LIBRARY OF CONGRESS

ICPN

Storing and Protecting Digital Content Over Time

Digital Preservation, Outreach and Education, Parts 3 & 4

Kris Stenson
Electronic Records Archivist
Illinois State Archives

Today's Agenda

- Introduction to the program
- How to best store your digital content
- Break
- How to protect your content from disaster

LIBRARY OF CONGRESS

ICPN

Topics Covered

- How to make crucial storage decisions
- The importance of metadata
- Options for storage media
- How others can help you achieve storage goals
- Identifying threats to your content
- Taking proactive steps to protect content
- Dealing with a disaster
- Disaster recovery

LIBRARY OF CONGRESS

ICPN

DPOE Modules

Identify – what digital content do you have?

Select – what portion of that content will be preserved?

Store – what issues are there for long term storage?

Protect – what steps are needed to protect your digital content?

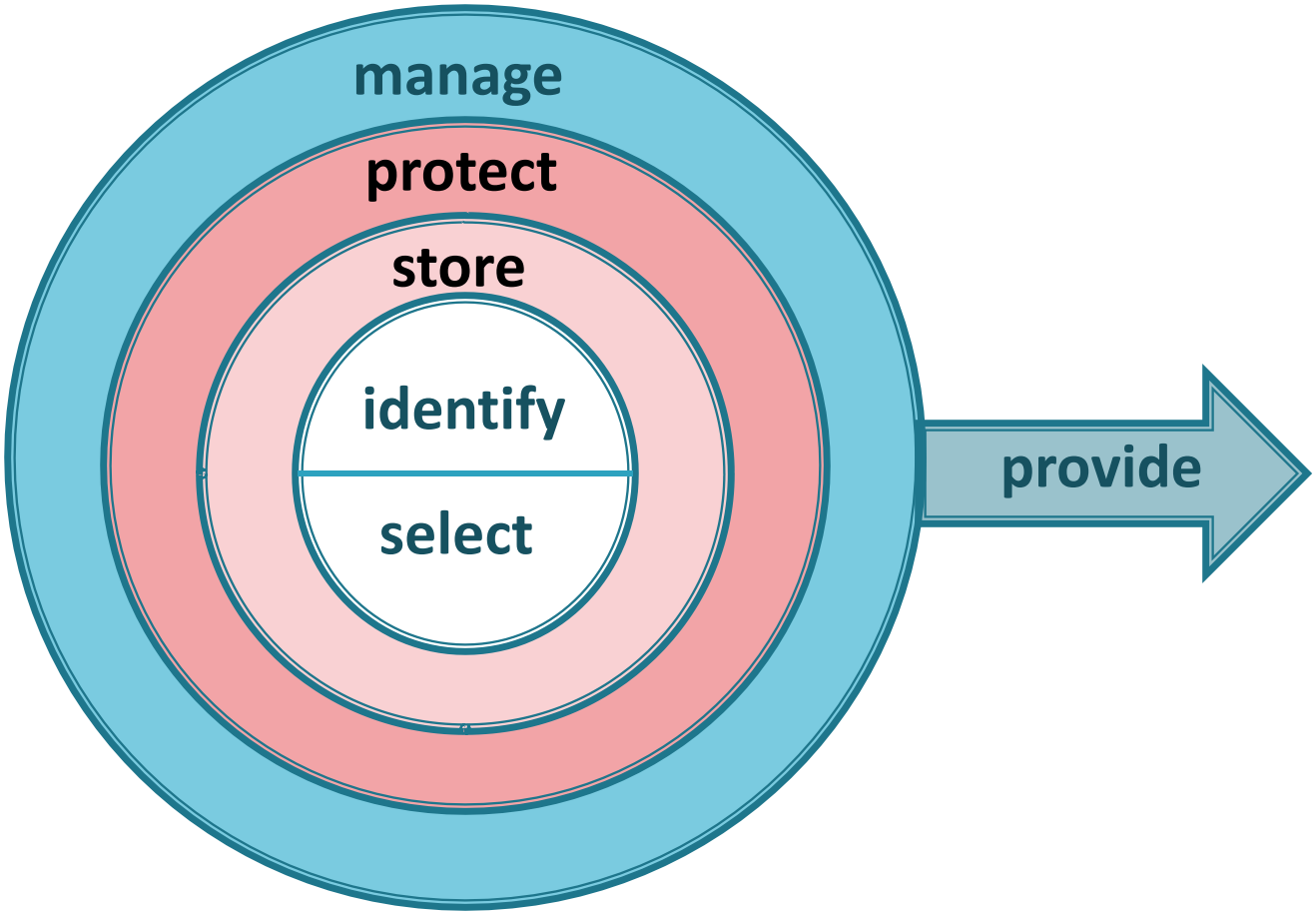
Manage – what provisions are needed for long-term management?

