THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

Guidance for Researchers Using Internet Cloud Computing Services and Apps

Table of Contents

Introduction

Researchers use cloud computing daily, in their private and professional lives, when using email (e.g., Gmail, Hotmail), social media (e.g., Facebook, Twitter), online storage (e.g., iCloud, DropBox), survey products (e.g., SurveyMonkey), real-time communications (e.g., WebEx, Skype), collaborative projects (e.g., Google Docs), or mobile apps. Broadly conceptualized, cloud computing may refer to third party online computing services accessed over the Internet (for formal definitions, see NIST SP 800-145 "Definition of Cloud Computing" – link in Appendix below). While cloud computing may also refer to deployments that are within an institution's firewall (private cloud) or combine intra-institutional systems with external networked computing systems that share connectivity over internet protocols (hybrid cloud), this guidance focuses on Internet cloud computing services or apps provided by third parties, often referred to as "Software As A Service."

While personal use of these services offers efficiency and convenience, their use for storing or processing institutional research data, especially individually identifiable research data, carries risks because the institution has no direct contractual control over information stored or processed in personal accounts on third party servers. Unauthorized uses or disclosures of electronic information may violate an individual's privacy, causing psychological or other types of harm, as well as creating an increased financial risk, due to the threat of identity theft. In addition to the risks to individuals, the institution where the researcher is affiliated may also be at risk of legal, regulatory, or reputational harm in the event of data misuse.

This document offers key points to consider when using cloud services, and addresses steps to take and controls to put in place in order to mitigate risks and alert appropriate parties in the event of data loss, breach, or other unlawful or unethical use or disclosure.

Cloud Computing: Setting Up Applications In the Cloud

In addition to "Software As A Service" applications that are offered by vendors directly in the cloud, such as the examples provided above, applications that have historically been run by an institution "inside the firewall" of the institution can be migrated to or set up in the cloud. That use of the cloud is often called "Infrastructure as a Service" because it is an alternative to building out data center computing systems.

Your local institution may offer secure research computing systems that are appropriate for your research project and do not require a cloud solution.

The following situations have been found to generally favor a use of the cloud for Infrastructure as a Service*:

- A project has high costs for computing, administration, space, and electric power in its current or envisioned state.

- A project requires variable amounts of processing and storage resources where the scalability benefits of cloud computing are fully utilized.
- A project has a heavy focus on sharing data with outsiders, and institutional security policies block outsiders' access to their local system.
- A project's resources (e.g. people) are limited where the success of your project depends on their time and effort being used for mission or front-facing implementation, rather than the ongoing maintenance and support of the IT infrastructure.
- A system requires off-site backups for data and/or archive.

* These examples should only be used as a guide and you should consider exceptions based on the specific circumstances of an institution's IT architecture.

Key Considerations

- **Responsibility for cloud implementation is shared between institutions and a cloud provider and therefore cannot be fully outsourced.** Most consumer-based cloud services and apps were not designed with research regulatory compliance or human subject protection ethics or considerations in mind. Cloud services for typical consumer use ordinarily have inflexible terms of use or non-negotiable end-user license agreements. Therefore, it is the responsibility of the researcher to identify and implement appropriate supplemental privacy and security safeguards into the protocol or research project when using these services. Staff with IT and data management and security expertise should be sought.
- **Data hosting services should comply with grant terms**. Researchers should check with their office of sponsored programs and grants office to determine if the institution has an operating agreement, service-level agreement, or other contractual arrangement with a specific cloud service provider or app before proceeding to use these services. If no such agreement or arrangement is in place, the researchers and institutional offices should work together to determine how to comply with grant terms.
- **Ownership of Data**. Cloud provider terms of service often stipulate that data handled or stored in accounts on the provider's servers becomes their property. To protect institutional data, researchers should consult their institution's intellectual property office, general counsel's office, or equivalent to ensure that institutional contracts and accounts with cloud service providers affirm institutional ownership of data and the right to retrieve and eradicate data from the provider's systems.
- **Anticipate current and future access to the data.** Are these uses compatible with the access controls of the cloud services provider or app? If you are using a personal account to store and share files and are the only "access administrator" for those files, what happens if you leave the institution? Will the institution lose access to all data, as well as the ability to manage others' access to those files?
- **Data deletion or retrieval.** Depending on the third party vendor's policies or willingness to negotiate, a researcher or institution may not be able to retrieve or delete sensitive information transmitted to the vendor.
- **Identifiable research data and institutional liability.** If sensitive data is identifiable, or

re-identifiable using other publicly-accessible data points, then exposure of elements of your research data to the public Internet by a cloud provider or app may subject the institution to legal sanctions. For example, a researcher's incorrect configuration of a folder as public versus private (the default setting in many of these consumer services) could expose data elements and the URL to access a file within general Internet search engine query results, such as a Google search.

Operational, legal and contractual issues

1. **Licensing and terms of use.** When you sign up to use third party services, whether they are free, low cost, or regularly priced, you may be agreeing to terms and conditions, terms of service, and acceptable use policies that are different from those of your institution. Personal licensing is between you and the service provider; they have not been reviewed and approved by your institution for official use. The service provider can attempt to hold you to what you agree to, even if it is just a "click-to-accept"-type agreement. Do you have delegated authority to enter into this type of agreement on behalf of your institution? If not, you may be in violation of institutional policy if you "click-to-accept" the terms of use.
2. **Data ownership.** It is essential to ensure that ownership of institutional data remains with the institution. Whenever you put data on a commercial service, ensure that the terms do not conflict with institutional policy or governmental contracts and grants in terms of data ownership.
3. **Institutional compliance.** Keep in mind that you may be required by the university to produce records relating to institutional business, including email, instant messages, files, etc., regardless of whether those records are stored on institutional or non-institutional systems or services. Using a cloud app does not relieve you of this obligation but may make it more difficult for you to comply.
4. **Data/account deletion.** There is no guarantee that deleted content or accounts will be eradicated completely from all systems. It may take weeks or longer before the content and the account are thoroughly flushed from all of the company's archives. Practices will also vary as to how long idle accounts may remain "active" before the account and associated data are flagged for automated deletion.
5. **Recourse.** If the service is free or if the end user license agreement or terms of use are not subject to negotiation, you probably have little or no recourse against the vendor if something goes wrong or they change their privacy or security practices.

