A Simple, Open, Scalable and Distributed Platform for Public Discourse

Thomas F. Gordon
Fraunhofer FOKUS

Abstract. [To be written]

Keywords. discourse systems, deliberation, e-democracy, e-participation, computer-supported collaborative argumentation (CSCA)

1. Problem Statement

There is some current debate about the relationships between "e-democracy", "e-government" and, more recently, "e-governance". The most widely accepted view, and the view we accept for the purposes of this paper, is that e-democracy is a subfield of e-government.

One of the main issues in the field of e-democracy if how to best use information and communications technology to facilitate public consultation, deliberation, participation or "engagement" in policy-making processes such as urban planning. A variety of discourse systems for the World Wide Web have been developed for this purpose [GeoMed, DEMOS, others]. Typically, these are client-server systems with a three-tier architecture: on the server side there is a web application which stores articles and other information in a database. On the client side, participants in the process access the system using a web browser. Figure 1 shows the basic architecture of this centralized type of participation system.

![Centralized architecture of typical participation platforms](image)

Figure 1. Centralized architecture of typical participation platforms

There are a number of problems with this approach:

**Scalability.** It is difficult to scale up to thousands of participants using replication and caching mechanisms, since all users access a single server and database and because
many views need to be generated dynamically to provide personalized views and support the transactions needed for interaction and participation.

**Motivation.** For several reasons, a centralized architecture is not optimal for facilitating large-scale participation. Participants need to access, learn and use a special purpose application, rather than being able to make contributions using some general purpose, familiar medium, comparable to a daily newspaper. And articles in a general purpose publication are likely to reach a wider audience and typically have the advantage of being archives in public libraries for a long time. Finally, the marketing potential of publishers is not mobilized for the participation process when the articles are published only a centralized participation server, rather as part of their own publications.

**Moderation Overhead.** Since the articles are stored and published on the participation server, the providers of the server must take full responsibility for the published content. In addition to the tasks of managing the participation process, moderators must take responsibility for checking and possibility editing every submission. The costs of moderation is by far the largest problem to overcome when trying to scale up to thousands of participants.

**Notification and Aggregation.** The centralized web server approach typically requires participants to manually browse the web site regularly in search of new articles of interest. Some systems make an effort to reduce this problem by enabling users to personalize the Web interface, in the manner of Web "portals". Other systems allow articles or reports to be "pushed" to the user's email address. Neither approach allows the user to fully aggregate and filter articles from many different channels.

**Accessibility ("Digital Divide").** Many citizens still not use computers, do not yet have access to the Internet or, even if they are regular computer and Internet users, the patience to learn a new user interface each time there is an opportunity to participate in some public discussion or deliberation process. These problems would be reduced if citizens could also participate via print media (e.g. newspapers) or web publications they already regularly read and use.

### 2. Solution Overview

The basic idea of the proposed solution to all of the problems identified above is to distribute the public discussion among existing print and web publications, making full use of letters to the editor, professionally written commentaries and other articles already being published via these media "channels". This approach raises the challenge of finding a way to tie all of these media together into coordinated and moderated, deliberation process. Three existing technologies provide the key to meeting this challenge:

**RSS.** Rich Site Summary (RSS) is an XML document type for news headlines. [BegeDov, 2001; Winer, 2001]. The headlines are published by a web site in an RSS file describing the new content of a "channel" published on the site. More than one channel can be published on each web site. Other sites and, more importantly, ordinary users, can "subscribe" to channels from one or more sites, and "aggregate" and filter the contents of these channels into a single, personalized news channel.
Using a RSS browser application, which could be a Web application or a separate desktop application, news items can be browsed and selected. Upon selecting a news item to read, the RSS browser asks the user's preferred web browser to retrieve and show the content of the full article.