Provide – what considerations are there for long-term access?

LIBRARY OF CONGRESS

ICPN

The DPOE Model



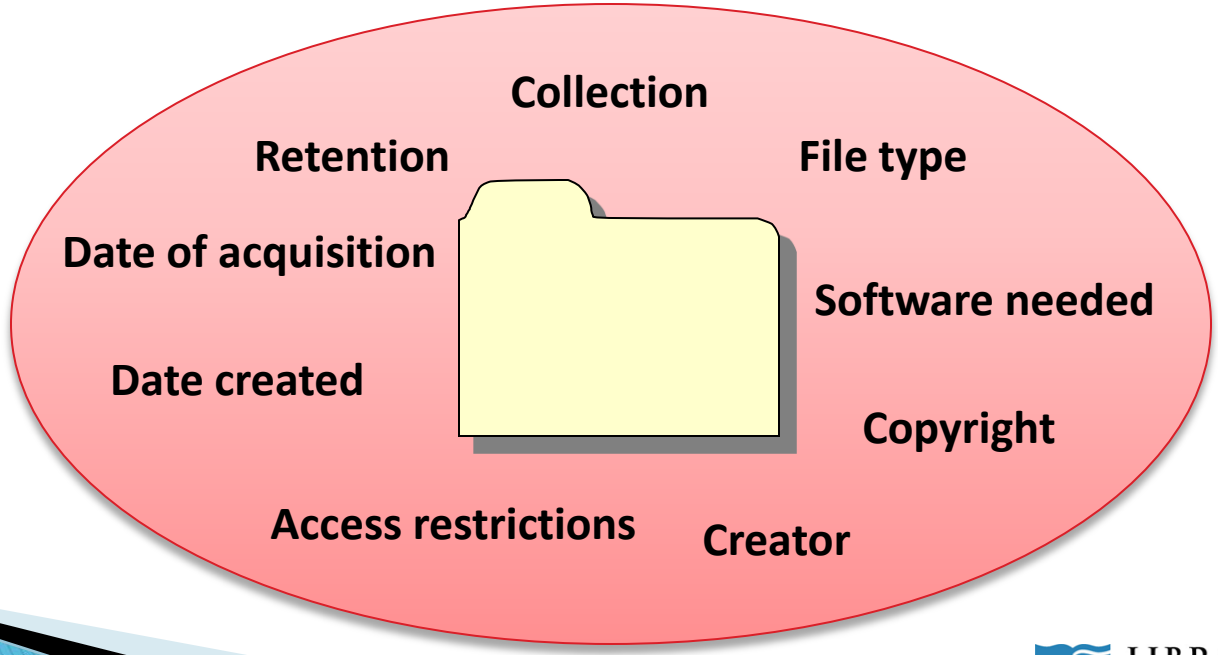
LIBRARY OF CONGRESS

ICPN

Overview of Storage Needs

Archival Storage manages digital content as objects, not just data

File + Metadata = Object

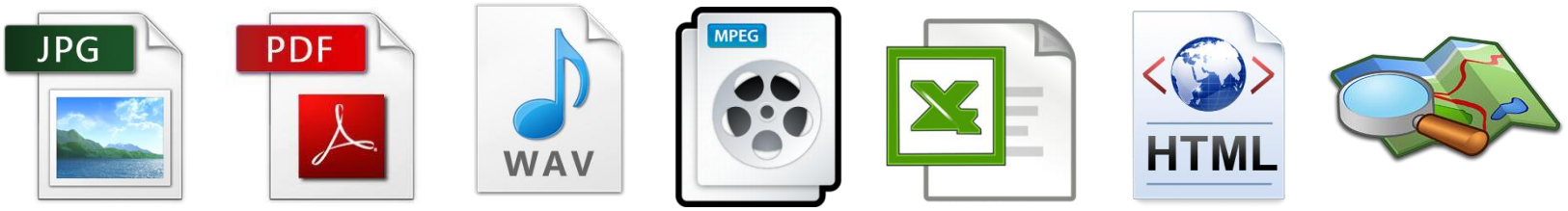


LIBRARY OF CONGRESS

ICPN

Overview, cont.

