked journal (case 4).

In this context, two assumptions have to be made. First, the author can determine the quality of his papers, and second, he/she is familiar with the reputation of all journals. During the simulation, when an author agent is determined to publish in OA, he/she additionally deposits every written article (rejected or accepted) in a repository, which can be accessed without restrictions by other authors in the system.

The agent group of journals comes into play when the review phase starts (*phase 5*). The journals accept a given number of articles (na_j). If the number of submitted papers is below this "issue capacity", all articles are accepted. If the number of submitted articles exceeds na_j, the journal chooses those articles for publication with the highest quality until the actual issue is complete. At this point, in order to avoid too much technical complexity in the simulation program, the following assumptions are made: the review process is accomplished within one period, the accepted articles are published in the period of acceptance (*phase 6*), and rejected articles cannot be resubmitted. Furthermore, it is assumed that the journals can determine the quality of an article exactly and try to fill every issue independent from the quality of the articles.

5.3 Interaction between journals and libraries

The pricing policy of publishers is a frequently discussed issue – especially in connection with advocating OA models. As depicted in the introduction, the so called "serials crisis" has been a main incentive for fostering OA. In their analysis of the cost implications of scholarly publishing King and Tenopir (1999) describe the effects of common pricing strategies enforced by the majority of commercial publishers. Because of the high specificity of their products they act like monopolists and try to absorb the maximum price. If subscriptions drop, the article production cost per subscriber increases and pub-

lishers raise prices in order to avoid decreasing contribution margins. In turn, this leads to cancellation of titles again. Despite not specifically taking OA into account – in 1999 the focus was more on effects of electronic publishing in general – the approach of King and Tenopir (1999), who used quantitative examples to demonstrate the cyclical effect of circulation, cost, and journal prices, can be taken as the basis for modeling the flaws of a continuance of the still predominant subscription-based system. The vicious circle does not end until the number of subscriptions falls below a critical mass. Since the prices for journals are still steadily increasing, this turning point has not been reached yet.

Whereas, in reality the pricing strategy can be described as "increase constantly", we implemented more economically worthwhile options – simply because of the fact that during the run of the simulation with the empirically observable strategies described above, the simulated market collapses after a certain time (see section 7.2). One the one hand, this important result supports the theory of King and Tenopir (1999) and provides evidence for the validity of our model, but on the other hand, showing that the scientific publishing market suffers from current pricing policies was not the main objective of our study. Hence, in order to analyze implications of alternative models over time, we provided the publishers with a pricing policy, which also allows for decreasing journal prices.

In the pricing stage of the simulation (*phase 7*), the publishers set the price of each journal ($Price_j(t)$), dependent on the number of libraries that subscribed (or unsubscribed) to the journal in the previous period (t = t-1):

$$\sum_{l=1}^{L}(\delta_{j,l}(t-1)) \quad with \begin{cases} \delta_{j,l} = 1, if\ library\ l\ subscribed\ to\ journal\ j\ in\ period\ t-1 \\ \delta_{j,l} = 0, if\ library\ l\ did\ not\ subscribe\ to\ journal\ j\ in\ period\ t-1 \end{cases}$$

The publishers increase the journals` price by the percentage Δ_{Price}, if no library had cancelled the subscription. Furthermore it is assumed that if over 50% of the libraries dropped a journal, its price decreases. If less than half of the libraries cancelled subscription to a journal, the price increases by a value below Δ_{Price}. We model the resultant pricing function as follows:

$$Price_j(t) = Price_j(t-1)\left(1 + \Delta_{Price} + 2\,\frac{\sum_{l=1}^{L}(\delta_{j,l}(t-1)) - L}{L}\Delta_{Price}\right).$$

In each simulation period, the libraries decide whether they subscribe to a specific journal or not (*phase 8*). The assumption that libraries are able to cancel journal subscriptions "immediately" is necessary in order to prevent the simulation program from getting too complex. Despite the rise of electronic journals, in reality libraries are still bound by mid-term contracts with publishers (so called "big deals"), which interdict the short-term cancellation of titles (Fladung, 2007). In our model, the libraries have a fixed budget per period, which they can use to subscribe to journals. No distinction is made between print and online journals.

As a recent survey of librarians on factors of journal cancellations (Ware, 2006) reveals, the subscription decision mainly depends on usage and price (we do not integrate the third important factor that the faculty simply may no longer require a journal). A library subscribes to those journals that on the one hand show the highest usage/price ratios and on the other hand do not exceed a certain price-cap (e.g., to avoid the case of subscribing to only one journal, which has the highest usage and therefore increases its price excessively). In the simulation model this price-cap is oriented towards the total budget of the libraries. The libraries subscribe to journals until the budget is exhausted. After the subscription phase output variables are calculated and a new simulation period starts.

5.4 Specification of variables and assumptions

Table 2 summarizes the parameters and variables used in the simulation model. We denote those variables that are fixed at a particular value during experimentation as parameters (Whicker & Sigelman, 1991). Valid ranges of those parameter values have been found through exhaustive sensitivity analysis

and calibration (see section 6). Besides the input variables (that are varied during experimentation) and the output variables, another type is listed in the table. That type includes variables that cannot be controlled independently, but are affected by the settings of the independent variables and parameters. According to Barton (2004) we denote them as intermediate variables. The third column lists the values of parameters, which were the basis for the simulation runs we conducted throughout experimentation. The input variables are represented by bit strings. Dependent on the strategies chosen for authors, libraries, and journals, they take the values 0 or 1.

Table 2: Parameters and variables of the simulation.

Symbol	Description	Values
Parameters (fixed values)		
T	simulation duration	150
I	number of authors	200
J	number of journals	8
L	number of libraries	3
σ_i	volatility of the article quality of author i	0.05
μ_i	average increase of article quality of author i	0.005
na_i	reading capacity per period of author i	5
nc_i	citation capacity per period of author i	10
nc^{new}	citation rate of read papers of period t	2
nc^{old}	citation rate of papers read in past periods	8
α_1	exponent in time-dependent term of citation probability function	1
α_2	exponent in time-dependent term of citation probability function	1.2
TC	crossover time of citation distribution function	20
na_j	issue capacity of journal j	12
Δ_{Price}	percentage of price increase per period	0.05
$B_{l,t=0}$	starting budget of library l	1000
Input variables (independent variables)		
$PubStrat_{i,t}$	publishing strategy of author i at time t	0= traditional non-OA 1= OA (self-archiving)
$PubStrat_{j,t}$	publishing strategy of journal j at time t	0= subscription-based 1= OA publishing (publication fees)
$SubStrat_{l,t}$	subscription strategy of library l at time t	0 = no additional budget available 1 = additional budget for OA journals
Intermediate variables		
q_a	quality of article a	
$k_{a,t}$	cumulated citations of article a at time t	
$\tau_{a,t}$	age of article a at time t	
$\Pi_{a,t}$	citation probability of article a at time t	
$P^R_{j,t}$	probability to read journal j at time t	
$r_{j,t}$	reputation of journal j at time t	
$\Delta_{j,l}$	Status of subscription of library l to journal j at time t-1	
$b_{l,t}$	budget of library l at time t (=$B_{l,t=0}$ if budget is constant)	
Output variables (dependent variables)		
$C_{i,t}$	cumulated citations of author i at time t	
$c_{j,t}$	citations of journal j at time t	
$Price_{j,t}$	price of journal j at time t	
$OAfee_{j,t}$	publication fee of (OA) journal j at time t (see section 7.4)	
$U_{j,t}$	usage of journal j at time t (= number of articles accessed via this journal)	
$Access_{a,t}$	accessibility of article a at time t	

Table 3 summarizes the assumptions specified within the computational representation. We distinguish between general assumptions that relate to boundary or scope conditions of the theory, and technical assumptions that are simplifications of the simulation itself and reduce unnecessary complexity (Davis, Eisenhardt & Bingham, 2007).

Table 3: Summary of model assumptions.

General assumptions	• when reading articles authors prefer journals with relatively high reputation • the quality of an article depends on the quality of the articles previously written by the author • journals' decision to publish an article depends only on the article's quality • reputation of journals depends only on the quality of the articles published • no distinction between print and online journals
"Technical" assumptions	• every author reads a certain number of articles per period • authors can only read those papers, which are OA or which are published in journals licensed by their library • authors write exactly one article per period • only single author papers • every new article cites a certain number of other articles • authors know the quality of their own articles (better/worse than other article) • authors know the reputation/quality hierarchy of the journals • every journal releases one issue per period • every journal has a certain capacity of articles per issue • the article's quality can be exactly determined by the referees • the review process is accomplished within one period • accepted articles get published in the period of acceptance • rejected articles are not submitted to another journal • journals try to fill every issue independent from the quality of the articles • libraries are able to cancel journal subscriptions in every period

6 Model verification and validation

Important steps of conducting simulation based research are the verification of computational accuracy (micro level) and the validation of simulation outcomes (macro level). With the overall goal to test the theoretical mechanisms behind the model, for both purposes empirical data can be used (Boero & Squazzoni, 2005). In the following we will give some insights how verification and validation was carried out within our study.

6.1 Verification of computational accuracy

Like many agent-based models, our simulation model contains a relatively large number of parameters and independent variables, for which adequate values or ranges of values had to be determined. Through sensitivity analysis it can be determined which input variables and parameters have significant impact on the desired measures of performance; calibration then determines the appropriate values (Balci, 1989; Law, 2005; Klügl, 2008). We refrain from describing these verification steps conducted during our research comprehensively due to lack of space. Instead, we concentrate on one typical example of parameterization and explain how citation parameters determine the distribution of the emerging network. Therefore, the values of nc^{new} (citation rate of papers read in period t) and nc^{old} (citation rate of papers read in past periods) are – ceteris paribus – varied as follows: (a) $nc^{new} = 10$, $nc^{old} = 0$; (b) $nc^{new} = 7$; $nc^{old} = 3$, (c) $nc^{new} = 3$; $nc^{old} = 7$, and (d) $nc^{new} = 2$; $nc^{old} = 10$. Figure 5 depicts the citation distributions of those different parameter settings at the end of a simulation run.

