Department of
Information Engineering
and Computer Science DISI

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)
http://www.disi.unitn.it

LIQUID JOURNALS: KNOWLEDGE DISSEMINATION IN THE WEB ERA

Marcos Baez, Fabio Casati, Aliaksandr Birukou and
Maurizio Marchese

February 2010

Technical Report # DISI-10-028

Liquid Journals: Knowledge Dissemination in the Web Era

Marcos Baez, Fabio Casati, Aliaksandr Birukou, Maurizio Marchese
Dipartimento di Ingegneria e Scienza dell'Informazione
University of Trento
Via Sommarive, 14 38100
Trento, Italy
{baez, casati, birukou, marchese}@disi.unitn.it

ABSTRACT

In this paper we redefine the notion of "scientific journal" to update it to the age of the Web. We explore the historical reasons behind the current journal model, and we show that this model is essentially the same today, even if the Web has made dissemination essentially free. We propose a notion of liquid and personal journals that evolve continuously in time and that are targeted to serve individuals or communities of arbitrarily small or large scales. The liquid journals provide "interesting" content, in the form of "scientific contributions" that are "related" to a certain paper, topic, or area, and that are posted (on their web site, repositories, traditional journals) by "inspiring" researchers. As such, the liquid journal separates the notion of "publishing" (which can be achieved by submitting to traditional peer review journals or just by posting content on the Web) from the appearance of contributions into the journals, which are essentially collections of content. In this paper we introduce the liquid journal model, and demonstrate through some examples its value to individuals and communities. Finally, we describe an architecture and a working prototype that implements the proposed model.

Keywords

Academic journals, Liquid journals, Enhanced Search, Multi-faceted artifacts

1. INTRODUCTION

The opportunities for knowledge dissemination today are fundamentally different than only 20 years ago. In the beginning of the modern scientific communication era (more than 300 years ago[1] the scarce and expensive resource was the *printing* and *distribution* of *papers*. Publishing was expensive. As a result, there was a need to screen contributions before they got published, and there was no other better means to do this than peer review. The process was the only reasonable one, and it was feasible as the research community was relatively small and the reviewing effort was low. Printing and distribution also meant that the journal had to be organized in volumes and *issues*, available periodically.

Today, the (social) Web has opened a world of possibilities for how the notion of "journal" could evolve to serve the need of scientist to learn about novel, interesting research ideas and results, in a way that is more efficient (less dissemination overhead) and more effective (makes it easier to find interesting, relevant, novel content). Specifically, publishing today is essentially real time and free. Printing and distribution are no longer the scarce resource in scientific dissemination. Furthermore, *papers* are no longer the only possible unit of scientific contributions that can be disseminated. Blogs with interesting ideas, scientific experiments, comments on somebody else's paper, reviews, slides, videos, demos, experiments, even publicly available data sets can all be interesting contributions to science.

In this paper we do not want to enter religious wars on whether blogs or data sets are a form of scientific contribution, or whether only properly written and peer-reviewed papers should be considered as "reputable" by society. We simply observe that the *original reasons* for having the current journal model (content including only papers, division in issues, submission and peer review process) are gone. At the same time, other challenges have come with the novel opportunities. The ease of publication along with an increase in the number of people doing research, have caused a tremendous increase of knowledge artifacts that are disseminated every day. This means that the scarce resource is now *attention* [11] and the real obstacle to dissemination - and the challenge for scientists - is not publishing, but rather making a contribution *visible* (on the author's side) and quickly identifying *interesting* contributions in an ocean of publications (on the reader's side). Yet, the world of scientific knowledge dissemination works essentially like the pre-web era.

The ambitious goal of this paper and of the line of research it describes is to initiate and facilitate a fundamental change in how scientific dissemination operates, with the goals of i) making it significantly more efficient and effective, with the meaning described above, of ii) facilitating behavior that we believe are good for science (such as early sharing of ideas and results), and of iii) establishing ways to assess reputation that go beyond citation-based metrics, that have well-known flaws [12] [10] [16].

[1]The first recorded peer review process was at The Royal Society in 1665 by the founding editor of Philosophical Transactions of the Royal Society, Henry Oldenburg

In this following we present an approach, a family of models, an architecture and an implementation for a notion of journal, called *liquid journal* (hereafter also denoted as LJ), that leverages the opportunities and the lessons learned from the social web. Through liquid journals, researchers can find and share "interesting" scientific content "related" to a certain area of research. It is a family of models because it supports, in terms of both concepts and implementation, a spectrum of model from the more traditional ones to the ones more social and web-aware, that we describe below. This is because we believe that the right balance can emerge with time and it will be the scientific community to select the point in the spectrum that will prove to best fit the needs of scientific dissemination.

1.1 Main concepts and motivations

The paper and the additional material, software, and videos referenced within, will describe in detail the model and the implementation and discuss its benefits (and also its limitations and risks of failure). We however summarize here the key ideas and motivations behind them. None of them per se is particularly surprising, but taken together can enable the fundamental shift mentioned above.

A first concept is the decoupling of the publication of an artifact (which can be achieved by just posting content on the Web) from its appearance in a journal. Specifically, we see liquid journals as a structured set of links to web-accessible content (whether freely available or only a under a commercial subscription will depend on the specific content and related access policy). A contribution can therefore appear in many journals (if the author and copyright policies allow, as discussed later). The benefit is that journal editors can select any content, not just the one explicitly submitted to the journal, significantly broadening the selection base. This decoupling also enables a possibility very important for LJ to be effective: everybody can in principle become an editor, set up editorial policies (and a board, if desired), start a journal, and select content for it. As we will see shortly, this is beneficial for content selection and assessment.