- May include any type of digital content
  - Images, text, sound, video, maps, etc.



- Requires some identification and description
  - Captured as metadata

- Needs at least **two** copies at least **two** places

LIBRARY OF CONGRESS

ICPN

Well-managed Collections

- Well-managed status makes preservation easier
- Sample characteristics:
  - Basic information about each deposit
  - Minimum metadata for objects (you define)
  - Common (or *normalized*) file formats
  - Controlled and known storage of content
  - Multiple copies in at least 2 locations

LIBRARY OF CONGRESS

ICPN

Let's Talk Metadata!

- How do you know what an object is?
  - Metadata uniquely identifies digital objects
- How do you know an object is authentic?
  - Metadata allows objects to be traced over time
- How do you use the object in the future?
  - Metadata makes digital objects understandable

Metadata enables long-term preservation

LIBRARY OF CONGRESS

ICPN

Object-level Metadata

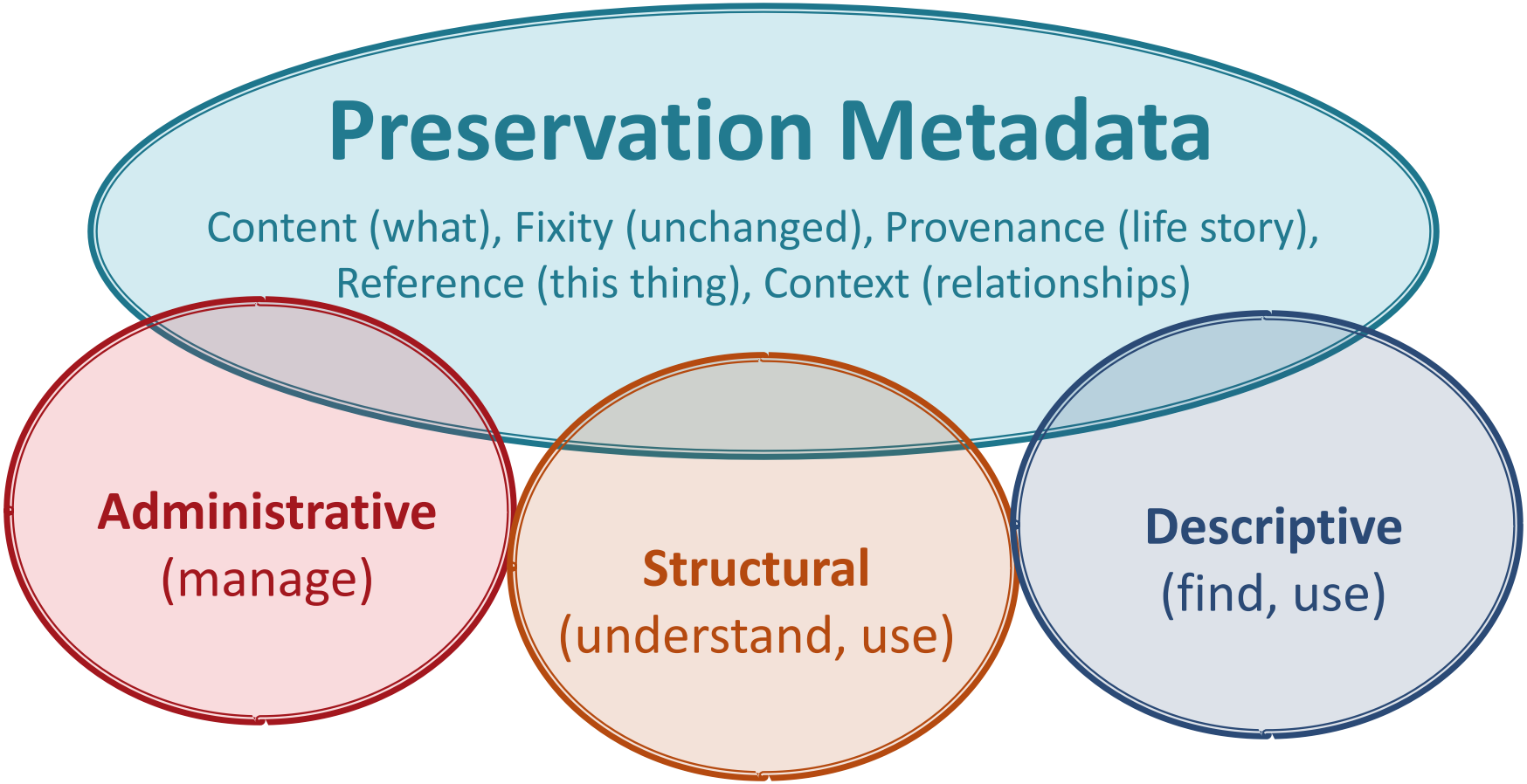


Diagram courtesy DPM Workshops

LIBRARY OF CONGRESS

ICPN

Object Metadata Properties

- **Content**: preserve the substance
- **Fixity**: demonstrate content is unchanged
- **Reference**: identify as this content and no other
- **Provenance**: trace to its origin (or to deposit)
- **Context**: preserve linkages with other objects

Original source: Preserving Digital Information Report, 1996