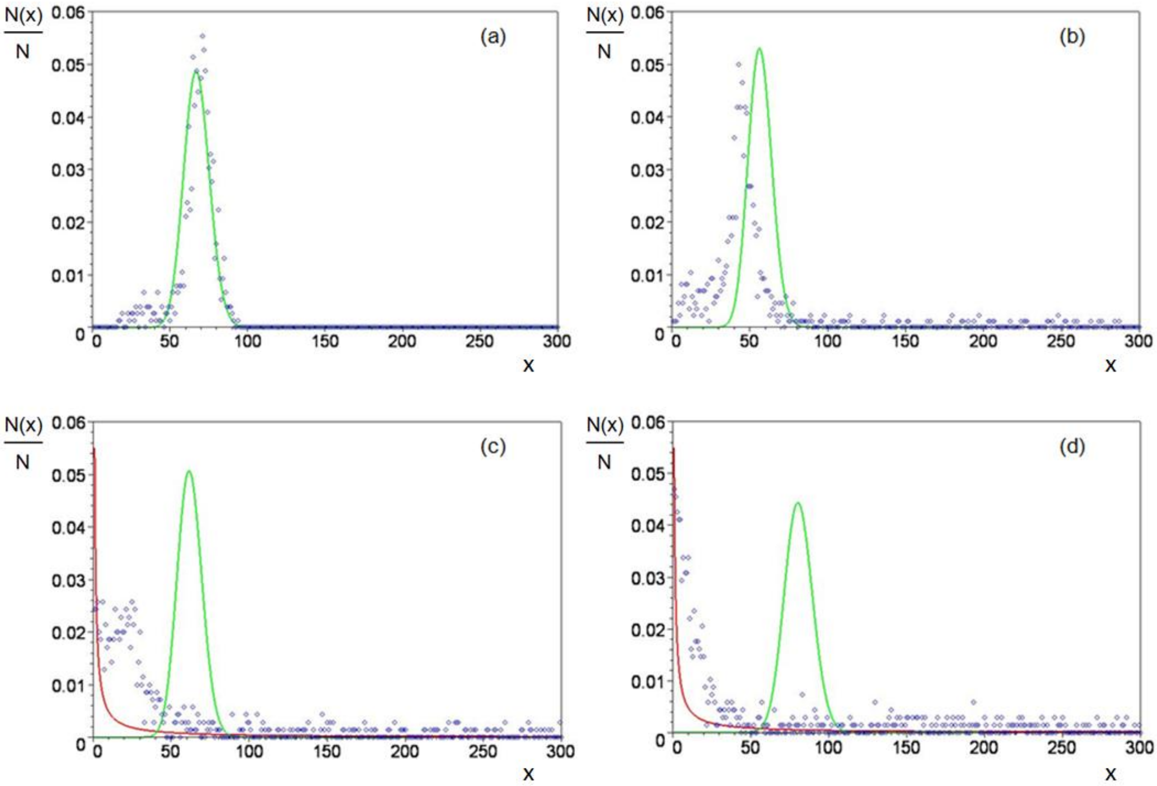

Figure 5: Transition from a Poisson-distributed random network (a) to a scale free network (d) caused by variation of parameters nc^{old} and nc^{new}; x-axis: number of citations, y-axis: number of articles with x citations divided by total number of articles in the network.

At parameter setting (a) authors only cite those articles, which they have read in the current period. In that case, the mechanism of preferential attachment is neutralized, so that the evolving network is completely random. In Figure 5(a) this is shown via the bell-shaped curve representing the Poisson distribution. The ratio of nc^{new} and nc^{old} are reversed step-by-step with the parameter settings (b), (c), and (d), so that in the latter case an author cites only two papers he has currently read, but 10 papers that he/she has read in the past and which have already collected citations. The more citations of those "old" papers (which differ in their citation probability) is allowed, the less random is the citation distribution. At parameter setting (d) the distribution obviously is no longer approximable with a Poisson distribution; it rather shifts towards a scale-free network, which is – as described in section 4.2 – a necessary condition for empirical citation networks. Hence, as in this example, sensitivity analysis and calibration with the distribution function as control measure is an adequate way of verifying the computational model.

6.2 Validation with empirical data

To determine that the output from the overall simulation model is valid, we primarily draw on the approach of comparing the results of the simulation to statistical results derived from large-scale empirical data (Davis, Eisenhardt & Bingham, 2007). The parameter configuration used as the basis for experimentation results in a typical citation distribution as depicted in Figure 6(a). Especially the asymptotic tail of the distribution shows the typical power-law characteristics with exponent α close to 3. Compared to the distribution of real citation networks (see section 4.2 and Figure 6(b)), it can be stated that the two sets of data compare so much that the model of the existing system can be considered as "valid" (Law, 2005). In other words, it can be used to make decisions about the system similar to those that would be made, if it were feasible to experiment with the real system itself (Law, 2003).

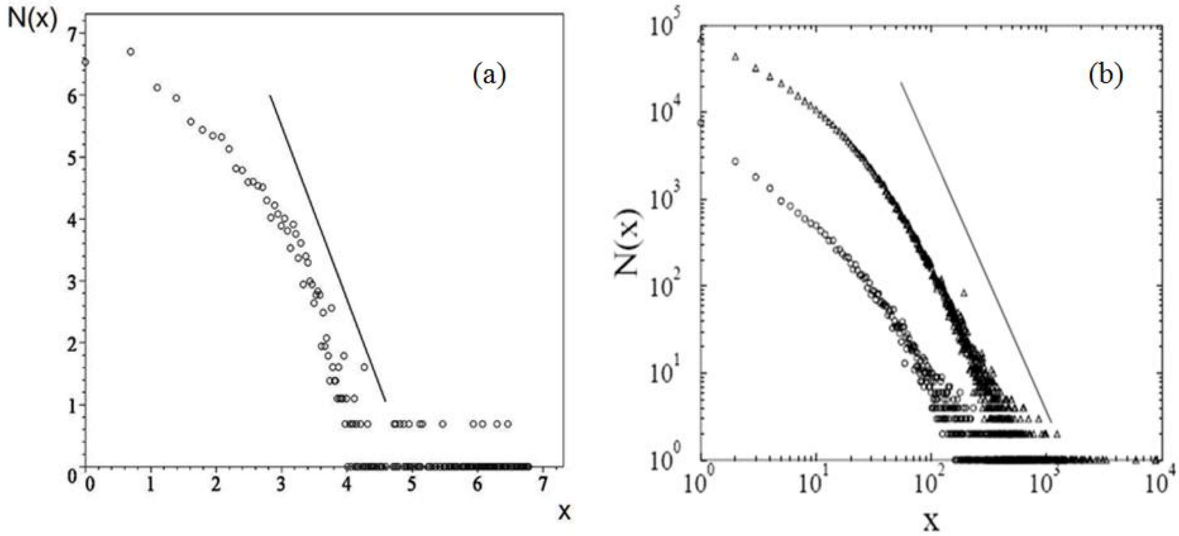

Figure 6: Comparison of (a) the citation distribution evolving during the simulation with (b) distributions of real citation networks found by Redner (1998; see section 4.2); double logarithmical representation with x = number of citations and N(x) = number of articles with x citations; for visual reference a straight line of slope -3 is also shown.

7 Results

7.1 Experimentation scenarios

Dependent on the variation of the input variables, different scenarios can be generated, which describe different evolutions of publishing patterns over time. With traditional publishing as base case, we simulate the transition of the scholarly publication market into a “Green OA world” (scenario A), into a “Gold OA world” (scenario B), and into a “Green and Gold OA world” (scenario C).

In **scenario A** the publication strategies of the authors ($PubStrat_{i,t}$) is modified in such a way that beginning with time period t=50 more and more authors are starting to deposit their articles in a OA repository – independent of the fact, if an article is accepted for publishing in a journal or if it is rejected. During a transition period that lasts 50 time-periods, 95% of the authors change their strategy towards OA.

Scenario B simulates the transition of traditional subscription-based journals into OA journals. This is achieved by manipulating the business model – here denoted as publishing strategy ($PubStrat_{j,t}$) – of some journals in such a way that they transform into OA journals at a certain time of simulation. In the simulation runs that underlie the results presented in section 6.4, two out of eight journals were selected to switch to OA in time period t=70. In this context, the analysis revealed that it is of importance, if these are low quality journals or top journals (with a relatively high usage). We focus on that point in section 7.5.

Scenario C represents a mixed model. Scenarios A and B are combined so that both self-archiving and OA journal publishing are included in the simulation runs conducted for scenario C. In particular, the transition period, where 95% of the authors switch to OA, is kept – as well as the number of journals that transform into OA journals. In order to keep “green” and “golden” effects distinguishable, the change in business models of these journals takes place in t=90, i.e. 40 periods after the self-archiving transition period starts.

Since it is unrealistic that only one model will survive in the future, we modeled the scenarios in such a way that traditional subscription does disappears completely. In this context, we agree with Hall (2010) who criticizes the Houghton approach for modeling and evaluating pure OA worlds. We try rather to account for the transition process that has to take place (and in some areas already has begun to take place) before OA can be broadly established.

To assure reliability of the results, for every scenario at least 100 simulation runs were processed. The diagrams presented in the following subsections either represent average values of the output measures, or show the results of one representative run. The latter is appropriate for visualization of the development over simulation time of those variables that are connected to individual agents. Within an agent based simulation, the development of the values of individual agents cannot (and should not) be *a priori* determined from the simulator. For example, whereas the overall citation pattern is similar for each run, it is not always the same agent/author that gets the most cites (see Figure 11). It is not possible to build average values of all processed runs for those variables.