The second concept is that LJ is based on a conceptual model for contributions in which the contributions can go beyond papers, and can include blogs, data, experiments, and in principle anything that can be referred by an URI. They are all first-class citizen, also for the purpose of assessing reputation. We believe this is important as it encourages the "publication" of well-crafted experiments and datasets, which indeed contribute to the advance of research. In addition, we model contributions as evolutionary objects. We realize that research does not only proceeds by fundamental shifts, but also by small increments. Today scientists often extends their research and publish a "brand new" paper for any extension. LJs supports a conceptual model in which papers and contributions in general can be "versioned" and in which reuse is seen as a good thing, as long as it is explicitly stated and tracked[2]. Contributions can also appear in multiple facets (a paper, a slide, a related video or demo, etc..) and the LJ model supports this, thereby facilitating the consumption of the scientific knowledge in the paper.

[2]Incidentally, this is the original reasons for calling these journals as "liquid", as opposed to the "solid" image of a paper which is modeled as something set in stone as opposed to an evolving description of an evolving research.

The third concept has to do with how the LJ infrastructure supports editors in selecting content. It is a key point also because it enables (and leverages, in a recursive fashion) reputation metrics which we believe are more significant than currently adopted ones and that encourage behaviors that are good for science. LJs are essentially a view over the web, and content can be decided with traditional (peer-review) processes, can be picked "by hand" by the editor, or can be selected automatically, by querying the web and ranking content based on reputation metrics, and then possibly refining the results by hand. The decision of whether a contribution belongs to a journal (and hence of what is interesting, inspiring, relevant) is the result of a mix of explicitly stated filters and rules, of recommendations and collaborative filtering approaches, and of peer reviews if desired. Because we have a potentially large number of editors thanks to the decoupling (people can create journals simply with a query, and can do so for their own reading, or for their group), we use the power of the community (the editors) as a way to filter and rank the sea of content on the web. The same large community that generates the noise problem can also solve it. Indeed, appearing in a large number of journals (especially in journals to which many people subscribe) is a measure of reputation.

But the most important reputation metric we consider is the sharing: the LJ infrastructure allows readers to share content they like with the individuals or groups (analogously to what we do today "by hand", by sending emails to colleagues). We take this sharing of a contribution as the prime measure of reputation for the contribution, as when we share content with colleagues we are asking them to spend time to read it, so we are doing an action that bears a significant cost in terms of time. Another form of reputation is when editors subscribe to content from an individual, because they want to be kept up to date with the research of a scientist of a group. These measures are in turn used to rank scientific content when searching the web and supporting therefore the "search" aspect which is part of the editor's activity in a liquid world. Furthermore, releasing scientific ideas early is rewarded because editors are more likely to subscribe or share content from a creative scientist.

A final goal and benefit of liquid journal is to enforce *diversity.* We have experienced, in this very same line of research, that we kept looking for prior art in our own community. By interacting with physicist, with philosophers, and with researchers from completely different communities, we realized that there was a large amount of related work in those communities we were not aware of. We also realize that those researchers were also not aware of research in the CS community. Hence a goal of the LJ infrastructure is that of offering results on a topic coming from diverse communities, as discussed in the following.

1.2 Benefits

In summary, the benefits of the proposed liquid journal model, as discussed above, are the following:

1. It allows researchers to have a tailored journal to read what they care about.

2. It allows real-time dissemination, that is, it exposes new ideas and brainstorming-like thoughts besides validated/reviewed research. It also exposes papers as soon as they "appear" on the Web.

3. It combines breadth and depth: It combines personalization with awareness to diversity as the journal selection model allows combining relevance, novelty, and interest as search criteria (so a reader may learn about very relevant research, but also about research less related but more novel or considered interesting by the community).

4. It exploits the filtering power of the community to help select interesting contributions. It also observes behaviors in the sense of reading/ tagging/ forwarding contributions to determine, also in real time, the interestingness of a contribution. In essence, the journal contains contributions that members of the community believe to be "worth reading". It is like a mechanical turk for finding good contributions.

5. It rewards creativity, early sharing of ideas, and collaboration: by considering (and therefore rewarding) blogs and in general non-reviewed thoughts as contributions, it encourages scientist to share their ideas. This is key as collaboration is a great catalyst to innovation. Today ideas are not shared early as they can be taken by others and turned into a paper, but if appropriate reward is given to seed thoughts, then the obstacles to sharing will be reduced.

6. It provide a complete, *lightweight*, and real-time assessment: Evaluation is a necessary aspect of research, not only to filter contributions but also to help select people for hiring or promotion. Having an LJ model allows looking into all aspects of a researcher's productivity (innovative thoughts contributions, reviewed contributions, etc...) and to have evaluation performed in real time (novel contributions, if interesting, can spread quickly like news). The real time aspect is particularly important for young researchers where citation count always takes a year or so for the uptake. In addition the model allows complementing the review-based evaluation with a lightweight evaluation, in the sense that peer review is possible but not necessary, while it is the community that while reading/tagging/-forwarding contributions of interest naturally provides a way to assess content.

7. it will naturally allow the community to select a dissemination model preferred by the community, and this will happens by looking at which journal models among the many variations available to LJ editors (selection process, kind of contributions, reputation metrics) are eventually successful.

8. Finally, the evolutionary and liquid approach encourages reuse (stated and tracked) of previous contributions, minimizing dissemination overhead and effectively allowing one to build (also in terms of dissemination) on the shoulder of giants, including themselves.

2. RELATED WORK

In the following we review the progress in dissemination and evaluation models, we analyze tools and approaches that help scientists to make use of the social web to collaborate and get interesting scientific content, and finally, the review search services and approaches to access scientific content available on the Web.

2.1 Novel dissemination models

In the recent years, Open Access movement lead to some progress in the traditional dissemination model. An indication of this is the publication of papers (or preprints) in open repositories and archives. One example is arXiv[3]; a service that provides archiving for academic publications in different fields. Using this service, authors can submit their publications to the archive and keep them up-to-date thanks to a versioning facility. Paper acceptance is decided by moderators, who, essentially, make sure the paper is properly categorized and is not spam. arXiv provides community features, like mail alerts and link to add content at social bookmarking services. However, the information about usage is not exploited by the system, since no recommendation mechanism or ranked search is provided.

