Infrastructure Policies Workspace

About

I am using this space to reorganize the Checklist requirements into meaningful categories (policies/documentation, staffing, workflow, technical infrastructure). The next step will be to determine which of the requirements we currently meet and which we need to focus on.

- Repository Preservation System wiki page
- RLG's Audit Checklist for the Certification of Trusted Digital Repositories
- D-Lib 12(12), Using the Audit Checklist for the Certification of a Trusted Digital Repository as a Framework for Evaluating Repository Software Applications

To do

- Look for integrity checking info
    - No useful info on how often one should run data integrity checks - the only references to it I could find used words like "regularly" and "frequently".
- Write up what info we need from MDSS
- Write up discussion questions for infrastructure meeting

Policies/Documentation

- ✅ Mission Statement (A1.1)
    - "reflects a commitment to the long-term retention of, management of, and access to digital information on behalf of depositors"
        - DLP: *The Indiana University Digital Library Program (DLP) is dedicated to the production, maintenance, delivery, and preservation of a wide range of high-quality networked information resources for scholars and students at Indiana University and elsewhere. The program supports efforts to provide open access to electronic information resources to the Indiana University community and beyond. The Digital Library Program is a collaborative effort of the Indiana University Libraries, the Office of the Vice President for Information Technology, and the university research faculty with leadership from the School of Library and Information Science and the School of Informatics. This collaboration capitalizes on the institutional capabilities of this university, focusing university resources on digital library services and projects that support the teaching and research of IU faculty, support the learning and research of IU students, and foster research about the digital library.*
        - DLP: *Strategic Area #12: Enhance preservation of digital collections and metadata:*
            - *Develop plans to insure the preservation of created digital content and metadata*
            - *Lead efforts to implement these plans*
            - *Explore and test methods of digital preservation*
        - MDSS: *The Distributed Storage Services Group (DSSG) provides a scalable, network accessible, standards-based storage infrastructure to support teaching, research, and administrative computing. Storage services at IU consist of a global file system for data sharing and storage for general research purposes, and the Massive Data Storage System (MDSS), used to store vast amounts of archival or near line data on a hierarchy of storage media.*
    - accessible to depositors and other stakeholders

Reviews/Certification/Ongoing Documentation (**Stacy and Jon/Infrastructure meeting**)

- Mechanism for reviewing and updating policies (A3.1)
- Commitment to formal, periodic reviews (A3.3)
- Document changes to operations, procedures, software, and hardware (A3.4)
- Document a change management process that includes creating records of changes made (D1.8)
    - Define process for testing the effect of critical changes to the system (D1.9)
- Commitment to regular certification and to notify certifying bodies "of operational changes that will change or nullify" certification status (A3.7)
- Policy about how and when to test understandability (C4.1, C4.2)

Business Planning (**Stacy and Jon**)

- Formal succession plan, contingency plans, and/or escrow arrangements (A1.2)
- Short- and long-term business plans to support sustainability (and regular planning process) (A4.1)
    - Business planning practices must be "transparent, compliant with relevant accounting standards and practices, and auditable." (A4.3)
- Commitment to "risk, benefit, investment, and expenditure analysis and reporting (including assets, licenses, and liabilities). (A4.4)
- Commitment to securing funds if necessary to maintain repository (A4.5)

One-liners

- Commitment to professional development
- Commitment to "transparency and accountability in all actions supporting the operation and management of the repository" (A3.5)
- Policy for staying current with security fixes (D1.10)

Depositors (**Michelle and Stacy**)

- Maintain contracts or deposit agreements with depositors as appropriate (A5.1)
    - Contracts, agreements, or documentation "specify and/or transfer appropriate preservation rights, as necessary." (A5.2)
    - Track copyrights and restrictions on use specified in agreements with depositors (A5.3)
    - Policy addressing situations where no ownership/rights specified
    - Agreements specify "all appropriate aspects of acquisition, maintenance, access, and withdrawl" (B1.2)
    - Agreements "specify exactly what digital object(s) are transferred, what documentation is associated with the object" (B1.3)

Designated Community (**infrastructure meeting**)