7.2 Base case scenario: Visualizing the serials crisis

We have mentioned above that since the simulation program is targeted on the prediction of long-term developments, the pricing strategy of publishers is modeled in an economically more reasonable way than it is often executed in present time. Nonetheless, we want to exemplify the effects of restrictive pricing. For reasons of clarity and comprehensibility the following example is kept simple (only four journals and 3 libraries with a constant budget over time) and the charts are taken directly out of the simulation applet.

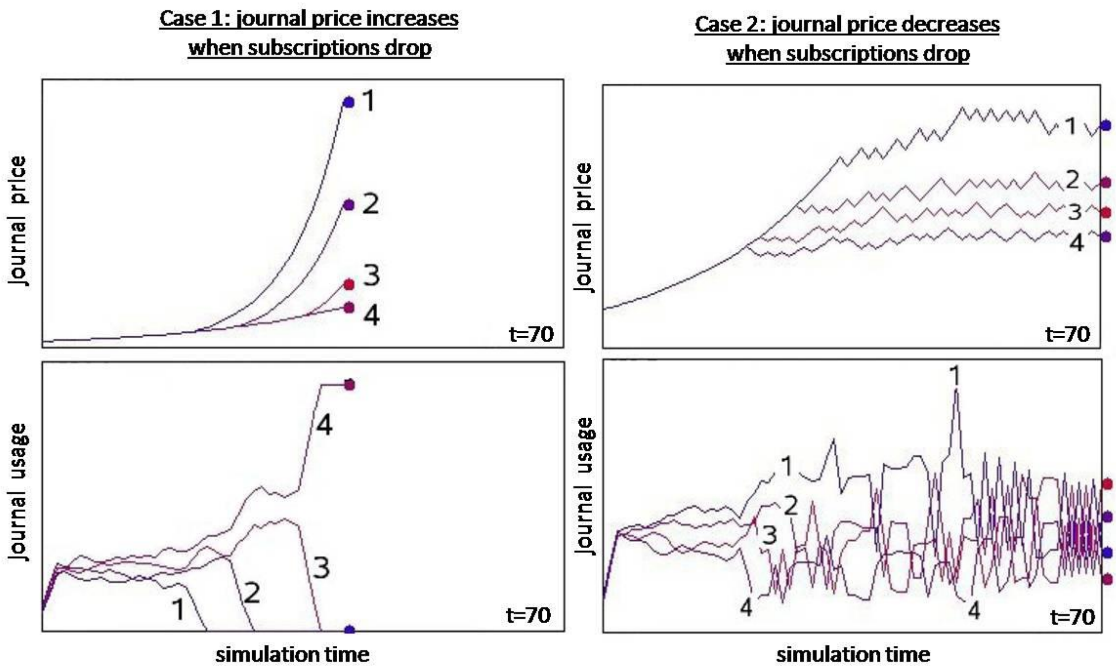

Figure 7: Comparison of long-term effects of alternative pricing strategies of publishers/journals; screenshots for both cases taken from the simulation applet at time period t=70.

In case 1 (Figure 7, left), the price of a journal is increased, if subscriptions drop. At the beginning of the simulation run, all journals are subscribed to by every library. The publishers increase their journal price to a constant rate per period ($\Delta_{Price} = 0.05$) and the journals begin to separate from each other in terms of quality (of the published articles) and thus in usage (articles read). When the budget of a library is no longer sufficient to subscribe to all titles, that journal with the lowest price/usage ratio (journal 1) is cancelled. To keep its profit margin, in the next period this journal raises its price over

the constant increasing rate, which again lowers the probability to be subscribed to again. The vicious circle for journal 1 ends, when no library subscribes to it any more. In consequence, the usage of this journal is zero from that point on. This process recurs over time, until only one journal survives. In the presented case, journals 1, 2, and 3 are cancelled, and journal 4 survives. With this economically counterintuitive pricing strategy those journals with the lowest reputation and usage have the highest prices. The simulation program is – in this simple case – determined to end when the price of the residual journal amounts to two-thirds of the total budget of the libraries.

Those results qualitatively do not change, if the number of journals and libraries are increased. A declining library budget, which can be observed in many institutions, only accelerates the monopolization tendencies. Library budgets increasing faster than the average price increase of journals are only a theoretical case. When comparing the simulation results with the real existing market for scientific journals, it can be said that the market is currently in a transitional phase and an extreme outcome as in this example will be prevented. Nonetheless, simulation is a fruitful method to visualize possible negative outcomes of a continuation of current pricing policies.

In case 2 (Figure 7, right), the price of a journal is decreased, if subscriptions drop. Again, the budget of the libraries is held constant over time. After the starting time span of steadily increasing journal prices, the journal with the lowest price/usage ratio is cancelled by some libraries, because they can no longer afford to subscribe to all journals. That journal (journal 4) lowers its price in the subsequent time periods, until it is subscribed to again – in this case to the disadvantage of journal 3. When journal 3 is cancelled, the usage of journal 4 increases again. This pattern continues throughout the simulation run, also including the remaining two journals. In this way, monopolization, as observed in case 1, does not occur. Whereas the journal prices in average increase to a slow rate over time, the journal usage fluctuates during the ongoing simulation run. The pricing mechanism of this second case is taken as a basis for the following scenario settings.

7.3 Transition to Green OA (scenario A)

In the simulation scenarios presented in the following, Green OA means that – parallel to the journal submission – authors deposit their articles in a repository. The repository therefore contains all articles produced by OA authors, independent of the fact if they are rejected or accepted. In reality, this case is comparable to the arXiv server or similar repositories of other disciplines. The left side of Figure 8 shows the linear transition process of the authors. Starting in period t=50 the researchers change their strategy, until in period t=100 a maximum of 190 authors self-archive.

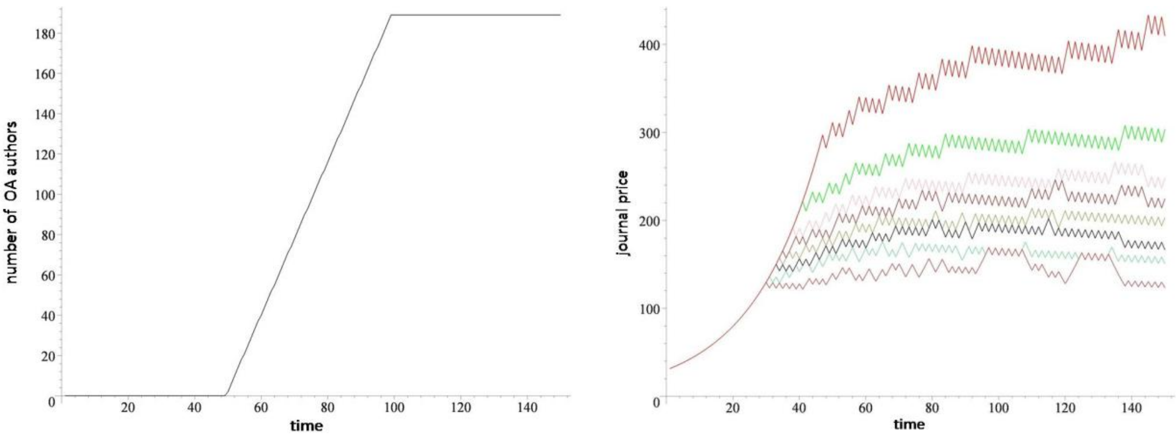

Figure 8: Distribution of authors switching to an OA strategy between t=50 and t=100 (left); development of journal prices over time (right).

In every scenario presented in the subsequent sections the basic parameters are fixed at 200 authors, 8 journals, and 3 libraries. Higher numbers that are more realistic can also be simulated with the program, but then the emerging article network as well as the computing time increases exponentially (while not changing the results qualitatively). To enable graphical depiction of the findings, we therefore selected the settings described above – not without making sure that the results are not biased because of the relatively small number of agents and nodes.

The right side of Figure 8 shows the development of journal prices over time. As above mentioned we based the simulation runs on the pricing strategies and budget parameters described in case 2 in the previous section. Since we did not try to model "real" numbers of the journal price, the amount of the values plotted on the ordinate is negligible.

The left diagram of Figure 9 depicts the development of the journals' usage compared with the repository usage. In this context, usage is defined as the number of articles that are accessed via the respective medium. For instance, if an author in the "read articles phase" (see section 5.1) chooses a specific article he/she wants to read and the journal is held by the library he/she belongs to, he/she accesses ("downloads") the article directly via the journal. If the journal is not subscribed to by the library, but the article is written by an OA author, it is accessible via the repository. If none of this applies, the author is forced to select another article. It can be seen that – as described in section 7.2 – the usage of the journals decreases after about 30 periods, when the libraries start to cancel titles. The fluctuation of the curves result from the model assumption that the libraries' decision whether a title is cancelled or subscribed to can be changed with every period.

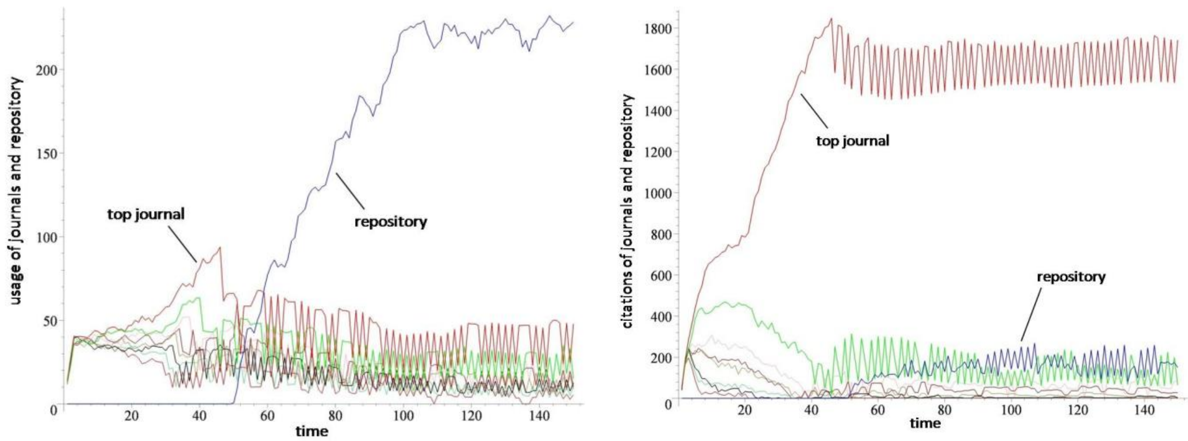

Figure 9: Development of journal and repository usage over time (left); development of citations per journal and repository articles over time (right).