LIBRARY OF CONGRESS

ICPN

File Formats

- Characteristics of good preservation formats:
  - Public documentation and open disclosure
    - Ensure future access
  - Widespread adoption and use
    - Lessens risk of sudden obsolescence
  - Self-describing
    - Limits software needed to interpret
  - Unencrypted, Uncompressed*
    - Encryption and compression jeopardize access

Storage Media Options

- Storage Media can include:
  - Physical, removable storage (CD, DVD, Tape, Flash)
  - Hard drives (HDD or SSD) in desktops or servers
  - RAID setups
  - Cloud services

- Types of storage:
  - Online: highest cost, fastest access
  - Near-line: medium cost, medium access speed
  - Offline: lowest cost, slowest access

Storage Media Considerations

- Cost (available resources for preservation)
  - Up front, future maintenance
- Space requirements (number and size of files)
- Expertise (skills required to manage)
  - IT, staff training
- Partners
  - Share repositories, geographic distribution
- External services (outsourcing)
  - Cost vs in-house
  - Who controls?

LIBRARY OF CONGRESS

ICPN

What is "The Cloud" exactly?

- Hosted services, storage partners
- Multiple, geographically distributed copies

Talk Amongst Yourselves

"Cloud Storage" is neither a cloud nor storage

©NBC

DISCUSS

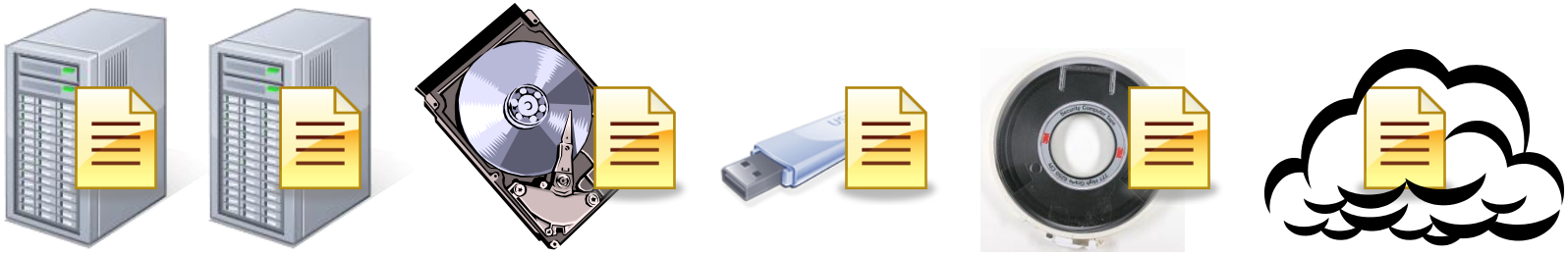
LIBRARY OF CONGRESS

ICPN

How Many Copies are Enough?

