Trustworthy Digital Repositories

A risk analysis for the Purdue University Research Repository (PURR)

Pascal Meunier

Greater Lafayette Security Pros, August 9, 2012

Trustworthy Digital Repositories

- Need for digital repositories
  - Something responsible for the long-term access to the world's social, economic, cultural, and intellectual heritage in digital form
  - Needs vary by community
- What makes a digital repository trustworthy?
  - Self-declaration not acceptable
  - Types of trust (direct, transitive, assumptive)
  - Measure and justification for trust: Assurance

Trustworthy Digital Repositories

- How do you know if a repository is trustworthy?
  - Is it a good idea to put your documents there?
  - Will you get back an exact copy?
  - Will everyone be able to access it easily, ”forever”?
  - Does it serve your needs?
  - What should a repository do, and what guarantees and processes should it use?
- Direct trust is burdensome
  - Need expertise
  - Need time and access

Open Archival Information System (OAIS)

- Effort to openly articulate what is responsible digital archiving, and the infrastructure needed
  - Defined terms and concepts
    - Archive == repository
    - Key: Designated Community
    - Key: Long term preservation
  - ISO 14721
- Specify what an AIS should do and should look like, in abstract form

Key Definitions

- Digital Archive: An organization that intends to preserve information for access and use by a Designated Community.
- Designated Community: An identified group of potential Consumers
- Consumers: Entities that find and access information in detail
- Long Term: Long enough for technologies, formats, media, and communities to change

Trustworthy Digital Repositories

- Organizations started calling themselves ”OAIS-compliant”
- How do you know if a specific AIS is a good implementation of the OAIS reference model?
- Answer: Certify it!
  - How?
    - Write a standard for it
      - This is the one I want to discuss today...
  - Who can certify?
    - Write a standard for it

Trustworthy Digital Repositories

- ISO Standards
  - Not free
- ”Recommended Practice” documents
  - Council of the Consultative Committee for Space Data Systems
  - ”Magenta books”
    - Hosted at the CCSDS web sites
    - Free

Trustworthy Digital Repositories

- ISO 16363 Standard for Trusted Digital Repositories
  - Criteria a trustworthy digital repository should meet
  - Not free, so we do without
  - Magenta Book:
    - CCSDS 652.0-M-1

Trustworthy Digital Repositories

- ISO/DIS 16919
  - Criteria for carrying out audits
    - Requirements to be an auditor
  - Not free
  - Magenta book equivalent:
    - ”Audit and Certification of Trustworthy Digital Repositories”
    - CCSDS 652.1-M-1

Trustworthy Digital Repositories

- Other documents
  - ISO 27001
    - Information Security Management Systems
    - Certification
  - ISO 27002
    - Security Techniques
  - Risk analyses for other trustworthy digital repositories

Trustworthy Digital Repositories

- Scholar's Portal
  - Trustworthy digital repository
  - Ontario Council of University Libraries
- Risk analysis available online
  - Draft version at http://spotdocs.scholarsportal.info/display/OAIS/Risk+Analysis+and+Management+Strategies

Magenta Book

- For the purposes of this presentation, we mean the one titled ”Audit and Certification of Trustworthy Digital Repositories”
  - http://public.ccsds.org/publications/archive/652x0m1.pdf
- Many more requirement types than InfoSec

Magenta Book Contents

- Organizational Infrastructure
  - Organizational viability
  - Financial
  - Contracts, licenses, liability

} Business plan

- Digital Object Management
- Infrastructure and Security Risk Management
- Annex: Security Considerations

Magenta Book Requirements

- Hierarchies of requirements
  - 3.5 Contracts, Licenses, Liabilities
  - 3.5.1 ”The repository shall have and maintain appropriate contracts or deposit agreements for digital materials that it manages, preserves, and/or to which it provides access.”
    - 3.5.1.1 ”The repository shall have contracts or deposit agreements which specify and transfer all necessary preservation rights, and those rights transferred shall be documented.”

Why Do Universities Care?