The influence of the transition to Green OA on the utilization of different download sources becomes obvious when looking at the development of the repository usage. Starting in period 50, the number of articles accessed via the repository increases with the proportion of OA authors in the network. The journals are still "used" in the sense that their articles are read and cited. As the right diagram of Figure 9 shows, the citations of articles are still attached to the journal, in which the article is published. Furthermore, the top journal is cited heavily; some of the low quality journals are cited less frequently. The repository only draws citations from rejected papers – citations of published papers would count for the respective journal. But the articles deposited in the repository are not necessarily of low quality. As in reality, rather weak papers, which are submitted to low quality journals, may have a higher probability to be accepted than papers of higher quality, which are submitted to a top journal. Authors who search and cite articles dependent on quality (see section 5.1) also discover articles self-archived

in the repository. This is why the repository even battles with the number two journal for second rank in citations.

A major advantage of a Green OA world is that the authors no longer have to rely on their libraries' subscriptions to specific journals. Provided that the publishers follow a considerate pricing strategy and the libraries collectively agree on continuing subscriptions on a certain level, the development of journal prices (right side of Figure 8) and usage (left side of Figure 9) over time indicate that such a change towards Green OA must not necessarily result in a collapse of the whole journal-based system.

The positive impact of Green OA on the supply of information in the network is depicted in Figure 10. In contrast to the diagrams presented above, the accessible information can be visualized in one curve – in this case by calculating average values of 100 simulation runs. Based on the parameter settings described above, the percentage of accessible information (articles in journals that are subscribed to by libraries together with articles deposited in the OA repository) is twice as high when self-archiving is broadly accepted (and performed) by the scientists. Starting with a value of 48% in a strictly traditional publishing system (200 authors, which submit their articles to 8 journals with a capacity of 12 articles per issue/period; 8*12/200 = 0.48), the accessible information begins to decrease when the first wave of cancellations takes place. With beginning of the transition to Green OA, the information supply increases until it levels off at about 75% at the end of the simulation. The degree of accessible information depends on the ratio of OA authors, which is 95% in this case. There is one simple reason why the information accessibility is about twenty points lower: A large part of the articles in the network is written by OA authors, but at a time period where these agents had not yet switched to OA (and thus did not self-archive those articles). The right side of Figure 10 shows that the quality of accessible papers increases to a similar extent. Since the quality is measured accumulatively it is necessary to couple the percentage of accessible information to it. This measure is not meaningful until other access-provisioning models are included in the analysis. We will refer to this point when discussing Gold OA and mixed models in the following subsections.

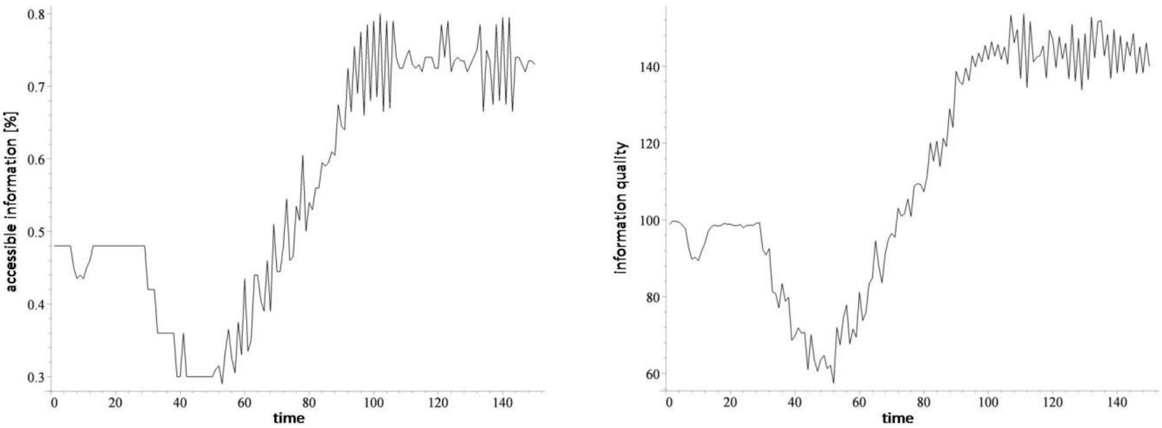

Figure 10: Percentage of accessible information (left); average quality of accessible information (right); average values calculated from 100 simulation runs.

The last issue we want to discuss based on scenario A is the impact of OA on the level of individual authors. Figure 11 depicts the development of citations per author over the simulation time. Similar to the citation distribution on the article level, a typical "long tail" distribution can be observed. A few authors collect most of the citations, whereas the majority of authors are seldom cited. The longer the simulation runs (and the citation network grows), the more obvious this structure becomes. The reason for this lies in the above-mentioned mechanism of preferential attachment: If one author has more citations than another, the probability that he/she gets cited again is also higher.

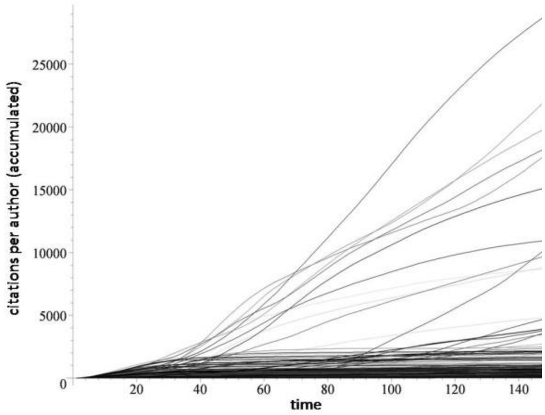

Figure 11: Accumulated citations per author.

The simulation findings show that authors benefit from a switch to OA because the higher visibility of OA articles in combination with the mechanism of preferential attachment results in a higher citation probability. If from two newly-written articles of the same quality, one is made openly accessible, that article has a higher probability to be cited in the future than the other article. These findings are in line with the conclusions of a recent article by Gargouri et al. (2010). The authors focus on the OA citation advantage by especially investigating the so-called self-selection bias, which suggests that the best authors are more likely to make their articles available under an open access model, and that they are more likely to do so with their best work (e.g., Craig et al., 2007; Moed, 2007). Gargouri and colleagues show that the OA citation advantage is greater for “citable” articles – in the case of our simulation model those articles with relatively higher quality and thus higher probability get cited. There is indeed a quality bias – however, it is not caused by the authors, but caused by the users self-selecting what to read and cite (Gargouri et al., 2010). In other words, OA does not make an “uncitable” article citable, but helps citable articles to get cited more often. Our results show that some authors who switch to OA during the transition of the community even manage to escape from the bottom of the long tail. Those chances of an increased standing in the community could be the main incentive for authors to change their publication behavior. However, OA citation advantages over colleagues certainly disappear, when all authors switch to OA. But because in most disciplines this is not to be expected in the medium term, “first movers” can benefit in the transition period. For a deeper analysis of these causal relationships we refer to Bernius and Hanauske (2009).

7.4 Transition to Gold OA (scenario B)

In the following we describe effects of a partial transition to Gold OA. In contrast to scenario A, where OA was initiated by the author, we now address a change of business models on the publishers’ side. In particular, two journals migrate to OA journals in simulation period t=70. In this section we focus on a case where the top journal and a journal with relatively low reputation switch to OA. Furthermore, in the presented run the libraries have no additional budget to finance the OA journals, i.e. they have to pay the publication fees for the articles published in the OA journals out of the same budget they use to finance subscriptions to traditional journals. We varied these settings when conducting the simulation runs, but due to space restrictions we focus on the findings related to those variations in section 7.5, where a mixed model is presented.