Minimum: Two (2) copies in two locations

Optimum: Six (6) copies in multiple locations



Hannes Grobe

LIBRARY OF CONGRESS

ICPN

Copy Considerations

- Size or number of files
  - Video files usually too large to have lots of copies
- Possible legal restrictions
  - Where permitted to store?
- Security concerns
  - Sensitive information: more copies = more risk
- Types of media used for storing the content
  - What is available to you?

Beyond Simple Storage

- Long-term storage requires a **Repository**
- Repositories:
  - Provide integrated storage and management
    - All in one place
    - Easier to provide access to users
  - Provide a centralized interface to work with
  - Allow for active maintenance of objects
    - Error checks
    - Migration/disposal scheduling

LIBRARY OF CONGRESS

ICPN

Choosing a Repository

- Build, join or buy?
- Range of types to consider:
  - General (any content) to special (format-specific)
  - Open source to proprietary
  - Easy to advanced installation and management
- Each option has pros and cons
- No system is fully compliant to standards
  - OAIS is the goal, but hard to achieve
- Select best option for your content – for now
  - Don't wait for the perfect solution

LIBRARY OF CONGRESS

ICPN

Planning for the Future

- Develop a storage management policy
  - Number of copies, locations, fixity means, metadata collected, file formats, etc.
- Specify storage service or partner agreements
  - In-house? Network of peers? Vendor?
- Monitor copies of content for errors/change
  - Checksums, periodic sampling
- Plan for media replacement
  - Budget for media and manpower
  - Make part of regular workflow, not project

LIBRARY OF CONGRESS

ICPN

LIBRARY OF CONGRESS

ICPN

BREAK

LIBRARY OF CONGRESS

ICPN

Protecting Your Digital Content

Protecting Content From…

Change and loss- accidental or intentional

LIBRARY OF CONGRESS

ICPN

Protecting Content From…

Obsolescence due to rapid technology change

LIBRARY OF CONGRESS

ICPN

Protecting Content From…

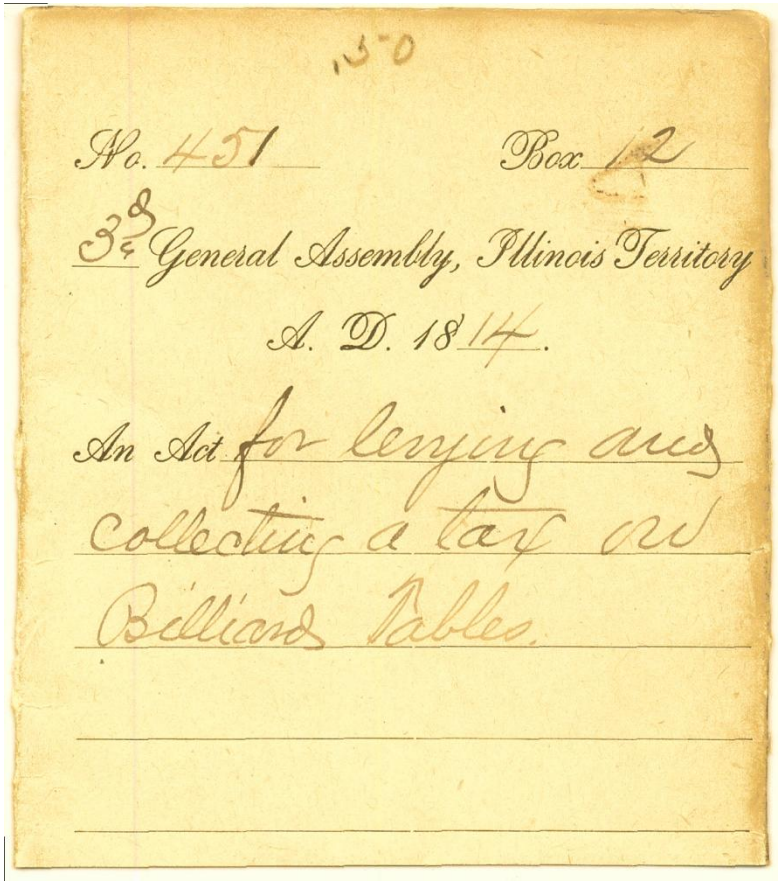
- Inappropriate access to data

LIBRARY OF CONGRESS

ICPN

Protecting Content From…

Non-compliance with standards or requirements



No. 451 Box 12

3d General Assembly, Illinois Territory

A. D. 1814.

An Act for levying and collecting a tax on Billiard Tables.