Several novel journal models (e.g. PLoS ONE[4], Frontiers[5], and many others) have emerged. All these models try to improve some aspects of traditional publishing by modifying the process, involving the community, etc. However, they are still based on the traditional notion of paper, on submission process, on specific review processes. For further discussion on the topic we refer the reader to [6].

The most relevant to LJ approaches are those of the deconstructed journal model [17] and the overlay journal model [15]. In the overlay journal model, the author writes a paper without a particular journal in target and makes it available in a repository. Then, overlay journals can take the paper from the repository and peer review it. Accepted papers are linked from the journal web page. In practical terms, this model has been explored by the RIOJA project [14] in the context of eprints, building an overlay journal on top of arXiv.

Similar in spirit, the deconstructed journal (DJ) model assumes that the different roles (editorial, quality control of content and form, conferring recognition, marketing and dissemination) could be played by different actors instead of a centralized publisher. The main idea of DJ is centered on the Subject Focal Point (SFP, the analogous to journal in the LJ model), which is defined as collections of relevant items of interest to readers. According to this model, multiple journals (SFP) can refer to the same item, and therefore to items already published in other journals (SFP).

LJ share with the overlay model the idea of decoupling the paper creation phase from the publishing, in selecting content from third-party repositories. LJ share with deconstructed journals the ideas of linking to the same content in different journals and of separating different roles played by the traditional journals. However, since we use a broader notion of scientific contributions, LJ provide a more general dissemination model. Moreover, LJ model takes into account personalization, i.e. fitting the content to someone's interests, and evolution of content.

In spite of the progress we have just mentioned, the current model of publishing and evaluating scientific contributions remains almost the same. Only some journals adopt usage metrics [3]. Tools are still constrained to the traditional notion of paper. Preprints and other kinds of non-conventional scientific contributions are rarely considered in the evaluation of authors' performance. In contrast, in LJ

[3]http://arxiv.org/
[4]http://www.plosone.org/
[5]http://www.frontiersin.org/aboutfrontiers/

we aim at evaluating different aspects of the research productivity and by doing so we consider and make available other types of scientific contributions.

2.2 Sharing scientific content

Social bookmarking has been used to share interest within communities. CiteULike[6], Mendeley[7], Zotero[8] and Connotea[9] are examples of social bookmarking services with the focus on sharing academic references. They allows storing, sharing and tagging references to publications. Moreover, similarly to the idea of LJ, Mendeley proposes a concept of shared collections. A shared collection is a collection of references collaboratively edited and tagged by a group of people. The collection focuses on a specific topic, and has a constant URL and subscription via RSS feeds, just like an LJ. Something similar to shared collections of Mendeley exists also in CiteULike and is called groups. A group on CiteULike is a set of people that created shared libraries of references. All members of the groups can see references and tags others posted there, but they cannot see .pdf versions of the articles (unless it is a special, invitation-only private group, where people can share .pdf files). Each group's page is accessible via a URL and updates are available via RSS feeds. One of the big shortcomings of groups in CiteULike that there are several hundreds of groups, sometimes overlapping in the topics and only very basic search facility is provided for searching groups matching one's interests. Connotea and Zotero also have groups, similar in functionality.

The main differences of Mendeley's shared collections and CiteULike groups with respect to the LJ are: a different notion of scientific contribution and issues; no restriction in memberships, a much wider notion of sharing. Moreover, another problem with the above services is that they rely on active users, that is, users who inject content into the system. Ideally, an automated discovery process as the one proposed in LJ will better support the sustainability of such services.

2.3 Search and access to scientific content

From an infrastructure point of view, search engine technology has been explored and applied to scientific content [13]. Specialized search engines, such as Google Scholar[10] and CiteSeer[11], have been developed for searching papers/books across multiple repositories using crawling techniques and protocols. BASE, an academic search engine, indexes the metadata from repositories which implement the OAI-PMH protocol. These services, however, provide only a partial view of what we consider as scientific contributions. Moreover, despite efforts on standardizing the access to scientific content (e.g., the OAI-PMH protocol[12]), these search engines, and other non conventional sources of scientific content, provide heterogeneous interfaces and metadata format.

Given this, the infrastructure of LJ builds on the abstraction of a scientific resource space management system [1], an abstraction inspired in the concept of Dataspaces, which extend DBMS concepts to reach heterogeneous data sources

[6]http://www.citeulike.org/
[7]http://www.mendeley.com/
[8]http://www.zotero.org/
[9]http://www.connotea.org/
[10]http://scholar.google.com/
[11]http://citeseerx.ist.psu.edu/
[12]http://www.openarchives.org/pmh

Figure 1: Example of a query for the creation of a Liquid Journal on peer-review

by allowing integration to be done in an incremental fashion [7]. In particular, building applications over this layer allows searching and operating with multiple data sources using a common interface.

3. EXAMPLE OF LIQUID JOURNAL

Before presenting the details of the liquid journal model and infrastructure, we first describe an actual example of a liquid journal. By going through this example we want to show the value of the liquid journal concept and motivate the ideas underlying this model[13].

Let us consider a large and multi-disciplinary research group doing research on a particular topic: peer-review. Obviously, important factors for the group to carry out its research and function as a group are, among others, to have access to relevant content on the topic, and the appropriate tools for sharing the relevant content within the group. Let us also assume that, as in most realistic scenarios, the group's requirements in terms of content go beyond papers and extend also to datasets, experiments, blogs, and other relevant materials. It is in this context that the group has created a *liquid journal on peer-review*. This journal has the dual purpose of sharing knowledge among the group, and also of creating a community around it to establish the group as a center of competence in the area.