- Written definition of the designated community/ies - who it is, what it is knowledge base is, what levels of service it expects. (C1.1)
    - This definition should be publicly available. (C1.2)
    - The repository commits to a certain level of "understandability" with the community in question (for example, if we commit to ensure that the content is 100% human-understandable, we will need to migrate to new formats as the community changes to use of those formats) (C1.3)
        - Must include a definition of the designated community's application tools that are to use the information (e.g. the community only uses Microsoft Word)
        - Must communicate with designated community what access and delivery options are available (C3.1)
- Procedure for monitoring or receiving notifications about changes in the needs of the designated community/ies (D2.3)

Minimum Metadata

- System **Ryan's working on it**
    - Identify properties to preserve for each class of digital object (B1.1)
    - Identify which pieces of descriptive metadata must be provided and who is responsible (B4.1)
    - Written definitions
        - For each SIP or class of information ingested (B1.3)
        - For each AIP or class of information preserved (B2.1)
- Designated community/ies **?**
    - Must articulate minimum metadata requirements for discovery (C2.1)

Technical Workflow Policies (**Jon and Stacy**)

- Policy addressing processes "to ensure that the information is acquired from the expected source" (B1.4)
    - Policy stating we will require authentication on all deposits
- Commitment to define, collect, track and provide information integrity measurements (chain of custody, checksums, etc.) (A3.6)
- Policy about what access information should be recorded (what is accessed, who is accessing, etc.) (C3.2)
- Policy for staying current with security fixes (D1.10)

Preservation Policies/ Documentation (**Ryan**)

- Documented preservation strategies (B3.1)
- Documented preservation planning (B3.11)
- Document strategies for AIP storage and migration (B3.2)
- Policy on copies (number, location, etc.) and backups (D1.2, D1.3)

Disaster/Loss Policies (**Ryan working on it**)

- Policy on what actions to take when data is corrupted or lost (D1.6)
- Written disaster preparedness and recovery plan(s), including at least one off-site copy of all deposited data (D3.4)
    - Policy of regular testing of disaster plans (D3.5)
    - Define processes for service continuity and disaster recovery (D3.6)

Technical Infrastructure

Access Rights (**Ryan**)

- Track and enforce copyrights and restrictions on use specified in agreements with depositors using appropriate authentication measures (A5.3, B5.1, C3.3, C3.4)
    - Log all access management failures and review inappropriate access denial incidents (B5.2)
    - Ensure that users that are allowed access to objects can access them in their entirety or that they get a response telling them that part or all of their request cannot be filled. (B5.3, B5.5)
- Ensure that "all access requests result in a response of acceptance or rejection." (B5.3, B5.5)
- Implement the access recording policy (C3.2)

Ingest/Submission (**Ryan and Muzzo**)

- Processes "to ensure that the information is acquired from the expected source" (B1.4)
- Ingest process verifies each SIP for completeness and correctness
- "Repository provides Producer/depositor with appropriate responses at predefined points during the ingest process." (B1.7)
- Demonstrate when preservation responsibility is formally accepted for the contents of the SIP (B1.9) (Is this a technical stage or a workflow process?)
- Record actions taken during ingest while they're happening (B3.8)
- Digital object submission interface requires from the user those pieces of metadata for which they are responsible and associates them with the AIP (B4.1)
- Acquire "Preservation Description Information (PDI) for AIP-associated Content Information." (B3.6)
- Representation Information Registries
    - Use "appropriate international Representation Information (including format) registries." (B3.3)
    - Record/register Representation Information ingested (B3.4)

Dissemination (**Ryan and Muzzo**)

- Ensure that "the process that generates the DIP is correct in relation to the request." (If they get the right image, but in the wrong format, that's a violation of this.) (B5.4)
- Ensure that the process that generates the DIP does not cause the loss or corruption of any content information (B5.6)

Data Integrity (**Ryan**)

- ✅ Obtain "sufficient physical control over the digital objects to preserve them." (B1.5)
- Collect and report information integrity measurements (chain of custody, checksums, etc.) (A3.6)
    - Analysis of digital content
    - Verification, analysis, and creation of metadata
    - Authentication and integrity checking
    - Creation of the AIP
- SIP tracking: "The accessioning procedures and the internal processing audit logs should maintain records of all internal transformations of SIPs, and thus demonstrate that they either become AIPs (or part of AIPs) or are disposed of" (B1.8)
- Actively monitor AIP integrity with Fixity Information (such as checksums) (B3.7)
- Create and ensure referential integrity between AIPs and descriptive information (no AIPs without information, no information without associated AIP(s)) (B4.2, B4.3)
- Track number and location (without ambiguity) of all copies of stored digital objects (D1.3)
    - Ability to synchronize multiple copies (D1.4)
- Mechanisms to detect and report data corruption or loss (D1.5)
- Appropriate hardware and software and procedure for monitoring for necessary changes or migrations (D2.1, D2.2)

Look for standards or reccommendations on integrity checking - frequency. (Checksums)
How much integrity checking we will do will determine what kind of performance we need out of HPSS.
Can we get disaster recovery information out of HPSS?