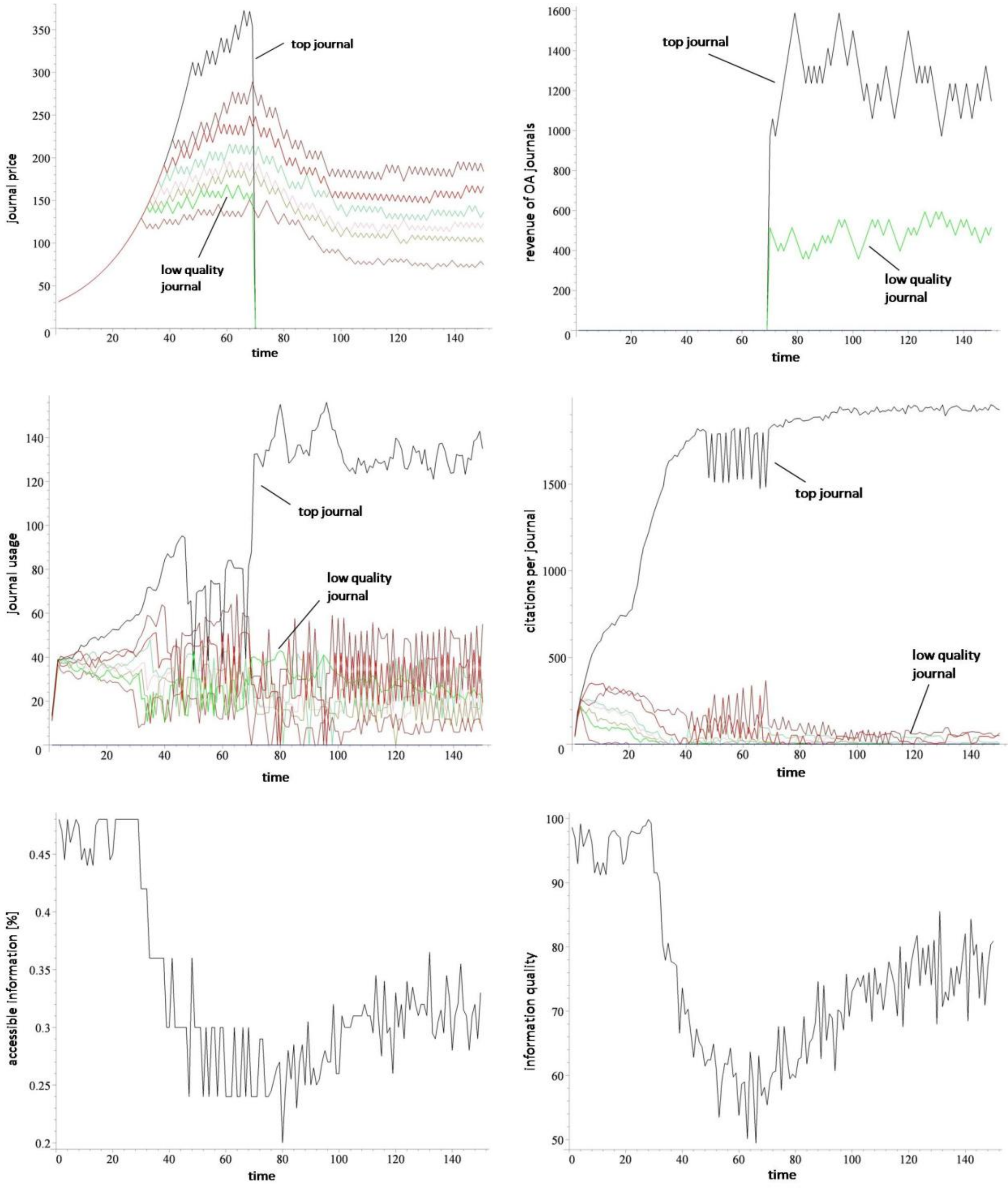

Figure 12: Simulation of a partial transition to Gold OA; development of journal prices over time (top left); revenues of OA journals (top right); development of journal usage over time (middle left); development of citations per journal over time (middle right); percentage of accessible information (bottom left); average quality of accessible information (bottom right).

Figure 12 combines the relevant graphical exemplifications of scenario B. The progression of the journal prices (top left) shows the switch to Gold OA particularly well: in period 70 the subscription price of the two OA journals falls to zero. From this point they generate income from the publication fees taken from the same public fund that allocates the libraries budget. The diagram top right visualizes the revenues of the OA journals. The starting values of the author fees were chosen in such a way that the journal generates the same income as in its last traditional period. In the following periods this

initial fee is kept constant, whereas the number of published articles per issue/period is variable. This setting is consistent with reality: On the one hand, OA publishers cannot change the author fee from issue to issue without losing credibility; on the other hand the number of articles per issue especially varies in pure electronic titles like the OA journals established to date. The progression of the OA journals revenue therefore reflects the number of published articles per issue. Looking at the price development of the remaining traditional journals after the OA transition, one can observe a clear price decrease between periods 70 and 100. This general decrease in prices is conditioned by the lower expenditures for journal subscriptions by the libraries. They have to pay for the articles in OA journals and therefore cancel traditional journals. Those are forced to lower their price until the price/usage ratio is on a level acceptable for the libraries to subscribe again. After this phase of decrease the prices adjust at a constant level, which is significantly lower than in the last period before the OA transition.

The impact of the transition to OA publishing on journal usage (middle left diagram of Figure 12) is obvious for the top journal: After the switch to OA, downloads of articles from the journal increase significantly. The reason for this lies in the fact that this journal publishes those articles with the highest quality. Per model assumption (and this is observable in reality, too), authors try to access those articles more frequently than low quality articles. Hence the low quality journal that switches to OA does not profit from the change as much as the top journal. Only in the 30 period long time span where the usage of the traditional journals decreases, the second OA journal can increase usage. Afterwards it falls back to “scientific insignificance”. Depicting citations per journal (middle right diagram of Figure 12) reveals the same prediction. The top journal assures its pole position through the change to OA, but the low quality journal is seldom cited. Summing up it can be said that the scientific community welcomes the change of the top journal to OA. Low quality journals may suffer from such a development – especially when there is no additional public budget for information supply available.

The information accessible in the network is depicted in the bottom left diagram of Figure 12. In the first 50 periods the curve progression is similar to scenario A. Starting with a value of 48%, the accessible information begins to decrease when the first wave of cancellations takes place. The OA switch in period 70 leads to a slow rise of accessible information (from about 0.25 to a constant value of about 0.31 at the end). In relation to this, the increase of information quality (bottom right diagram) is slightly higher, which is due to the fact that the top journal is beyond the OA journals in this case. Since we only let 25% of the journals change their business model towards Gold OA, the results concerning the information quality and supply are not directly comparable with scenario A, where 95% of the articles get deposited freely accessible in the end. Determining more journals to change to OA would unsurprisingly lead to a higher degree of accessible information. Unfortunately, because of the absence of reliable data, one can only estimate the growth of Gold OA – especially in comparison with the progress of Green OA.

7.5 Mixed model (scenario C)

We finish this section by showing the results of a scenario that includes both OA models, green and golden. Again starting with the traditional setting, self-archiving is initiated in period t=50, followed by the switch to author-paid OA journals in period t=90. We selected this development (with a stronger representation of Green OA), because we consider it being the most likely way a transition towards OA will occur in reality. As stated above, exact data of the growth of different models are hard to find, but some papers provide an informative basis for estimation. Using Ullrich's Periodicals Directory for their study, Björk, Loos and Lauri (2009) conclude that on the basis of published articles the increase of Gold OA is likely to be lower than 0.5% per annum. They estimate the overall percentage of articles published as “true” Gold OA in 2008 around 7% – compared to about 12% available as Green OA. Whereas the growth of Gold OA is modest, the increase of articles deposited in OA repositories remains more significant – amongst other things elevated by the number of institutional and funder’s OA mandates which has grown rapidly in the last two years. Looking into the future, when authors have to change their business models it seems more likely that they will be forced to choose self-archiving than publishing in ÒA journals. Furthermore, initiating the two OA models in the way described above is necessary for segregating the respective effects on the output variables. For instance,

a simultaneous initiation of gold and green leads to overlapping effects that impede clear interpretations. Since the simulation environment supports designing different mixed scenarios by just changing parameter settings, comparison of different mixed scenarios is a central part of research in progress and in future.

A main problem of a system transformation towards Gold OA is the financing of the transition phase. On the one hand, publishing an article in an OA journal requires paying author fees. On the other hand, subscription costs for (in the transition phase still existing) traditional non-OA journals have to be funded in parallel. We thus modeled two cases for scenario C. In C1 the existing budget for information supply is held constant, meaning that author fees have to be paid out of the same budget as the subscription fees. In this case, the overall system costs do not increase. C2 describes the case, when author fees are funded by an extra budget – the subscription budget remains untouched. Technically, this differentiation is simply reached by varying the binary input variable SubStrat (see section 5.4).

To ensure comparability of the two sub-scenarios, it is furthermore necessary that journals of a similar reputation level switch to OA. We present that case in which the two top journals switch to Gold OA, because we assume transformation of the most influential journals to cause more significant effects. Figure 13 depicts the diagrams for journal price, OA journals revenue, and usage. On the left, scenario C1 is visualized, the graphs on the right present scenario C2.

Until simulation period t=90 the development of journal subscription prices (Figure 13, top) is similar for both sub-scenarios. Switching the two top journals to OA results in an immediate price decrease to zero again. From time period t=91 the OA journals generate revenue through author fees (Figure 13, middle). Because of the facts that the submission rate to these high reputation journals is relatively high and we assume their issue capacity to be variable, the revenue then increases slightly for the second best and significantly for the top journal. This behavior is not directly implemented in the simulation program. It results simply from the fact that the OA journals are not bound to a fixed issue capacity: As in reality, it is unusual to provide a print version of an OA journal. And for strictly electronic journals it is much easier to publish a greater number of articles in a specific issue – for example caused by a large number of submitted manuscripts that fulfill the journals quality requirements for a special issue. What is inherent here is the apprehension that OA publishers could give in to the temptation to increase their revenue at the cost of quality. In this context, the danger of increasing author fees in order to “create” revenue (Bernius et al., 2009) seems to be the lesser of two evils.

An interesting issue when comparing cases C1 and C2 is the price development of the remaining traditional journals. If there is only one (fixed) budget for financing the subscription model as well as Gold OA, then the remaining traditional journals could face the problem of cancellations and in consequence sinking prices (Figure 13, top left) – particularly, if the top journals switch to OA. The increasing subscription prices in C2 (Figure 13, top right) result from the interim surplus that is available when funding author fees via an additional budget. Since the OA journals no longer have subscription prices, all the remaining traditional journals can be subscribed to by libraries for a certain time span. This also leads to a short time decrease of repository usage (Figure 13, bottom right). As the number of cancellations caused by the increase in journal prices (at about t=120) grows authors will revert to the repository as the main source for article downloads. Besides, for both journals in both sub-scenarios the switch to OA leads to a significant increase of usage (Figure 13, bottom). Without migration of the two journals to OA, the repository usage would be as high as in the pure green scenario (left diagram of Figure 9).