Incident response and mitigation

- **Know your data stakeholders.** Keep at hand the contact information of your research subjects, internal and external research collaborators, study team and other research sites in the event of an incident where research data may be compromised, lost, stolen, or subjected to unauthorized use.
- **Know the incident response policy of your institution and the office's contact information** or other offices that need to be notified in the event of an information

incident.
- **Consider training** of staff on incident response and mitigation measures

Acknowledgements and Further Reading

- Content adapted from UC Irvine's Guidance on the Use of Cloud Services, which relies on UCLA's Guidance on the Use of Cloud Apps by Individuals, UC Santa Cruz's Use of Free Services, with additional input from UC Berkeley and Lawrence Berkeley Lab.

APPENDIX A

Brubacher, Steve (U. Wisconsin Milwaukee), Practical Strategies for Managing Risk in Cloud Environments https://net.educause.edu/ir/library/pdf/ERB1103.pdf

U.C. Irvine Guidance for Use of Cloud Services http://www.security.uci.edu/cloud.php

Michigan State – Appropriate Use of Online Software Tools
https://itservices.msu.edu/guidelines-policies/cloud-computing-appropriate-use.html

Internet2 - Cloud Data Storage Solutions: Dropbox Security & Privacy Considerations
https://spaces.internet2.edu/display/2014infosecurityguide/Cloud+Data+Storage+Solutions

NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing Policy (GDS Policy)
http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf
NIST definition of "Cloud" Special Publication 800-145

National Institute of Standards and Technology (NIST) Definition of Cloud Computing
http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Architecting for Genomic Data Security and Compliance in AWS
https://d0.awsstatic.com/whitepapers/compliance/AWS_dBGaP_Genomics_on_AWS_Best_Practices.pdf

Other NIST resources:

- NIST SP 800-12, An Introduction to Computer Security: The NIST Handbook
- NIST SP 800-14, Generally Accepted Principles and Practices for Security Information Technology Systems
- NIST SP 800-16, Information Technology Security Training Requirements
- NIST SP 800-18, Guide for Developing Security Plans for Information Technology Systems
- NIST SP 800-26, Security Self Assessment Guide for Information Technology Systems
- NIST SP 800-27, Engineering Principles for Information Technology Security (A Baseline for Achieving Security)
- NIST SP 800-30, Risk Management Guide for Information Technology Systems
- NIST SP 800-34, Contingency Planning Guide for IT Systems
- NIST SP 800-37, Guide for Security Certification and Accreditation of Federal Information Systems
- NIST SP 800-44, Guidelines on Security Public Web Servers

- NIST SP 800-47, Security Guide for Interconnecting Information Technology Systems
- NIST SP 800-53 Revision 4, Recommended Security Controls for Federal Information Systems
- NIST SP 800-60 Vol. 1 & 2, Guide for Mapping Types of Information and Information Systems to Security Categories
- NIST SP 800-63, Electronic Authentication Guideline: Recommendation of the National Institute of Standards and Technology
- NIST SP 70, The NIST Security Configuration Checklists Program
- NIST SP 800-81, Secure Domain Name System (DNS) Deployment Guide
- NIST SP 800-83, Guide to Malware Incident Prevention and Handling
- NIST SP 800-86, Guide to Integrating Forensic Techniques Into Incident Response
- NIST SP 800-92, Guide for Computer Security Log Management
- NIST SP 800-95, Guide for Secure Web Services
- NIST SP 800-97, Guide to IEEE 802.111: Establishing Robust Security Networks (this is related to wireless network deployment)

Attribution, Sharing and Adapting

We encourage broad dissemination of this guidance document, and incorporation of these practices into clinical trial operations. We would also appreciate feedback and additional contributions so that we can continuously improve this work product.

We encourage you to:

- **request** — email us and request the materials
- **share** — copy, distribute, and transmit the work
- **adapt** — adapt the work to suit your needs

Under the following conditions:

- **Attribution**: In freely using the materials, we require that you acknowledge Harvard Catalyst as the publisher and that you give appropriate credit to any named individual authors.

- **Suggested citation:** *This material is the work the Harvard Catalyst Data Protection Taskforce and subcommittee of the Regulatory Foundations, Ethics, and Law Program. This work was conducted with support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health Award 1UL1 TR001102-01 and financial contributions from Harvard University and its affiliated academic health care centers). The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.*

With the understanding that:

- **We might contact you**: We are interested in gathering information regarding who is using the material and how they are using it. We may contact you by email to solicit information on how you have used the materials or to request collaboration or input on future activities.
- **When reusing or distributing, make clear the above terms**: For any reuse or distribution, you must make clear to others the terms of this work. The best way to do this is with a link to the web page containing this guide.
- **When adapting**: Please share improvements to the tool back with us so that we may learn and improve our materials as well.

CONTACT US

Copies of all materials are freely available. Please send your requests, questions and comments to *regulatory@catalyst.harvard.edu* and visit the Harvard Catalyst Data Protection Subcommittee web page.