Fall 2006

01001100 01010100 01000101 01010010
LTER DataBits
Information Management Newsletter of
The Long Term Ecological Research Network
01001100 01010100 01000101 01010010

Feature Article: Three Challenges in Supporting Shared Workspaces

- Mason Kortz (CCE)

In the past months the Ocean Informatics team at Scripps Institute of Oceanography has worked to create shared workspaces for the PAL and CCE sites. For a discussion of the technologies behind these shared spaces, and their place in a collaborative infrastructure, see the Spring 2006 Databits issue (Kortz, 2006). The initial setup for these shared workspaces required design work. Below are some of the challenges we faced, along with our thoughts on how they could be solved, mitigated, or at least prepared for. The ideas below are not necessarily best practices, or even strong recommendations - they are rather local lessons learned about many challenges, and potential solutions, that exist in supporting shared workspaces.

1. Defining the purpose and scope of a shared space. In a collaborative environment, especially one with multiple shared workspaces, it is important to define the role of each shared space - its purpose and its scope. The purpose of a shared space may be archiving, dissemination, collaboration, or some combination of purposes. You should also consider whether policy or technology will limit use of your space to one type of content, such as raw data, documents, or images. The scope of a shared space is typically defined in terms of its user base. The user base may be a group of individuals, a pre-established organization, or simply 'everyone'.

Once the role of a shared space has been defined, that information must be made available to the user base. Providing a purpose for a space prevents the 'what-goes-where?' syndrome that often leads to shared spaces becoming disorganized, misused, or abandoned. Providing a scope helps users avoid posting sensitive material in a public space, or putting information in a place where the intended recipient cannot access it.

The process of defining shared spaces can also be illuminating for administrators. If two or more spaces are defined as having the same purpose and scope, then it may make sense to merge them into a single shared workspace. Alternatively, you may find that a particular purpose-scope combination is not present, which may prompt further questions about the use cases of shared workspaces in your environment.

2. Insuring integrity of shared information. In a shared workspace, many users may be working within the same space and editing the same information. Frequently, in shared environments, especially in those designed for collaborative work, users will have privileges that allow them to alter or remove the information created by other users. Because of this, there is an increased need to protect the integrity of shared information. An advanced shared space could support version control - the ability to track the addition, deletion, and editing of information, as well as to revert to a previous state if an unwanted change is made. In some shared spaces, the underlying technology offers this support; in other cases, an ad hoc change-tracking system can be implemented as a set of best practices.

Shared spaces that support version control at a technology level provide a robust solution for information integrity. Almost any change collision can be sorted out with no loss of information. However, such systems can be restrictive, in that the technology used defines what type of shared space is created. For example, website content management software can provide a version-controlled space, but only for the purpose of creating web pages. Further, version control technology creates additional overhead, as it essentially adds a new layer into any interaction with your shared space. Version control at the technology level makes the most sense for spaces with a focus on collaboration, where frequent change collisions are possible.

In some cases, implementing version control at the technology level seems to be over-engineering. Both in a space where collaboration is not the focus, and in a collaborative space with a small enough scope, a small set of best practices can be defined. Best practices that encourage users to avoid altering others' work and to leave a record of changes made essentially work as an ad hoc versioning system.

3. Maintaining consistent support for multiple users. A shared space, which by definition must support many users, can raise issues of consistency. One of the most obvious, but also potentially most frustrating, is the issue of supporting consistent access to your shared space. Because this depends on both the technology that supports your shared space and the platforms the user base uses to connect to the access the shared space, there are a great number

of variables to take into account. Testing access to a shared space before launching can catch many of these problems, but changes to the user base or your infrastructure are always possible.

Other issues with consistency arise from practices, rather than technology. With many people accessing the same space, different choices will be made in terms of organization and nomenclature. Enforcing organizational structure can be difficult, and with smaller shared spaces it often isn't necessary. When it is necessary, it is best to anticipate the needs of the space and create the structure ahead of time, rather than relying on users to create it as they go. Similarly, there are many cases in which enforcing nomenclature is unnecessary, as long as information is readable to the user browsing it. In cases where information may need to be sorted or searched by a program, however, a strict nomenclature should be used.

Example: WebDAV Share. Many of the issues above were encountered while setting up a WebDAV file sharing space for both the Palmer and California Current Ecosystem sites. We began by considering the purpose of the shared space. We needed a space where the entire research community could work jointly on various projects. As such we defined the purpose as quick, easy collaboration and file exchange. Because of technological limitations with WebDAV, we did not have the option for a granular permission system, so we added the caveat that this space would not be used for sensitive material. We expanded this purpose to include use of a general dropbox, by which researchers could get data, publications, and other information to other researchers and the information management team. We defined the scope as all participants associated with a site (i.e. researchers, graduate students, technicians, outreach, and administrative staff).

Next, we considered what we could do to ensure data integrity. The WebDAV server was already part of our backup system, so disaster recovery was covered. For short-term data integrity, we considered software version control through the application Subversion, but decided that the additional layer of complexity did not fit the 'quick and easy' formulation of the shared space's purpose. Instead, we decided to implement a few best practices. First, when editing another user's information, do not overwrite their work - create your own copy of the resource and edit that. Secondly, when making additions or edits, include your name and a timestamp (using a specified format) in the file name, so that other users can quickly identify the most recent version of a shared resource. Thirdly, don't delete shared files; instead, move them to the 'to be deleted' space, which is regularly cleared of files older than one month.

With these ideas in place, we implemented the shared space. Early on we

noticed a cross-platform issue: Macintosh OS X users could upload files, but not folders. We realized that the server was blocking the hidden .DS_Store files in the Mac folder, causing the copy operation to fail. We also added some structure to the shared space, creating four directories at the top level: data, metadata, individual, and other. We created further structure in the data and metadata directories so that researchers would know where to upload files, and created a convention for uploaded data file names. The other directories we left as an open area without any enforced structure or nomenclature. Finally, we created a README file in the top level of the share to detail the purpose and scope of the shared space as we had defined it, as well as the best practices and conventions we had created. The file also lists the names and email addresses of the users that with access to the share. This README file also contains instructions on connecting to the shared space, and is emailed out to new users when they are granted access.

The WebDAV spaces, as well other shared workspaces, are now part of the PAL and CCE information infrastructure that supports collaborative work. As we both expand and refine our support for collaborative work, new challenges arise that are both technical and social in nature. As with many shared tools, technical foresight and explicit best practices - both gained from our experiences in the past - help to ease the introduction of these shared resources into scientific practice.