To start a liquid journal on peer review, during the journal definition phase, the *editor* gives shape to the journal by defining the editorial board (i.e., the group members), the properties of the content to be included in the journal and the publication process. The content is defined by the editor by means of a query as shown in Figure 1. The content resulting from the query will serve as input to the selection

[13]A demo video showing the motivating ideas for this example can also be found at http://project.liquidpub.org/research-areas/liquid-journal

procedure. In this case, the editor has selected *papers, blogs and datasets* coming from *Springer and ArXiv* that are related to *peer review.*

Once defined, the journal moves to the next phase: content gathering. The editorial board starts getting "interesting" and "relevant" scientific content as results matching the query definition. The results are essentially links to contributions, and readers can access them based on subscriptions of their libraries or institutions. As indicated in Figure 1, the editors can select from a number of properties of the content to be included in the journal: different *type*, e.g.paper, blog, dataset, according to the shown query; *maturity*, which ranges from early (initial ideas) to finalized (published journal paper); and *certification* ;e.g., peer reviewed paper vs non-reviewed blog. These properties are part of our broad notion of *scientific contribution*. Note that ranges of maturity can depend on the type of contribution, e.g., for a paper levels could be workshop, conference, journal, while for a dataset they can be: raw, cleaned, verified, depending on the quality of data.

Let us now assume one of the editors has received a suggestion from her friend by email: an interesting article on "review behavior". As the editor trusted her friend, she proceeded by adding the article manually. Then, after doing some research, this editor finds in the author's webpage the slides and dataset used in the paper. Instead of putting three unconnected items, the model provides the tools for gluing them together as different facets of the same contribution. The multi-faceted property provides a much more powerful representation, in lines with the actual nature of contributions. Analogous links are also provided to express different versions and variations of a paper (e.g., pre-print, workshop paper, journal paper).

The results are updated according to the policy specified by the editor (e.g., daily, monthly, real-time), as new contributions become available on the web, similar in the spirit to Liquid Queries [4] that change results over time. The advantage of this automatic procedure is that there is no need to go to different sources and search for contributions - they will be delivered to the journal. Results are kept private until editors apply the selection procedure to filter out irrelevant or not interesting results. Another important property of the results is the diversity. In the example in Figure 1, we can see that results comes from different communities (computer science, physics and biology). Liquid journals ensure diversity by supporting queries from a variety of sources.

The next phase of the process is the selection of results. Having an ever-growing stream of contributions makes the proper organization of the collected material a fundamental aspect of the liquid journal. Therefore, the model provides the support for tagging and annotating contributions so group members can quickly refer to them and perform an easier selection. For example, an editor might want to use tags to classify contributions by the specific approach used. Tags and annotations can be private or public, and editors are able to share their ontologies with others, allowing them to view the content through their glasses.

The final phase for editors is publication. For instance, the editors need to prepare a special collection of the most relevant contributions to a particular subtopic, let say, "novel review processes", or to create a collection of best contributions over a particular period of time. Liquid journals provide the notion of issues as the tool for capturing snapshots

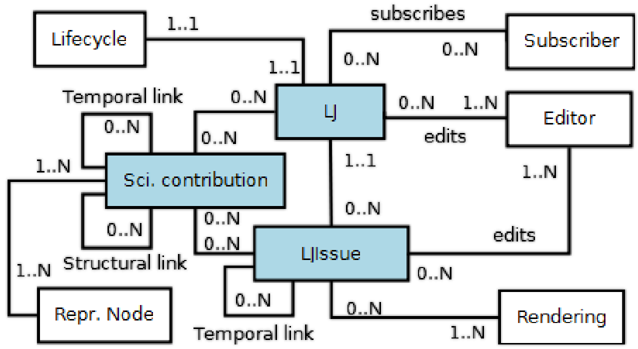

Figure 2: Liquid journals conceptual model

of the evolving liquid journals. When creating a journal issue, the editor defines which contributions are included from the evolving liquid journal and ensures links will not change once the issue gets published.

With respect to the purpose of sharing knowledge within the group, the journal on peer review constitutes a hub for group members to share contributions they find interesting and relevant to the topic of peer-review. Editors can share contributions with all group members by simply putting them into the stream of the journal or share contributions with selected members - that might have a particular interest on specific contributions - just by forwarding them.

Such collective editing and sharing contributes to the overall quality of the journal. People using liquid journal system can subscribe to journals to get updates, e.g., new contributions added, new journal issue released. The quality of the liquid journal, assuming the group does a good job in selecting and organizing the content, will attract subscribers and thus bootstrap a community around the journal. The more people subscribe to the journal, the more visibility the group will have. We call this last phase consumption of a liquid journal.

4. LIQUID JOURNAL MODEL

The liquid journal model builds on three basic elements that replace their counterparts in the traditional model. These basic elements are depicted in the conceptual model in Figure 2. They are:

- **Scientific contributions**. The basic unit of the liquid journal model is the scientific contribution. In contrast to the traditional model, this is not constrained to papers.
- **Liquid journals**. Collections of scientific contributions that can evolve in their definition and content. From now on we refer to these as journals or liquid journals interchangeably.
- **Liquid journal issues**. Issues are snapshots in the journal evolution that do not change once published.

In the following we describe these elements in detail.

4.1 Scientific contributions

Departing from the traditional model of scientific contribution, which considers only papers, we want to define new dimensions that will allow us to understand, classify and evaluate all the range of potential contributions. To this end, we characterize contributions using the following dimensions:

1. **Nature of the contribution**. This dimension refers to the format, type and other essential characteristics by which a scientific contribution is recognized. Examples of categories in this dimension are blogs, datasets, scientific workflows, and combinations thereof (complex artifacts).

2. **Degree of maturity**. The degree of maturity refers to how elaborated and digested is the knowledge expressed in the scientific contribution. It can take the following values: early (vague ideas in blogs), tentative (position papers) and finalized (complete materialized concepts, like in a journal paper), as explained in [8].

3. **Level of certification**. It refers to the degree of scrutiny the scientific contribution was subject to, i.e., at what extend it was validated and reviewed. In this regard, we plan to have different certification types, including of course, the community certification and peer review [5].