LIBRARY OF CONGRESS

ICPN

Protecting Content From…

Disasters (natural and man-made)

NIH Record
http://nihrecord.od.nih.gov/newsletters/2007/07_27_2007/story1.htm

LIBRARY OF CONGRESS

ICPN

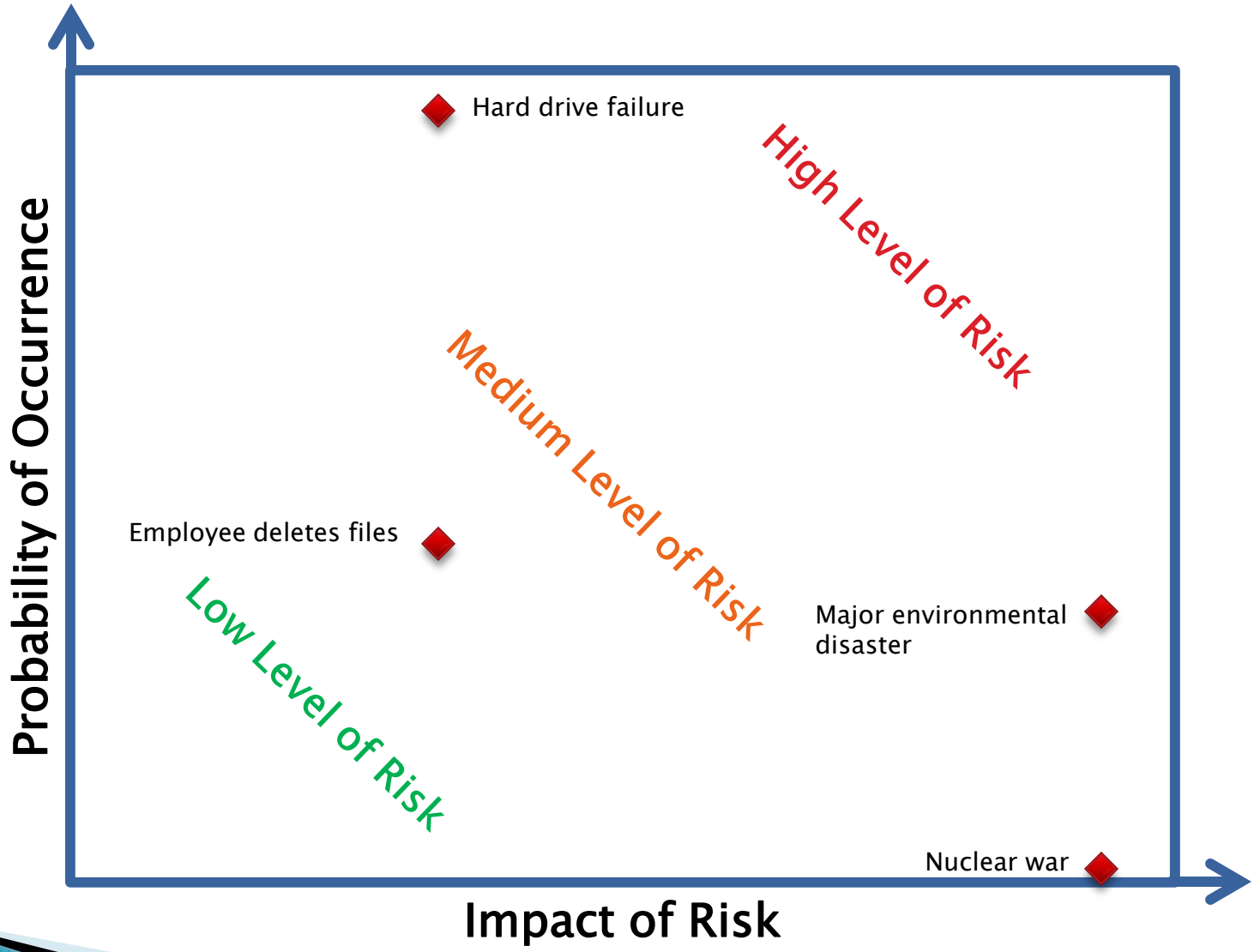
Prevention & Treatment

- Proper planning should allow you to:
  - Predict – most likely risks and threats
  - Prevent – undesirable outcomes
  - Detect – errors, problems, damage
  - Respond – with appropriate measures
  - Repair – damage or possible loss