**Autonomous Citation Indexing.** In newsgroups and discussion forums the concept of a "thread" is a tree of linked articles, where each article cites and replies to a single prior article. A "citation index" is a catalog of cross-reference information in a body of literature. Thus, an outline or table of contents of a thread can be considered to be a simple kind of citation index. Citation indexes are designed to facilitate information retrieval and allow navigation through the body of literature in many ways. Citation indexes reveal relationships between articles and can be used for various kinds of analysis, such as identifying trends and finding out how often a particular article or author has been cited. In the past, creating a citation index has been a labor intensive process. But recently autonomous citation systems, in particular CiteSeer [Lawrence, 1999], have been developed which can automatically create a citation index from articles in electronic formats. CiteSeer can "autonomously locate articles, extract citations, identify citations to the same article that occur in different formats, and identify the context of citations in the body of articles." The CiteSeer system has been used to build a fairly comprehensive online index for the World Wide Web of over 500,000 computer science articles [CiteSeer Scientific Literature Digital Library].

**Issue and Argumentation Mapping.** An important task is to summarize and visualize the discussion at various phases of the deliberation process, usually by applying some model of deliberation or argumentation, such as IBIS [Kunz, 1970] or Toulmin diagrams [Toulmin, 1958]. Using a centralized online discussion forum, this has typically been done by annotating the articles and modifying the threads of the discussion, by copying, moving or deleting articles. Of course, this approach cannot be applied in the proposed distributed model, since moderators will not usually have permission to change or restructure the original articles. The alternative proposed here is to use general purpose outlining, "idea processing" and diagramming software, such as Inspiration\(^1\) or Tinderbox\(^2\). Many if not most of these applications now provide some way to embed URLs and to export graphical maps to HTML, using the URLs to create links to the original articles, or to collections of (references to) original articles.

These three technologies are put together into a complete architecture for distributed deliberations as follows. The moderation team uses a Web site (the "participation server") to announce the process, calling for participation, and to subsequently post information about the state and status of the process, including summaries and maps of the discussion thus far. This participation server should also publish news about the process in RSS format. Participants do not register with the participation server or post articles there. Rather, the editors of online newspapers or other publications are

---

encouraged to register their RSS news feeds with the participation server. This could be done using an online form. Participants submit their contributions to one of the registered channels (i.e. web publications), presumably one they already read regularly, in the usual way. For printed newspapers, this could be done with a traditional (paper) letter to the editor, helping to bridge the digital divide (albeit via a media break). The editors of the participating publications are encouraged to promote the participation process in their publications, in particular the print version of the publication, if there is one. Authors of personal "weblogs" [Lasica, 2002] could also register their RSS channels with the participation server. This provides a way to submit articles directly, bypassing edited publications.

The publication server uses an "aggregator" to combine the channels and select articles dealing with topics relevant to the process. These articles are combined into a channel which is then published on the publication server, also in RSS format. This channel can be subscribed to by participants or anyone else. They would use their preferred aggregation applications to combine this channel with the other channels they read regularly, in order to be notified about and follow the parts of the discussion of interest to them. The RSS channel for the discussion would also be feed into an autonomous citation index, to create an online index. The index, which replaces the threads of traditional discussion forums, can be published on the participation server, or be "outsourced" to some other server.

Finally, the task of creating summaries and visualizations of the discussion would be delegated to professional "analysts", relieving moderators of this responsibility. The analysts need not have moderation or mediation skills. Their task is analyze, categorize and organize (references to) the articles, to "reconstruct" and visualize the arguments and issues of the debate. The analysts can use the diagramming and mapping tools of their choice, so long as it can export the map to HTML. The HTML maps could be published on the participation server by the moderation team. One advantage of this approach is that it allows for alternative discourse analyses. For example, each participation edited publication could create and publish its own discourse analysis. This is important, since discourse analysis is a highly interpretative task; multiple analyses are not only possible but to be expected.

3. System Design

Use Cases
[to be written]

Data Flow
[diagram below to be explained]
Components and Architecture

[diagram below to be explained]
4. Software Tools

5. Analysis and Evaluation

**Scalability**

**Efficiency**

**Personalization**

**Motivational Issues**

**Legal Issues**

**Business Model**

6. Related Work

7. Conclusion

References