- NSF mandate for Data Management Plans
  - National Science Foundation Grant Proposal Guide (GPG) Chapter II.C.2.j, available at
    - http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp
- Universities have big library systems
  - Need to archive theses
- Move for publications to become digitally archived instead of in paper books
  - Journals make us pay private parties to access publicly-funded research

PURR

- Purdue University Research Repository
- Can we give an advantage to our faculty with a HUB that's also a Trustworthy Digital Repository?
  - HUBzero platform developed here at Purdue
    - nanoHUB
    - NEEShub
    - cceHUB
  - Idea: Let's support projects, incl cooperations with it
    - Host the project's data, unpublished manuscripts

Magenta Book Applied to PURR

- ”3.5.1.3 The repository shall have written policies that indicate when it accepts preservation responsibility for contents of each set of submitted data objects.”
- Does PURR want HIPAA-covered data sets?
  - No.
- Other regulated content?
  - No
- So the data submission form asks questions

Magenta Book Requirements

- Confidentiality is not ever mentioned
  - Implied for authentication secrets
    - Public resource, so for administrators
- Integrity
  - Hash of files to be stored
  - Hash recalculated periodically and compared
- Availability
  - Worried about all the things that could go wrong and decrease the quality of service

PURR

- PURR is much more complex than a regular TDR
- PURR has unique confidentiality requirements
  - Host unpublished data
    - Controlled sharing
    - Archival
  - Works in progress
    - Discussions about
  - Serve faculty and collaborators
  - As well as a public archive

Magenta Book

- 5. Infrastructure and Security Risk Management
- 5.1. Technical Infrastructure Risk Management
  - 5.1.1. The repository shall identify and manage the risks to its preservation operations and goals associated with system infrastructure.
    - 5.1.1.1 The repository shall employ technology watches or other technology monitoring notification systems.
    - 5.1.1.2 The repository shall have adequate hardware and software support for backup functionality sufficient for preserving the repository content and tracking repository functions.

Magenta Book

- 5.1.1.3 The repository shall have effective mechanisms to detect bit corruption or loss
  - e.g., SHA256 hash of uploaded documents
    - Stored in database
    - Verified periodically
- 5.1.1.4 The repository shall have a process to record and react to the availability of new security updates based on a risk-benefit assessment.

Magenta Book

- 5.1.1.5 The repository shall have defined processes for storage media and/or hardware change (e.g., refreshing, migration).
- 5.1.1.6 The repository shall have identified and documented critical processes that affect its ability to comply with its mandatory responsibilities.

Magenta Book

- 5.2. Security Risk Management
- 5.2.1 The repository shall maintain a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant.
  - Risk Analysis
- 5.2.2 The repository shall have implemented controls to adequately address each of the defined security risks.

Magenta Book

- 5.2.3 The repository staff shall have delineated roles, responsibilities, and authorizations related to implementing changes within the system
- 5.2.4 The repository shall have suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an offsite copy of the recovery plan(s).
  - This is interesting given that ITSO has stopped using tape backups...

Magenta Book

- 6. Other Considerations
  - Who is the auditor
    - Defend against social engineering
  - Safe handling of audit results
- 4.6.1.1 The repository shall log and review all access management failures and anomalies
  - Concerns regarding the identification of security threats by reviewing logs

Digital Object Identifiers

- DOIs
  - Permanent reference to data or publication
    - Prefix
      - Identify the owners (”registrant”)
    - Suffix
      - Identify the object
  - Can have a URL that changes
    - Availability
    - Integrity
      - What if the URLs are maliciously altered?
    - No confidentiality