We propose the above dimensions to reflect the vision that maturity and certification should be properties orthogonal to each other and to the nature of scientific contributions. Thus, by decomposing contributions following these dimensions, we offer the community the flexibility of choosing the characteristics they want. For example, some journals would want to select contributions of a given maturity and certification degree, regardless the format of the contribution. From the readers' perspective, dimensions will allow filtering content and expressing their interest by mean of subscriptions. Finally, we can propose evaluation methods that take into account these dimensions to compute the reputation of contributions and authors.

Note that we also consider feedback and review as potential contributions. As such, feedback (and especially useful, constructive feedback), contributes to the reputation of the person providing it. In doing so, we encourage community participation in reviewing and filtering information.

Concerning the multi-faceted aspect we mentioned in Section 3, papers represent just one facet of the potentially various ones the research work has (e.g. slides, experimental protocols, datasets ..). This is represented in or LJ model by *structural links* between artifacts. To support the evolution part, we model the temporal dimension as different lines of research work that can evolve and branch in time. We represent this dimension using *temporal links* between artifacts. The notions of multi-facetedness, versioning, branching are elaborated in [9].

Finally, what the above definition implies is that authors will have the possibility of exposing different aspects of their research work, at different points in its development, In doing so, authors will be able to spend more time on doing actual research, and getting early feedback.

4.2 Liquid journal

At the core of the proposed model is the liquid journal element. A liquid journal is an evolving collection of interesting and relevant **links** to scientific contributions available (freely or not) on the web, and it is the analogous to the traditional notion of journal in the proposed model[14].

Considering journals as *collections of links* opens a world of new possibilities in terms of content creation, licensing and copyright models, and services for the scientific community. It means, in other words, that journals not necessarily owns the contributions. Indeed, as captured in our conceptual model, multiple journals can refer to the very same scientific contribution. This "appearance" of contributions in journals is an important information we exploit for capturing interest (see details in Section 5.2).

Another important property of liquid journals is the *evolving nature*. Unlike traditional journals, which are fixed compilation of papers, liquid journals' links to scientific content can change (appear/disappear) in time and so enabling real-time dissemination. In doing so, scientific contributions can be put into the stream of journals to be linked, consumed, reviewed, as soon as they become available; avoiding in this way, time delays related to the periodicity of the traditional journal, and at the same time leveraging the nature of the Web in the dissemination

The driving force of the liquid journal is the editorial board. The board is composed of editors, which can be either organizations or individuals, running and giving shape to the journal. As mentioned earlier, the model provides flexibility in terms of content and processes (e.g., for review and publication). Hence, it is up to the editorial board to decide what to include, how to visualize it and the processes to follow. In the example of Section 3, the editorial board (integrated by the group's members), defined these preferences in the creation phase. Further details of how these preferences can be defined, in Section 5.

Finally, *subscribers* play an important role in the model. They represent the community of people following the liquid journal by means of subscriptions. It is by combining this information that we can leverage community attention as parameter for filtering out the noise and identifying good journals, and by propagation, editors, contributions and also authors.

4.3 Liquid journal issues

The evolving nature of liquid journals is a strong element of our model. However, in some cases, "snapshots" of the evolving links can be useful. For example, as in Section 3, to publish a selection of contributions on "novel review processes". The model supports this by means of *liquid journal issues*. Liquid journal issues are fixed collection of links that cannot change once published. Nonetheless, the scientific content being linked can still change if it is an *evolving contribution*. Therefore, it is up to the editor whether to link a specific version or the latest version (the evolving one).

In the conceptual model of Figure 2, structural links connect the journal issues with their respective liquid journals. Moreover, given that journal issues represent points in the evolution of a liquid journal, the model provides temporal links to connect different issues and capture this relation in time.

The notion of periodic journal publication schedule is optional. One editor may think of periodic issues (essentially snapshots of the "interesting content related to certain areas"), while another may simply desire a "continuous" model

[14] Please note the difference between the "liquid journal model" (general model), and a "liquid journal" (collection)

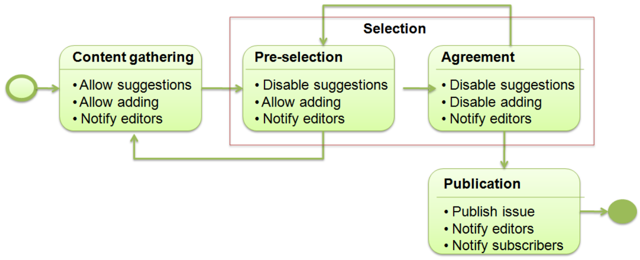

Figure 3: Example of the selection and publication lifecycle for a liquid journal

where content is updated as new contributions, more "interesting" or "relevant" than the current ones are published. The choice also depends on the amount of interesting contributions, e.g. there might not be need to create issues of a journal on a very narrow topic, having only several dozens of contributions.

5. LIQUID JOURNALS LIFECYCLE

In the previous subsection we have described the general model. Here we identify the general aspects of the liquid journal model lifecycle, and describe them in terms of: definition, evolution and consumption.

5.1 Journal definition and publication

In the liquid journal definition phase editors define their preferences in terms of content, collaboration group and selection process. This definition is performed using the *liquid journal definition language*, which is composed of different parts, namely:

General information. This refers to the information the editor compiles when defining the journal and that describes its purpose. For example, the name of the journal and description.

Editorial board. The editorial board defines the collaboration group, the group of people that will collaborate, edit and share the specific journal.

Selection and publication process. Selection and publication process, in its simplest form, can be performed by removing /adding contributions and then making them available, with no other additional phases or restrictions whatsoever. There are, however, many possibilities in the way editors can carry out the selection and publication procedure. In the liquid journal model (and the related language) we want to support editors in the use of different services, potentially dispersed, in order to describe a composite process involving all different roles present in the journal. At present, we rely on a flexible lifecycle management approach [2] that allows users to easily define lightweight workflows. In Figure 3 we illustrate an example of lifecycle for the *journal on peer review*.