Preservation/ Security

- Preserve the content information of AIPs (B3.5)
- Backup functions appropriate to content and services (D1.2) **HPSS**
- "Repository has defined processes for storage media migration" (D1.7) **HPSS**
    - Appropriate hardware and software and procedure for monitoring for necessary changes or migrations (D2.1, D2.2)
    - Implement strategies for AIP storage and migration (B3.2)
- Security **Ryan**
    - Implement "mechanisms (processes) to adequately address each of the defined security needs." (D3.2)
    - Process for staying current with security fixes (D1.10)
- Representation information **Ryan**
    - Monitor Representation Information (including formats) and notify staff when it approaches obsolescence (B3.9)
    - Use "appropriate international Representation Information (including format) registries." (B3.3)
    - Record/register Representation Information ingested (B3.4)
- Acquire "Preservation Description Information (PDI) for AIP-associated Content Information." (B3.6) **Ryan**

Hardware and Software

- Well-supported operating systems and other core infrastructural software (D1.1)
    - Ours is, HPSS's isn't
- Appropriate hardware and software and procedure for monitoring for necessary changes or migrations (D2.1, D2.2) **Ryan**

Ongoing Analysis/Monitoring/Testing (**Ryan**)

- Maintain "a systematic analysis of the environment: data, systems, personnel, physical plant, security needs, etc." (D3.1)
- Appropriate hardware and software and procedure for monitoring for necessary changes or migrations (D2.1, D2.2)
    - Monitor Representation Information (including formats) and notify staff when it approaches obsolescence (B3.9)
    - Ability to test understandability (C4.1, C4.2)
- Mechanisms to detect and report data corruption or loss (D1.5)
    - Actively monitor AIP integrity with Fixity Information (such as checksums) (B3.7)
- Monitoring and feedback mechanisms to ensure operation and resolve problems (A3.2)
- Process for staying current with security fixes (D1.10)
- Process for testing the effect of critical changes to the system (D1.9)

Workflow

Ongoing Planning/ Reporting (**Jon and Stacy**)

- Mechanism for reviewing and updating policies (A3.1)
- Regular short- and long-term business planning to support sustainability (A4.1)
- Regular "risk, benefit, investment, and expenditure" analysis and reporting (A4.4)

Designated Community (**Stacy**)

- Mechanism for eliciting new requirements from depositors (A3.2)
- Procedure for monitoring or receiving notifications about changes in the needs of the designated community/ies (D2.3)

Ongoing Testing/ Review/ Certification

- Formal, periodic reviews (A3.3) (**Jon and Stacy**)
- Regular certification; between certifications notify certifying bodies "of operational changes that will change or nullify" certification status (A3.7) **Jon and Stacy**
- Regular testing of disaster plans (D3.5) **Ryan**
- Process for testing the effect of critical changes to the system (D1.9) (Change management process (D1.9)) **Ryan**

Preservation/ Migration/ Security

- ✅ Process for staying current with security fixes (D1.10)
    - Brian takes care of this

- Procedure for testing understandability (C4.1, C4.2)
- "Repository has defined processes for storage media migration" (D1.7) **HPSS**
- Take action when notified that Representation Information (including formats) approaches obsolescence (B3.10) **?**
- Take action when data is corrupted or lost (D1.6)

User Interaction

- "Repository provides Producer/depositor with appropriate responses at predefined points during the ingest process." (B1.7) **Muzzo**
- Review inappropriate access denial incidents (B5.2) **Ryan and Brian**

Staffing

Need to document the core competencies we need and then be able to show that we have appropriate staff (**Jon and Stacy**)

- Staff have skills and expertise appropriate to their duties (A2.1)
- Appropriate number of staff (A2.2)
- Staff have delineated roles, responsibilities, and authorizations (D3.3)