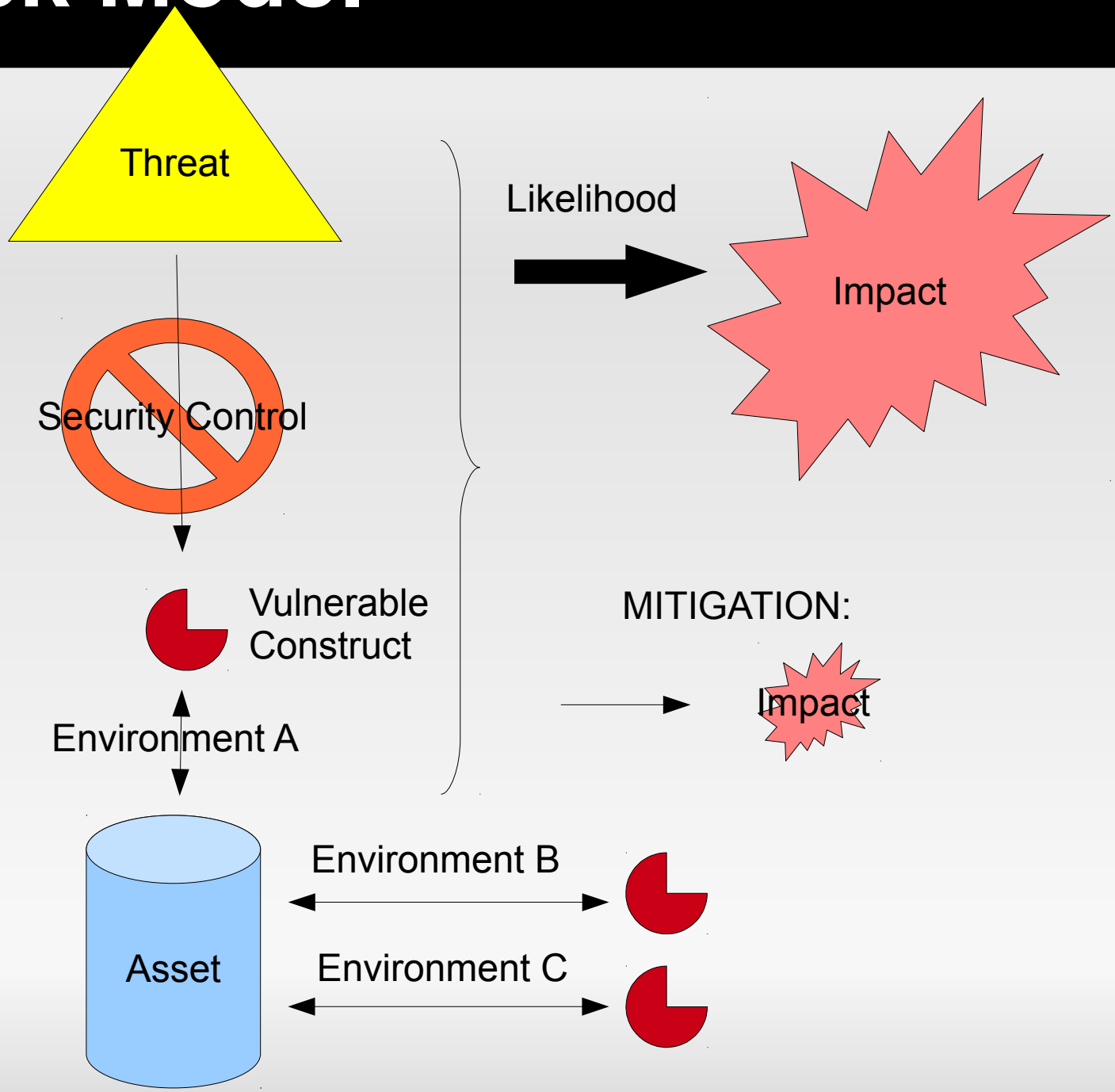
Threat Typologies

- Needed for risk analysis
- Several in the area of digital curation
- Scholar's Portal Types:
  - Economic, political, social, or legal threats
  - Technology-related failures
  - Man-made threats
  - Natural threats
  - Utility or environmental/building systems failures

PURR Threat Types

- Dimensions of PURR
  - People
  - Technologies
  - Environment
    - Cyber
    - Physical
    - Legal
    - Financial
    - Educational
    - Political

PURR Risk Model



Conclusions

- AISes based on the reference model of OAIS have been around for about a decade
- Certification is new this year (2012)
- Risk management approach overlaps HIPAA's
  - More emphasis on availability, little confidentiality
- PURR breaks new ground
  - It's also a HUB
    - Unique confidentiality requirements
    - Additional complexity and security challenges