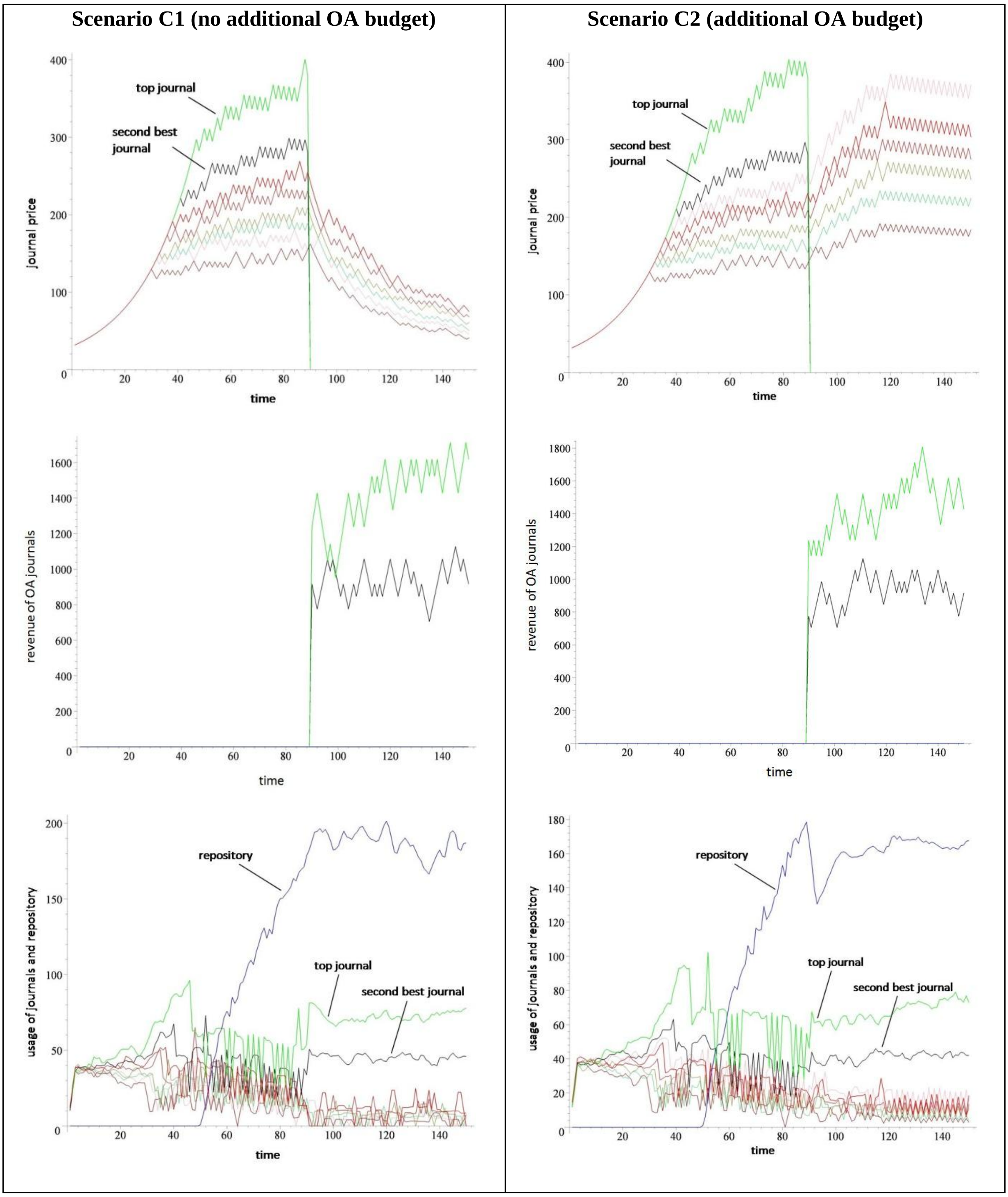

Figure 13: Simulation of a transition to Green and Gold OA; comparison of two settings varying the subscription strategy of the libraries (C1: no additional budget; C2: additional budget for author fees); development of journal prices over time (top); revenues of OA journals (middle); development of journal usage over time (bottom).

System costs are not explicitly modeled in these scenarios. It is evident that C2 leads to higher costs, but to make conclusions in terms of efficiency, it is necessary to include other measures like the amount of information accessible through different models. A comparison of scenario A (solely Green OA) with scenario C1 concerning the average level of accessible information at the end of a simulation run (Figure 10, left vs. Figure 14, top left) does not reveal significant differences. In the modeled case, financing author fees of OA journals without the use of an additional budget does not affect the

percentage of accessible information in the network. The reason for this is the simultaneous cancellation of traditional journals starting with period t=90. Financing Gold OA with an extra budget leads to an increase of accessible information of about 5%. Thus, these results indicate that when Green OA is widely adopted, an additional (costly) support of OA journals seems questionable. Though, the quality of the potential OA journals has to be taken into account.

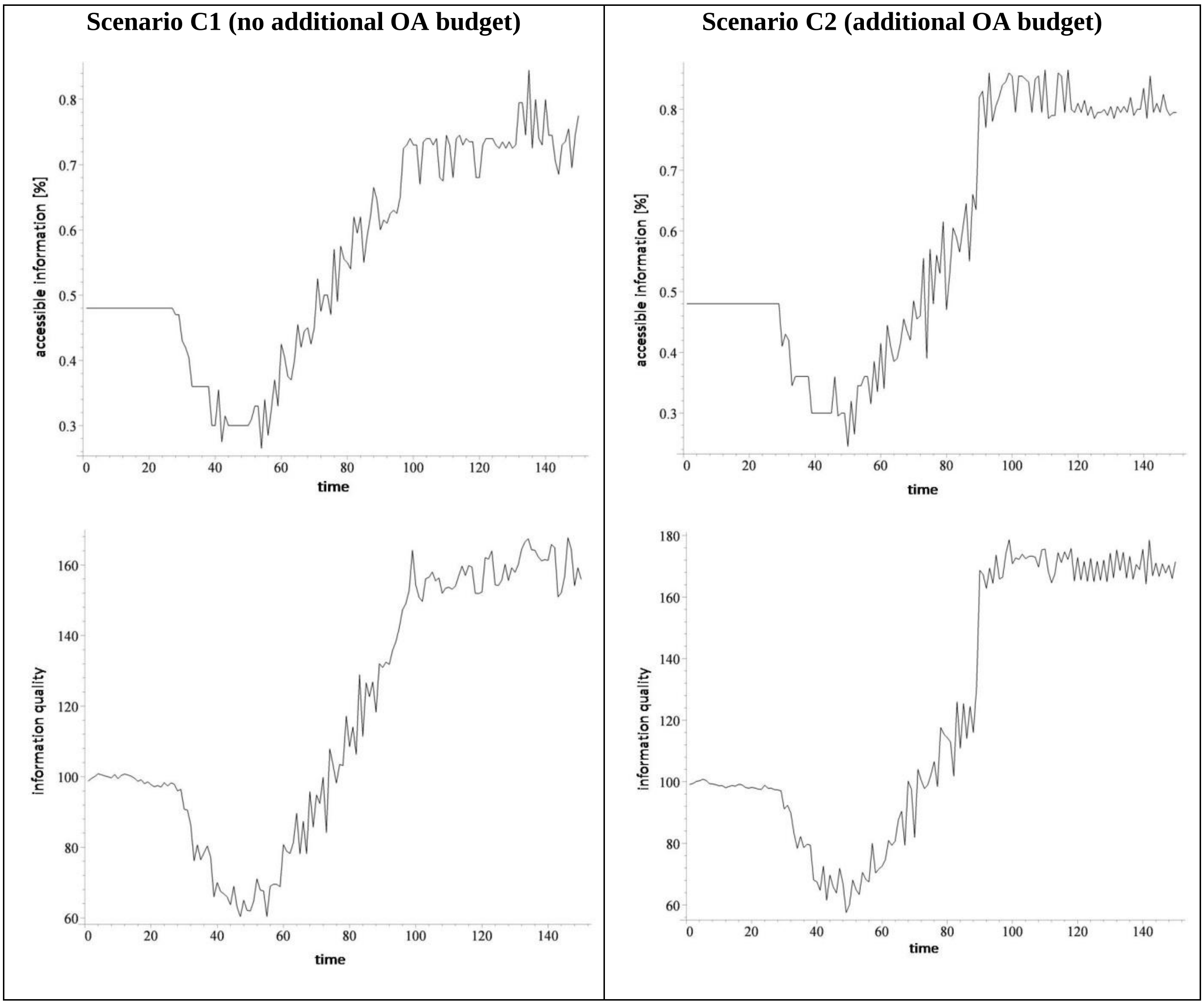

Figure 14: Simulation of a transition to Green and Gold OA; comparison of two settings varying the subscription strategy of the libraries (C1: no additional budget; C2: additional budget for author fees); percentage of accessible information (top); average quality of accessible information (bottom).

The progress of the average article quality (Figure 14, bottom) over the simulation time closely follows the development of information accessibility. This is due to the fact that in the numerator we simply add the qualities of accessible papers. Therefore the absolute level of the quality is of interest. Comparing C1 with C2 shows a slight advantage of C2 – in analog to the information accessibility curves. Significant effects are observable when comparing two cases where the switch to OA is accomplished by journals with differing levels of quality/reputation. Other things being equal, if the two journals with the lowest quality switch to OA, the quality of accessible information levels off about 20 points lower than in the scenario settings addressed here. We conducted these runs through experimentation, but refrain from visualizing the associated results in order to avoid “over-complication”. Summing up it can be stated that in terms of communicating high quality research it is indispensable that the top journals of a discipline convert to OA. The same holds for authors: a scientific community benefits particularly from self-archiving conducted by top scientists who publish high quality papers.

8 Conclusions

8.1 Summary of results

Before discussing contributions of this study to theory and practice as well as limitations and future research opportunities, we want to give a short summary of the main results derived from experimentation with the simulation model. Oriented on the research questions specified above, the following propositions can be drawn:

- *Continuing the current practice of increasing journal prices in combination with shrinking (or stagnant) library budgets results in a long-term system collapse.* This development is not really surprising, but the two main goals set in this context were fulfilled. On the one hand, the observed outcome is an important indicator that the simulation "reacts" in the right way. On the other hand, we wanted to explore and visualize the development of critical variables over time, which – if not counteracted – leads to a crash of the traditional publishing system.