Query. The model and language supports users to express their preferences on the nature of scientific contributions they want. The purpose of this is to pre-filter the information by defining a view over the scientific content on the Web. Thus, the query part of the language deals with the explicit editor's preferences in the following dimensions:

- *Type of content*: This refers to particular combinations of nature, certification and maturity levels. i.e. the dimensions of contributions introduced in Section 4.1.

- *Properties of the content*: Each type of content have its own set of attributes. In defining these properties, the editor filters out the content and focus the query to the properties he explicitly searches in the scientific contributions. These properties are defined as a set of n-ary relations on the attributes (e.g., equals, not equals, and any other attribute-specific relation) and connected by logical operators (i.e., conjunction, disjunction, negations).

- *Relations*: Scientific contributions are not unconnected entities. By defining relations (temporal, structural or semantic) among contributions we allow editors to establish their preferences by describing the context of contributions, i.e., how they interact with other entities. Similarity stands up as an important type of relation and, as such, we consider it as first class citizen.

- *Sources*: editors can select the number and type (open, commercial, certified, etc..) of sources in order to enlarge or reduce the scope of the query.

- *Ranking*: query results will provide all contributions available on the Web. In order to support the editors, these contributions need to be ranked and cut at some point according to a relevance measure. Editors can define the ranking criteria, which can be traditional metrics such as citation count, h-index, g-index, or new metrics as discussed in Section 5.3.

- *Clustering*: Clustering is also an important feature, with the purpose of providing query results in a way that it is easy to consume, and deliver its user. It can be done using differente parameters such as sources, authors, community topics etc. depending on the available information.

With the above, we have defined the general properties to be provided when creating a liquid journal. The formalization of the language for expressing such properties is still in development. Nonetheless, in Section 6.2 we show an XML definition based on the early results.

Besides formalisms, the key here to identify the mechanisms and interface metaphors that can assist editors in the process of defining a liquid journal. This implies the definition of interaction schemas that can better be translated into a journal definition. Our primilinary proposal, as seen in Figure 1, relies on an interactive web-based wizard.

5.2 Journal consumption

Liquid journals bring "interesting" and "relevant" scientific content in the form of an informed selection of scientific contributions made initially by the collaborative effort of the editorial board. Interestingness and relevance are thus implicit properties that add to the explicit preferences in the journal definition to provide tailored scientific content to be consumed i) initially by the editors in the editorial board and ii) finally - when the editorial board decide to go live - by a number of readers/subscribers to the liquid journal. We note here again, that the actual content, however, is not in the journal. A liquid journal is defined (see Section 4.2) as a collection of links, and as such, it relies on the actual sources and on the reader's[15] capability to access them. Access permission are always based on the reader's permissions and on what the specific source of the link allows.

In both period of the consumption phase (inside the editorial board or open to subscribers) people will provide feedback. By selecting elements from the results or adding new elements during the selection process, they will collectively establish what is worth reading. Content is also to be shared. In this respect, liquid journals act as a hub for sharing content among people. By adding content to the journal, it immediately becomes available to others. From this perspective, the selection process establish also a sharing policy. But not all contributions are for everyone, and therefore, there are other ways of sharing. Forwarding is a more focused way of sharing content. It allows to take a contribution and share it with the selected sub-group of people that would have a particular interest on the topics of the contribution. Sharing thus enables collaboration through a network of trust. As in social networks, individuals can share their "scientific stories" and get suggestions from the peers they trust and so collectively focus their attention on the things they care about.

The elements we have discussed above are key for determining what individuals consider interesting. Interestingness can be seen as a combination of quality (general property) and what is relevant to the user (personal meaning). This personal aspect thus range from the explicit preferences to implicit ones captured by feedback, sharing, forwarding; and in addition to collaborative filtering and diversity analysis approaches that help *group* and *rank* contributions. Note that the trade off of diversity and interestingness is important for providing content of interest but still allowing people to discover new content.

Thus, a strong point of our model is that we do not rely on people explicitly rating content (providing reviews or giving one or five stars to a contribution), but we rely on things individuals do for themselves (bookmarking, sharing, forwarding,...).It is by combining this selfishness that our model leverages all the features. Finally, this whole model enables the filtering power of the community to select good content and to identify good editors. Furthermore, from subscriptions to appearance, these elements will also make possible new ways of evaluating the work of editors, authors and contributions.

5.3 Journal evolution

Once editors have defined their preferences in the journal definition, the liquid journal is always evolving as new content (matching the definition) becomes available. More precisely, new elements can join this evolving collection as they appear on the Web or based on the update rate defined by the editor (e.g., daily, weekly updates).

Despite the automatic nature of the query, the selection procedure still relies on the ability of editors to identify interesting and relevant scientific content. The editor will refine the content to adjust it to the subjects of the journal and to her personal interests. In other words, the evolution strategy can be characterized as *semi-automatic*, i.e., it is part automatic (query) but with a strong human component (user editing).

We want to underline here, that liquid journals are compatible with the traditional model in that of defining fixed collections. Journal editors can define issues for liquid journals by specifying fixed views of their content on a period of time limiting to peer reviewed content. Note that in our model, however, journals do not necessarily *own* the scientific contribution, and therefore journal content may overlap.

The evolution is then driven by the lifecycle defined by the editorial board. For example, the editors may want to work with submission and perform peer-review before selecting contributions. Or they may rely directly on the query results and their internal filtering.

6. INFRASTRUCTURE

In the previous sections we have drawn the conceptual framework of our proposed model. In this section, instead, we take a practical perspective and describe the implications of the liquid journals model on a specific supporting infrastructure.

6.1 Architecture Overview

Designing and implementing an infrastructure for supporting the LJ model requires solutions and strategies for the different aspects of the model: i) managing the lifecycle of the journals, ii) journal definition, evolution, consumption and sharing; iii) access to scientific content in the Web, iv) computing the reputation of contributions and scientists (for evaluation and ranking), and v) the projection of these features to an appropriate and user-friendly graphical user interface. The LJ architecture relies on specialized components designed for each of the aspects mentioned. In Figure 4 we illustrate these components.