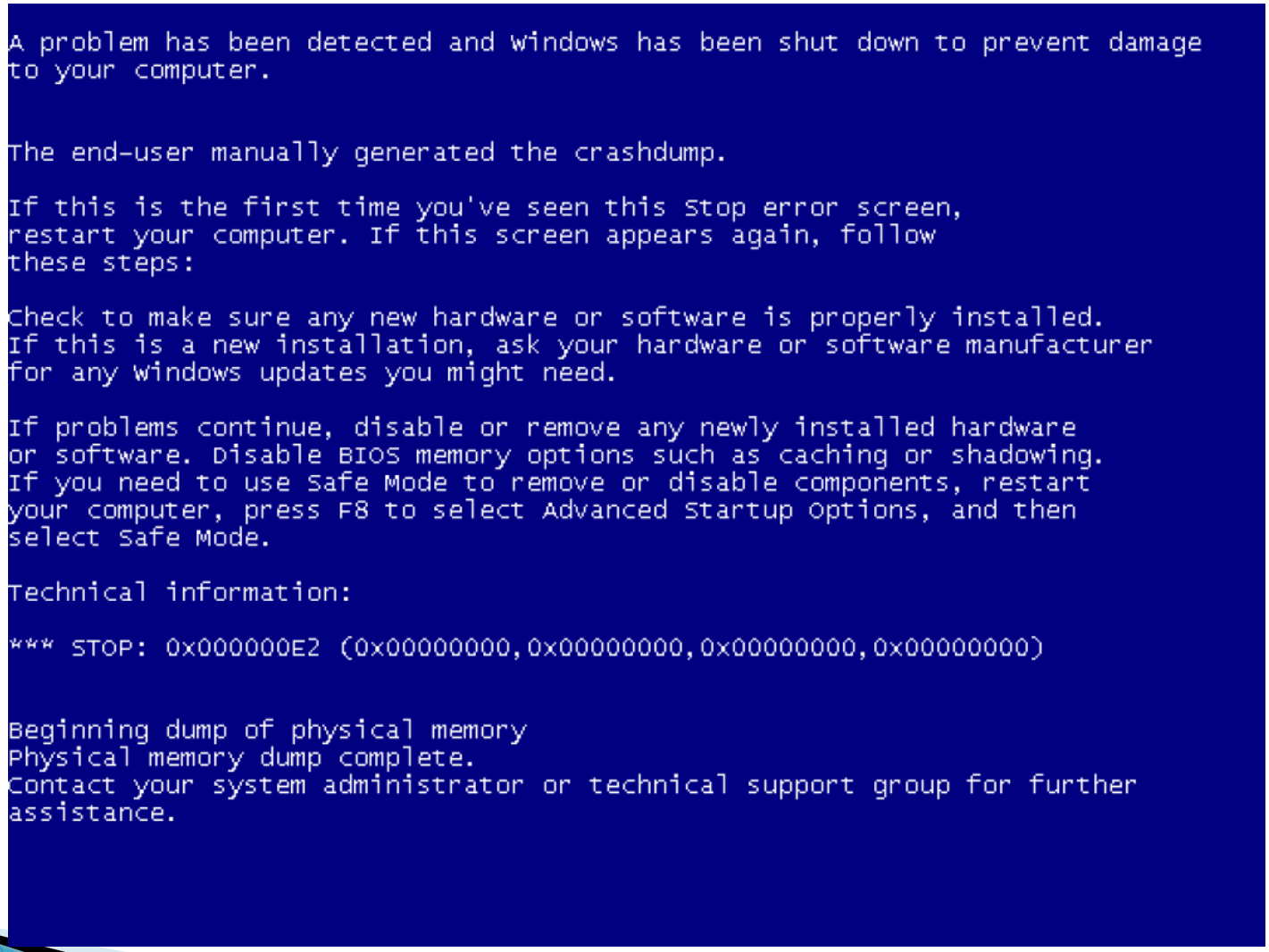
Risk Management

1. Identify possible risks
2. Define those risks (nature and scope)
3. Assess potential impact (possible damage)
4. Develop appropriate, feasible responses (plans)
5. Respond to risks, threats (implement plans)