- *A continuation of the traditional subscription-based system (provided that publishers pursue an "appropriate" pricing strategy as well as constant library budgets) leads to a decrease in accessible information.* The fact that the percentage of articles accessible for authors in the network begins to decrease when libraries are forced to cancel titles is the second important effect of the serials crisis, which can be visualized with the use of the presented simulation program.

- *From the perspective of an individual scientist, unrestricted access to his/her work can broaden an author's impact in terms of increasing citations.* This OA citation advantage is predominantly caused by the mechanism of preferential attachment and particularly effective in the transition phase towards OA. Thus, especially in a non-OA community, where the majority of scientists do not yet make use of OA, "first movers" may benefit from path-dependencies induced by an early adoption of an alternative access-provisioning strategy. However, it must be pointed out that OA does not make "uncitable" low quality articles suddenly citable.

- *Compared to the traditional publishing model, the amount of information accessible in the network increases in both OA models, self-archiving and OA journals.* A direct comparison of the two models is difficult, because of the fact that they are initiated by different market players (the agent groups "journals" vs. "authors").

- *A broad adoption of Green OA does not necessarily mean the end of the traditional journal-based scholarly communication system.* If publishers and libraries compromise on a licensing strategy comprising adequate journal prices along with subscription rates that guarantee continued existence of traditional journals, self-archiving can co-exist with the subscription-based system. However, it requires publishers and libraries to both agree to such a mutual co-existence.

- *When aiming for a broad transition to Gold OA, with regard to cost efficiency, financing of OA journals is efficient only if no extra-money is spend.* Indeed, compared to financing without additional budget, financing with additional budget has the potential to lead to a further increase in the degree of accessible information. But in relation to the required amount of expenditures, this "higher increase" is rather small.

- *From the publisher's viewpoint, a transition to Gold OA is beneficial particularly for top journals that possess a high degree of usage before the switch to OA.* Whereas these journals increase their usage after a switch to OA, journals with a lower reputation may suffer from a broad Gold OA adoption. The latter especially holds true for the case where no additional budget for financing publication fees is established. Furthermore, since highly selective journals in particular have to

factor the cost of refereeing many rejected articles into the costs of every accepted article; one could also think of replacing publication fees with refereeing fees (Harnad, 2010).

- *With regard to information supply and quality, it is important that top journals switch to OA.* In terms of reading papers and as target for the publication of their scientific work, authors normally frequent a community's top quality journals to a relatively high degree. If those journals change their business model towards Gold OA, the scholarly communication system in general may benefit, because authors gain access to those papers they actually want to read. However, if publishers of top quality journals allowed their authors to immediately self-archive the papers, the public would additionally benefit in terms of a cost reduction.

- *From the viewpoint of scientists, both OA models fulfill their needs, but for the scholarly communication system in general a transition to Green OA seems to be the most beneficial development.* Our results indicate that Green OA combined with a certain level of subscription-based publishing (scenario A) would provide a sufficient level of accessibility while being less expensive than a broad adoption of Gold OA. The latter should only be supported against the background of getting top journals to provide openly accessible scientific work.

8.2 Contributions to research and practice

In methodical terms the main contribution of this study is the development of an agent-based simulation of the scholarly publishing market that comprises typical actions of the market participants as well as the interplay between them. This research instrument enables closing a research gap that exists due to the rather static viewpoints of existing literature towards evaluation of alternative access-provisioning strategies. Our aim was to discover mechanisms and predict developments that may lead to (and explain) specific outcomes of possible market transformation scenarios. In this context, simulation is the adequate approach, because it enables comprehensive experimentation with a model of the real market.

Furthermore, this study contributes to theory of agent-based modeling by developing and documenting an example of how the emergence of macro-level outcomes can be achieved as a consequence of well defined and plausible sociological micro-processes. This can be illustrated, for example, by looking at the large number of intermediate variables, which emerge during a simulation run but cannot be directly controlled. The way the simulation model was built and validated can furthermore be seen as a successful application of the building blocks approach. Using the method of a stepwise implementation of agent groups and relating actions allows controlling complexity when designing large artificial experimentation environments.

We also contributed to the theory of complex networks by showing how power-law network emergence based on accumulative advantage can be induced without explicitly modeling the mechanism of preferential attachment (which would be much simpler, because it requires only few lines of code). In our approach the agent is more than just a mindless node with complete information of the network structure. The power-law distribution emerges as a result of the behavior of the individual agents and the interaction between agent groups on the micro-level. In our case, a direct implementation of preferential attachment would have considerably biased validation on the macro-level and thus interpretability of the simulation results.

This study contributes to theories related to OA in different ways – for example, by substantiating the argument of a citation advantage of OA articles. At the individual level we showed how authors can benefit from self-archiving by increasing their reputation measured in the number of citations to their work. In addition, the simulation results visualized the mechanisms of a journal system collapsing in the long-term because of a continuation of the serials crisis. Hence, the vicious circle argumentation in the literature could be corroborated and visualized over time.

The practical contribution of this research stems from the integration of all market players. Of course, besides our stakeholders, we particularly seek to provide researchers and the public with guidance concerning publishing policies. For example, decisions regarding potential financial support of OA models can be aligned with our findings. Nonetheless, commercial publishers can also draw their conclusions from the presented results. The implication here is that especially high quality journals, which may already be heavily frequented, can benefit from a transition of their business model into an OA journal. On the other hand, as Waltham (2010) found out, most of the journals are currently not in need of forcing changes: Even in the digital age they are still selling a large number of print versions and are constantly increasing their revenue. In reverse, this fact should alert the public to actually increase endeavors to fundamentally restructure scholarly communication. We hope that our results are somewhat helpful for adjusting future strategies.

8.3 Limitations and future research

This study should help in understanding the effects of OA on the dynamics of the scholarly communication market. However, our study has some limitations that predominantly stem from the research method we used. Like other methodologies, a simulation-based approach has shortcomings that are typical for the method. Some of these problems we have already counteracted as much as possible, others serve as references for performing future research.

The most obvious problem of simulation models is the fact that the results are sometimes based on rather unrealistic assumptions (Davis, Eisenhardt & Bingham, 2007; Whicker & Sigelman, 1991). In order to avoid hardly manageable complexity we had to shorten processes, define agents in such a way that they (to be more precise: their actions) are homogeneous and assume all strategic rules as to be effective. Hence, some constructs which are sometimes complicated even in real life– like the negotiation of so called “big deal” licensing packages between libraries and publishers – are not adequately covered in our simulation model. Second, simulation findings are dependent on the initializing values of parameters and input variables (Whicker & Sigelman, 1991). We addressed issues of parameterization and model validation in section 6. In fact, simulation findings can only be demonstrated for the parameter regions that form the basis of experimentation. As stated by Harrison et al. (2007), generalizations beyond this space can at best be referred to as conjectures. In this context, an important starting point for future research is advancing the empirical grounding of the scales and ranges. As stated above, simulation should not be seen as substitute for empirical research, but as a complement to it.

Another important point is discussed by Chaturvedi et al. (2005): Because of the emergent nature of agent-based models, exact mechanisms that lead to outcomes observed at a macro level (e.g., the efficiency measures discussed in the results) are hard to identify. We responded to this problem by thoroughly determining the micro processes and rooting them in theory. Nonetheless, especially in a rather comprehensive simulation environment as created for this study, there is always potential for refining simulation constructs such as the “0” and “1” bit strings we use as representation for strategies. Besides a further empirically grounding of micro processes, with regard to validity and reliability it would also be fruitful to adjust the size of the emerging network: More agents and more exactly defined time periods would bring the simulation model closer to the real world. Whilst this is also a question of more computing power, on the other hand the trade-off between standardization and complexity has to be taken into account. Our primary goal was to take a first step in analyzing effects of developments in scholarly communication from a dynamic viewpoint. In our opinion, this was achieved by the approach we chose. However, our future research is oriented at extending and refining the experimentation scenarios in order to gain more detailed insights in the complicated interferences taking place during the current transformation of the scholarly communication system.

Acknowledgements

This research was supported by grants from the German Research Foundation (DFG, Grant No. GZ554922). We gratefully acknowledge the financial support.

Literature

Axelrod, R. (1997). Advancing the art of simulation in the Social Sciences. In R. Conte, R. Hegselmann & P. Terna (Eds.), Simulating Social Phenomena (pp. 21-40). Berlin: Springer.

Axelrod, R. (2007). Simulation in the Social Sciences. In J-P. Rennard (Ed.), Handbook of research on nature inspired computing for economy and management (pp. 90-100). Hersey, PA: Idea Group.

Axtell, R. L. (2000). Why agents? On the varied motivations für agent computing in the social sciences. Working paper 17, Center on Social and Economic Dynamics. Retrieved March 23, 2012, from http://www.econ.iastate.edu/tesfatsi/WhyAgents.RAxtell2000.pdf

Balci, O. (1989). How to assess the acceptability and credibility of simulation results. In E. A. MacNair, K. J. Musselman & P. Heidelberger (Eds.), Proceedings of the 1989 Winter Simulation Conference (pp. 62-71). Washington D. C.

Barabasi, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. Science 286, 509-512.

Barton, R. R. (2004). Designing simulation experiments. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), Proceedings of the 2004 Winter Simulation Conference (pp. 73-79). Washington D. C.

Bernius, S., Hanauske, M., König, W. & Dugall, B. (2009). Open access models and their implications for the players on the scientific publishing market. Economic Analysis and Policy, 39(1), 103-115.

Bernius, S. & Hanauske, M. (2009). Open access to scientific literature – Increasing citations as an incentive for authors to make their publications freely accessible. In Sprague, R. H. (Ed.), Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, IEEE Computer Society Conference Publishing Services.