LJ provides a view of the scientific content available of the Web. As scientific contributions, in the broad meaning of the paper, fall outside traditional sources (e.g., digital libraries) where standars can be applied, the infrastructure requires of an access layer that provides us the abstractions for easily accessing and searching content residing in non conventional, dispersed and heterogeneous sources on the Web. In order to address this requirement, we rely on the abstraction of *Resource Space Management Systems* (RSMS) [1] applied to the scientific domain.

The ResMan[16] system, an implementation of a RSMS, provides a uniform access layer to resources available on the Web. It abstracts applications on top of the heterogeneity of the underlying services on. The approach followed by the system is to realy on adapters, components that map the specifics of different and non compatible services to a commmon and uniform protocol. ResMan then allows upper layers to interact with a registry of adapters (resource manager API) and operate on the resources using different levels of abstractions. Details of the system can be found at the system webpage.

On top of ResMan, a scientific RSMS (Karaku)[17] provides a common and extensible conceptual model for scientific resources as well the set of basic services for searching and operating on these resources (as defined in our model in **Section 4**). As seen in Figure 4, we have separated the query language and the access layer from the journal-specific features, as this component provides very intersting features on its own.

On these foundational components, the liquid journal core

[15]the consumer of the content.

[16]http://project.liquidpub.org/resman

[17]http://project.liquidpub.org/karaku/

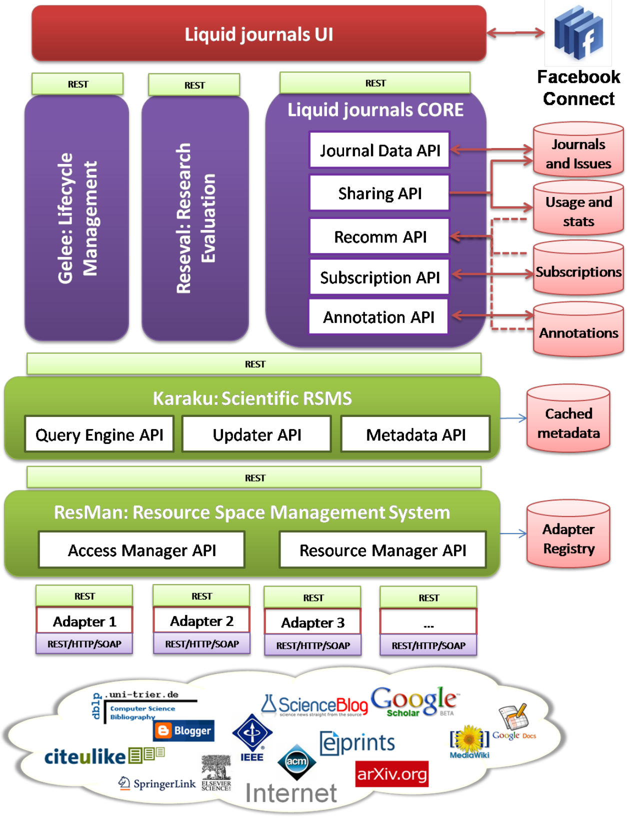

Figure 4: Liquid journals architecture

component builds the services that support the model introduced in this paper. These services are organized in the following groups of APIs [18]: Journal Data API, Sharing API, Recommendation API, Subscription API, Annotation API.

As we need to provide LJ editors the possibility of defining their own processes, the architecture includes a simple lifecycle management component, the Gelee system[19]. An extensible tool for computing metrics for contributions and papers (and any other user-defined entity) - the Research Evaluation tool (ResEval)[20], completes the backend infrastructure. In this context, ResEval takes information about scientific entities from the scientific RSMS and applies the algorithms for computing the reputation of editors, authors, contributions and journals, which are later used in liquid journals for ranking. This is possible due to the flexibe nature of the architecture that allow us to plug the services (and sothe data) of liquid journals as any other source by registering it in ResMan (once the correspondent adapter has been implemented).

Services are very important in our architecture but to fully exploiting them it is necessary to provide a Web interface that facilitates the search, content consumption and sharing. Withouth an appropriate interface, services are worthless. In this approach, we pay special attention to this issue and provide a rich Web application on top. In addition to this, we integrate our application with the Facebook social network with goal of facilitating the sharing, and making it easier for people to use and connect with the system. This is possible through the Facebook Connect service.

6.2 Liquid journals at work

The architecture introduced above is not only conceptual but functional in a working prototype. Thus, following the example introduced in Section 3 below we show how the LJ backend handle a LJ definition. The definition is captured by the web interface shown in Figure 1, translated into the XML below and posted to the REST interface of the Journal Data API.

Listing 1: Example of definition for a "Liquid Journal on Peer Review"

```
<liquid-journal>
  <!-- General info -->
  <name>Journal on peer-review</name>
  <description>
    Most relevant content about peer review
  </description>
  <!-- Editorial board -->
  <editor-list>
    <editor>
      <user-ref
        href="liquidpub.org/user/1482912814"
        value="Alex"/>
      <join-date>06/01/2009</join-date>
    </editor>
        ...
  </editor-list>
  <!-- Selection and publication process -->
  <lifecycle-ref
    href="liquidpub.org/gelee/api/model/5"
  />
  <!-- Query definition -->
  <update-rate>real-time</update-rate>
  <query> GET papers, datasets, blogs
          FROM arXiv, Springer
          WHERE keyword = "peer␣review"
          ORDER BY num-citations
          GROUP BY source
  </query>
</liquid-journal>
```

As above, the rest of the APIs accept and provide resources in XML (and JSON), consumed by the web interface. For space restrictions we cannot put screenshots of the whole system, so we prepare a collage in Figure 5 showing the various features (editing, updates, tagging, lifecycle management, ...). Nevertheless, the best way to see a tool in action is by using it, so we invite the reader to try out latest beta version at `http://project.liquidpub.org/research-areas/liquid-journal`.

