Chapter 5

Pitfalls and What Not to Do

Cloud services can be used to build a professional-level product, but a lack of knowledge regarding their proper use can lead to difficulties. Even with proper attention and consideration, problems can arise during development. Here are some potential challenges to prepare for when using cloud resources.

Stick to Resources That Suit the Scope of Your Collection

Cutting-edge cloud services can provide engaging experiences and exciting visuals, but they do not need to be used for every collection. They can be expensive, time-consuming to implement, and require expertise beyond what is found in typical IT teams. Some cloud resources may seem obviously out of scope for most repositories; quantum computing, neural machine learning, satellite communication, and other next-generation computing resources may be intriguing but are perhaps not necessary for meeting the needs of most repository users. When creating a collection, look to include resources that will not overextend the capabilities of the team responsible for its creation. For example, managing video resources requires more work than simply hosting still images. If the effort needed to provide video digitization, playback, and transcription seems large compared to the number of videos intended for inclusion in the repository, consider prioritizing the addition of still images before working on the video. If the collection features a large audio component, focus on delivering a rich set of playback tools before expanding into other areas. An advantage of cloud resources is that they can often be easily added to a project after its creation, making it possible to hold off on adding new features until a later time.

Cloud Service Vulnerabilities

Cloud services have teams of trained professionals maintaining their virtual and physical security. Yet for all the precautions in place, they are not invulnerable to hackers, disasters, or business and political forces. With some cloud service providers, the security of a cloud-hosted service is left to the user, and therefore is only as secure as the capabilities of the project developers and systems administrators. With other providers, security is fully managed, and while this can be a great asset for those who are not knowledgeable in this area of IT, it does require trust that the company providing the service is taking the appropriate measures to prevent attacks to its servers. There are other potential risks as well; a company that is providing any resource crucial to the operation of a repository can financially collapse, experience service-disrupting disasters, or be barred from doing business in a country due to changing laws and political influence. In these cases, off-site backups of repository data, source code, and detailed setup documentation may be the only safeguards against a total loss of the project. However, in most cases, the security solutions offered by cloud service providers should be enough to keep a repository safe from data loss and secure from the majority of malicious attackers.

Manage Costs Carefully

Without precise understanding of billing structures and resource costs, a cloud solution can easily become expensive; with the vast number of options available to choose from, the cost of cloud infrastructure can increase wildly if left unchecked. Data redundancy features, high-powered servers, and resources that scale automatically can all incur costs quickly and

uncontrollably. Even an increase of traffic to a repository can increase its operational costs; not only will an institution be billed for the resultant resource utilization, but potentially for the increased inbound and outbound data transfer as well. The simplicity of using cloud services also poses a risk to cost stability, as it can be very easy to provision a high-powered, cutting-edge stack of cloud resources that may not immediately appear on a billing cycle.

A thorough understanding of chosen cloud resources will help when budgeting for a repository project. Knowing the expected behavior and pricing options of the services in use can keep a repository designer from making choices that may cost more than expected after the project is deployed. Knowing how much speed is truly required for a repository can make a large difference financially, as the cost difference between various cloud servers and databases can be significant. It is also possible to put caps or limits on certain resources to keep them from growing too large or using too much bandwidth. Sophisticated users can even use timed events to have resources increase their speed and performance on a predetermined schedule, thereby paying only for the necessary resources at the necessary times.

It may be helpful to implement some measure to help prevent unnecessary expenses. Some providers offer cost alarms for their suite of services—e-mails, text messages, or other notifications can be triggered when a monthly spend surpasses a chosen amount. Setting a threshold somewhat lower than the absolute top of the budget can allow a designer to quickly make changes before costs get out of hand while maintaining continuous uptime. These alarms can also provide insight to an institution for long-term budgeting and can provide extra information regarding resource utilization when working on the next version of the repository or when designing other projects.

Conclusion

The creation of a digital repository can be a complex, ambitious undertaking. Ideally, a repository must thoughtfully and accurately display the depth and variety of resources that it contains, and it must possess flexibility to accommodate new resources and collections as they are added. It must be technologically capable of providing audio and video representations of cultural artifacts, academic scholarship, and documents of historical relevance. It must strive to meet the standards of academia, government entities, and other regulatory organizations; be responsive to the needs of its users; and be accessible to all people regardless of their abilities. The ideal repository must be a searchable, extensible tool, ready to scale and change along with its user base.

Building a repository requires its creators to satisfy a long list of requirements. There can be stakeholders at multiple points in the creation process, from design to deployment, each with unique requirements. Design can be hindered or helped by budgetary constraints, institutional regulations, and staffing resources. The process will vary drastically depending on the available technical knowledge. The project may seem financially unfeasible, or the technology required may seem out of reach for individuals and smaller institutions. Just maintaining a repository will require an ongoing commitment that may appear too large for some to assume. And without a guide or a starting point, the entire process may feel like a goal existing only in the distance, too great to achieve.

Creating a repository can seem daunting, especially when faced with the challenges of social distancing and remote working requirements. Without office collaboration, in-person IT support, and access to traditional server rooms, it may seem impossible to even start such a project, much less see it to completion. Fortunately, cloud technology provides the tools to make it possible for any library or other institution to create a repository, built to unique, precise specifications. Cloud technology not only allows for the creation of repositories under these circumstances, it provides the potential to create a repository with capabilities beyond what was possible before. It is a valuable set of resources and should be seriously considered when building a repository or any IT project for a library.