Bernius, S. (2010a). Speeding up the spiral – Analysis of the effects of open access on scientific knowledge creation. Proceedings of the 18th European Conference on Information System (ECIS), Pretoria, South Africa.

Bernius, S. (2010b). The impact of open access on the management of scientific knowledge. Online Information Review, 34(4), 583-603.

Björk, B-C., Roos, A. & Lauri, M. (2009). Scientific journal publishing: Yearly volume and open access availability. Information Research, 14(1), paper 391. Retrieved March 23, 2012, from http://InformationR.net/ir/14-1/paper391.html

Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T. & Gudnasson, G. (2010). Open access to the scientific journal literature: Situation 2009. PloS One, 5(6), e11273.

Boero, R. & Squazzoni, F. (2005). Does empirical embeddedness matter? Methodological issues on agent-based models for analytical Social Science. Journal of Artificial Societies and Social Simulation, 8(4). Retrieved March 23, 2012, from http://jasss.soc.surrey.ac.uk/8/4/6.html

Carley, K. M. (2001). Computational approaches to sociological theorizing. In J. Turner (Ed.), Handbook of Sociological Theory (pp. 69-84). New York, NY: Kluwer Academic/Plenum Publishers.

Chaturvedi, A., Mehta, S., Dolk, D. & Ayer, R. (2005). Agent-based simulation for computational experimentation: Developing an artificial labor market. European Journal of Operational Research, 166, 694-716.

Cole J. R. & Cole, S. (1973). Social Stratification in Science. The University of Chicago Press, London, Chicago.

Craig, I. D., Plume, A. M., McVeigh, E. M., Pringle, J. & Amin, M. (2007). Do open access articles have a greater research impact? A critical review of the literature. Journal of Informetrics, 1(3), 239-248.

Creaser, C., Fry, J., Greenwood, H., Oppenheim, C., Probets, S., Spezi, V. & White, S. (2010). Authors' awareness and attitudes toward open access repositories. New Review of Academic Librarianship, 16(1), 145-161.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing theory with simulation methods. Academy of Management Review, 32(2), 580-599.

Dooley, K. (2002). Simulation Research Methods. In J. Baum (Ed.), Companion to Organizations (pp. 829-848). London: Blackwell.

Epstein, J. M. & Axtell, R. (1996). Growing Artificial Societies. Cambridge: MIT Press.

Eysenbach, G. (2006). Citation advantage of open access articles. PLoS Biology, 4(5), 692-698.

Fladung, R. B. (2007). Scientific communication – Economic analysis of the electronic journal market. Stuttgart: ibidem.

Gargouri, Y., Hajjem, C., Lariviere, V., Gingras, Y., Brody, T., Carr, L. & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. PLoS One, 5(10), e13636.

Gilbert, N. & Troitzsch, K. G. (2005). Simulation for the social scientist. Berkshire: Open University Press.

Hall, M. (2010). Minerva's Owl – A response to John Houghton and Charles Oppenheim's "The economic implications of alternative publishing models". Prometheus, 28(1), 61-71.

Hall, S. (2010). A commentary on "The economic implications of alternative publishing models". Prometheus, 28(1), 73-84.

Hanauske, M., Bernius, S. & Dugall, B. (2007). Quantum game theory and open access publishing. Physica A, 382(2), 650-664.

Hajra, K. B. & Sen, P. (2005). Aging in citation networks. Physica A, 346, 44-48.

Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y, Oppenheim, C., Stamerjohanns, H. & Hilf, E. (2004). The green and the gold roads to Open Access. Nature Web Focus. Retrieved August 13, 2011, from http://www.nature.com/nature/focus/accessdebate/21.html

Harnad, S. (2010). No-fault peer review charges: the price of selectivity need not be access denied or delayed. D-Lib Magazine, 16(7/8).

Harrison, J. R., Lin, Z., Carroll, G. R. & Carley, K. M. (2007). Simulation modeling in organizational and management research. Academy of Management Review, 32(4), 1229-1245.

Hein, O., Schwind, M. & König, W. (2006). Scale-free networks – The impact of fat tailed degree distribution on diffusion and communication processes. Wirtschaftsinformatik, 48(4), 267-275.

Hitchcock, S. (2011). The effect of open access and downloads ('hits') on citation impact: A bibliography of studies. Retrieved August 13, 2011, from http://opcit.eprints.org/oacitation-biblio.html

Houghton, J., Rasmussen, B., Sheehan, P., Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, M. & Gourlay, A. (2009). Economic implications of alternative scholarly publishing models: Exploring the costs and benefits – A report to the JISC. Retrieved March 23, 2012, from www.jisc.ac.uk/media/documents/publications/rpteconomicoapublishing.pdf

Houghton, J. & Oppenheim, C. (2010). The economic implications of alternative publishing models. Prometheus, 28(1), 41-54.

Jubb, M. (2011). Heading for the open road: Costs and benefits of transitions in scholarly communications. Liber Quarterly, 21(1), 102-124.

Kim, J. (2010). Faculty self-archiving: Motivations and barriers. Journal of the American Society for Information Science and Technology, 61(9), 1909-1922.

King, D. W. & Tenopir, C. (1999). Evolving journal costs: Implications for publishers, libraries, and readers. Learned Publishing, 12(4), 251-258.

Klügl, F. (2008). A validation methodology for agent-based simulations. Proceedings of the 2008 ACM symposium on Applied computing (pp. 39-43). Fortaleza, Ceara, Brazil.

Laherre, J. & Sonette, D. (1998). Stretched exponential distributions in nature and economy: ‚Fat tails' with characteristic scales. European Physics Journal B, 2, 525-539.

Law, A. M. (2003). How to conduct a successful simulation study. In S. Chick, P. J. Sanchez, D. Ferrin & D. J. Morrice (Eds.), Proceedings of the 2003 Winter Simulation Conference (pp. 66-70). New Orleans, USA.

Law, A. M. (2005). How to build valid and credible simulation models. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong & J. A. Joines (Eds.), Proceedings of the 2005 Winter Simulation Conference (pp. 24-32). Orlando, Florida.

Law, A. M. & Kelton, D. W. (1991). Simulation Modeling and Analysis (2nd ed.). New York, NY: McGraw-Hill.

Lehmann, S., Lautrup, B. & Jackson, A. D. (2003). Citation networks on high energy physics. Physical Review E, 68, 026113.

Leicht, E. A., Clarkson, G., Shedden, K. & Newman, M. E. J. (2007). Large-scale structure of time evolving citation networks. The European Physics Journal B, 59(1), 75-83.

Mann, F., von Walter, B., Hess, T. & Wigand, R. T. (2009). Open access publishing in science: Why it is highly appreciated but rarely used. Communications of the ACM, 52(3), 135-139.

Marks, R. E. (2007). Validating simulation models: A general framework and four applied examples. Computational Economics, 30(3), 265-290.

McGuigan, G. S. & Russell, R. D. (2008). The business of academic publishing: A strategic analysis of the academic journal publishing industry and its impact on the future of scholarly publishing. Electronic Journal of Academic and Special Librarianship, 9(3).

Merton, R. K. (1968). The Matthew Effect in Science. Science, 159(3810), 56-63.

Moed, H. F. (2007). The effect of "Open Access" upon citation impact: An analysis of ArXiv's Condensed Matter Section. Journal of the American Society for Information Science and Technology, 58(13), 2047-2054.

Odlyzko, A. M. (1997). The economics of electronic journals. Firstmonday, 8(2). Retrieved March 23, 2012, from http://www.firstmonday.org/issues/issue2_8/odlyzko

Oppenheim, C. (2008). Electronic scholarly publishing and open access. Journal of Information Science, 34(4), 577-590.

Park, J.-H. (2009). Motivations for web-based scholarly publishing: Do scientists recognize open availability as an advantage? Journal of Scholarly Publishing, 40(4), 343-369.

Pujol, J. M., Flache, A., Delgado, J., & Sanguesa R. (2005). How can social networks ever become complex? Modelling the emergence of complex networks from local social exchanges. Journal of Artificial Societies and Social Simulation, 8(4). Retrieved March 15, 2012, from http://jasss.soc.surrey.ac.uk/8/4/12.html

Ragin, C. C. (1987). The comparative method: Moving beyond qualitative and quantitative strategies. Berkley: University of California Press.

Redner, S. (1998): How popular is your Paper? An empirical study of the citation distribution. European Physics Journal B, 4, 131-134.

Repenning, N. (2002). A simulation-based approach to understanding the dynamics of innovation implementation. Organization Science, 13(2), 109-127.

Sargent, R. G. (2004). Validation and verification of simulation models. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), Proceedings of the 2004 Winter Simulation Conference (pp. 17-28). Washington D. C.

Shockley, W. (1957). On the statistics of individual variations of productivity in research laboratories. Proc. IRE 45 (1957).

Swan, A. (2010). The Open Access citation advantage: Studies and results to date. Retrieved August 13, 2011, from http://eprints.ecs.soton.ac.uk/18516/

Wagner, B. (2010). Open access citation advantage: An annotated bibliography. Issues in Science and Technology Librarianship, 60(Winter 2010). Retrieved August 13, 2011, from http://www.istl.org/10-winter/article2.html

Waltham, M. (2010). The future of scholarly journal publishing among Social Science and Humanities Associations: Report on a study funded by a planning grant from the Andrew W. Mellon foundation. Journal of Scholarly Publishing, 41(3), 257-324.

Ware, M. (2006). ALPSP Survey of Librarians on Factors in Journal Cancellation. Association of Learned and Professional Society Publishers. Retrieved March 15, 2012, from http://www.marywaltham.com/ALPSPresultsANDconclusions_.pdf

Whicker, M. L. & Sigelman, L. (1991). Computer Simulation Applications – An Introduction. Newbury Park, California: Sage Publications.

Xia, J. & Sun, L. (2007). Assessment of self-archiving in institutional repositories: Depositorship and full-text availability. Serials Review, 33(11), 14-21.