7. CONCLUSIONS

In this paper we have introduced a new liquid journal model for the web era. This model separates publication and dissemination of a scientific contribution, which in the model goes beyond the notion of a paper and includes, e.g., datasets, blogs. We have presented a liquid journal infrastructure, which supports editors in selecting content, and favors diversity of contributions.

The final quality of a liquid journal depends on how good the editors are in selecting content. Therefore, we propose to use a reputation system that would do proper credit attribution for editors by considering subscribers of the journal and articles included in the journal. Of course, as with any metrics there is a risk that people will try to tweak the system. At the moment, we do not know how serious this risk

[18] http://project.liquidpub.org/research-areas/liquid-journal
[19] http://project.liquidpub.org/gelee/
[20] http://project.liquidpub.org/reseval/

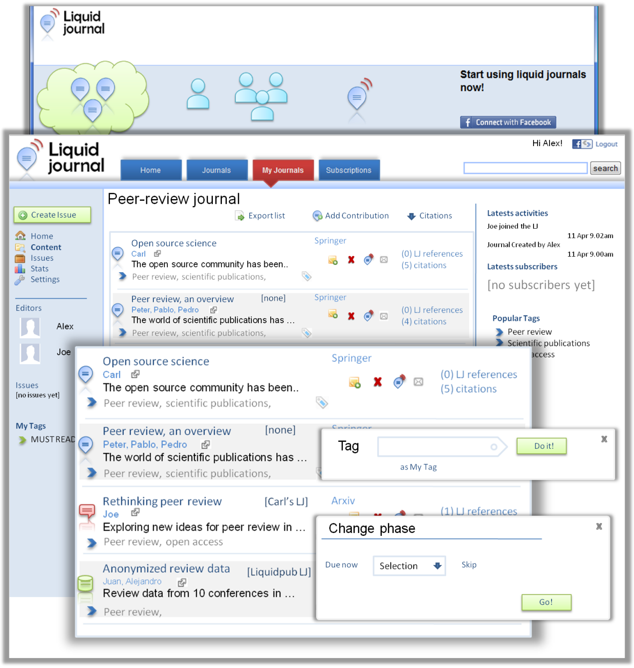

Figure 5: Collage of the various features of LJ

is, since we have not done research on this. However, such risk occurs also on the Web (fake Amazon reviews, promoting certain viewpoints in Wikipedia), yet, somehow it seems to work. In this line of research, we collaborate with IIIA-CSIC, Barcelona.

Search for relevant contributions is one of the most challenging parts. Here we rely on community-generated content about scientific articles (tags, links, etc.), but, this approach surely relies on how the liquid journal model starts up and get adopted by others. To facilitate the start of the model and attract more users a possible approach could be pre-populating some of the journals using similar in spirit initiatives available on the Internet.

The current model relies on traditional sources of publishing (conference proceedings, journals) as on sources of articles. One may argue that if the novel model of liquid journals is going to replace traditional journals, there will be no more sources of articles. However, liquid journals also take articles from online repositories such as eprints, and the tendency is that such repositories become more and more widely used (and might replace journals in the future).

Among the future and ongoing works we can mention addressing the scalability of the approach and the study of the implication of the liquid journal model on licensing, copyright, and business models of publishers. We are currently implementing a prototype supporting our model and the latest beta version is available at `http://project.liquidpub.org/research-areas/liquid-journal`.

8. ACKNOWLEDGMENTS

The authors would like to thank Claudio Bartolini, Azzurra Ragone, Joe Wakeling, Alejandro Mussi, Sheryl Leong and Florian Daniel for their continuous suggestions and comments to the ideas of this paper. This work has been supported by the EU ICT project LiquidPublication, under FET-Open grant number 213360.

9. REFERENCES

[1] M. Baez and F. Casati. Resource Space Management Systems. In *Proceedings of the European Conference on Web Services (ECOWS 2009).*

[2] M. Báez, F. Casati, and M. Marchese. Universal Resource Lifecycle Management. In *Proceedings ICDE.*

[3] P. Binfield. Plos one: background, future development, and article-level metrics. In *ELPUB 2009*, 2009.

[4] A. Bozzon, M. Brambilla, S. Ceri, P. Fraternali, and I. Manolescu. Liquid queries and liquid results in search computing. In *Search Computing Challenges and Directions*. Springer, 2010.

[5] R. Cuel, D. Ponte, and A. Rossi. How science 2.0 is affecting the scientific publishing industry: An analysis of the web2.0-based scientific knowledge production initiatives. 2009.

[6] P. Dall'Aglio. Peer review and journal models. *Arxiv preprint physics/0608307*, 2006.

[7] M. Franklin, A. Halevy, and D. Maier. From databases to dataspaces: a new abstraction for information management. *ACM Sigmod Record*, 34(4):33, 2005.

[8] F. Giunchiglia and R. C. Abente. Scientific knowledge objects v.1. Technical Report DISI-09-006, University of Trento, January 2009.

[9] F. Giunchiglia and R. ChenuAbente. Scientific Knowledge Objects v1. Technical Report.

[10] A. Haque and P. Ginsparg. Positional effects on citation and readership in arXiv. *Journal of the American Society for Information Science and Technology*, 60(11), 2009.

[11] B. Huberman, D. Romero, and F. Wu. Crowdsourcing, attention and productivity. *Journal of Information Science*, page 0165551509346786v1, 2009.

[12] M. Krapivin, M. Marchese, and F. Casati. Exploring and Understanding Citation-based Scientific Metrics. 2008.

[13] N. Lossau. Search engine technology and digital libraries. *D-Lib Magazine*, 10(6), 2004.

[14] M. Moyle and A. Lewis. Rioja (repository interface to overlaid journal archives) project: final report. 2008.

[15] S. Pinfield. Journals and repositories: an evolving relationship? *Learned Publishing*, 22(3):165–75, 2009.

[16] M. Simkin and V. Roychowdhury. Stochastic modeling of citation slips. *Scientometrics*, 62(3):367–384, 2005.

[17] J. Smith. The deconstructed journal–a new model for academic publishing. *Learned Publishing*, 12:79–91, 1999.