Risk Impact/Probability Chart



Exercise



LIBRARY OF CONGRESS

ICPN

Control Your Content

- Revisit your inventory
  - Where is content? On- or offline?
- Know who has access to content
  - Administrators, IT staff, others?
- Manage authentication information
  - Control access for staff, depositors, users
- Track and review usage then adjust practices
  - Safety vs. utility
  - Web use, internal use and activities, maintenance

LIBRARY OF CONGRESS ICPN

Reduce Your Risk

- Choose stable media and formats
  - Refresh/migrate on schedule
- Have appropriate backups
  - Practice geographic redundancy
- Keep informed
  - New tech developments
  - Know your area's disaster risks
  - Know the rules/laws that apply to your collections

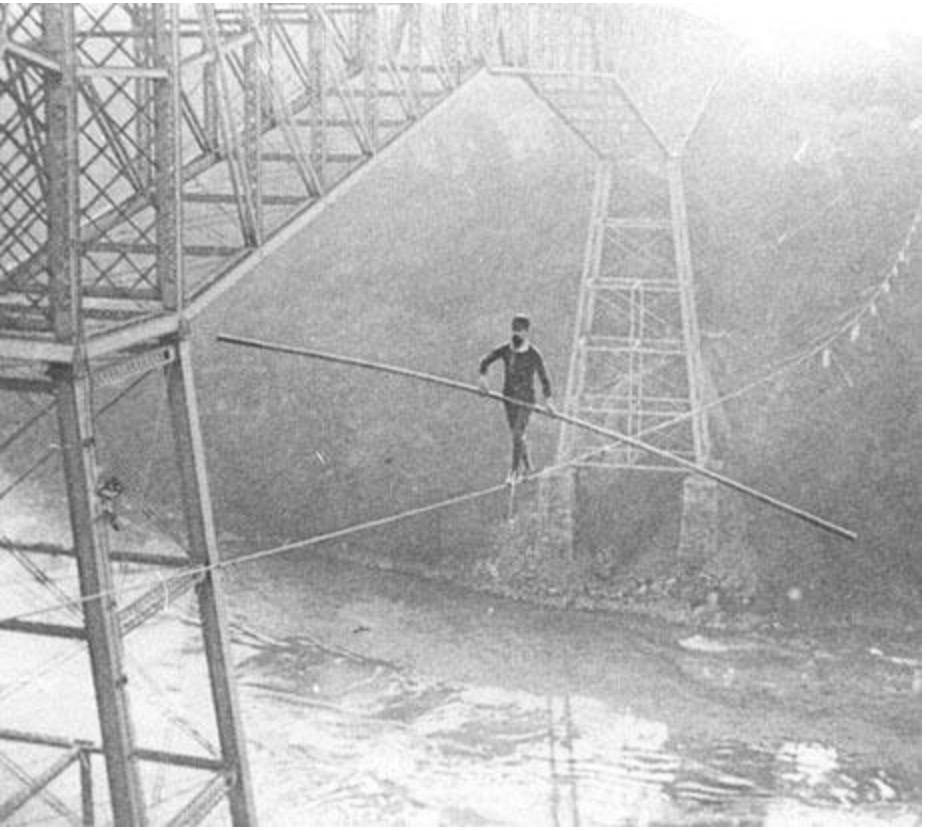
LIBRARY OF CONGRESS ICPN

Detect Problems Early

- Physical monitoring
  - Moisture sensors
  - Smoke alarms
  - Inspection schedule

- Digital monitoring
  - Checksums
  - Regular system fixity checks
  - System diagnostic tools
  - Regular format & media review

Planning for the Worst



- Engage in ongoing disaster planning
- Establish committee and share information
- Develop and maintain (and read!) documents
- Identify possible outcomes and prepare
  - server goes down, media is damaged

LIBRARY OF CONGRESS

ICPN

Planning Components

"Ultimately, an organization would use a **suite of plans** to properly prepare response, recovery, and continuity activities for disruptions affecting the organization's **IT systems, business processes, and the facility**."

NIST Contingency Planning Guide for Information Technology Systems, pg. 7.

LIBRARY OF CONGRESS

ICPN

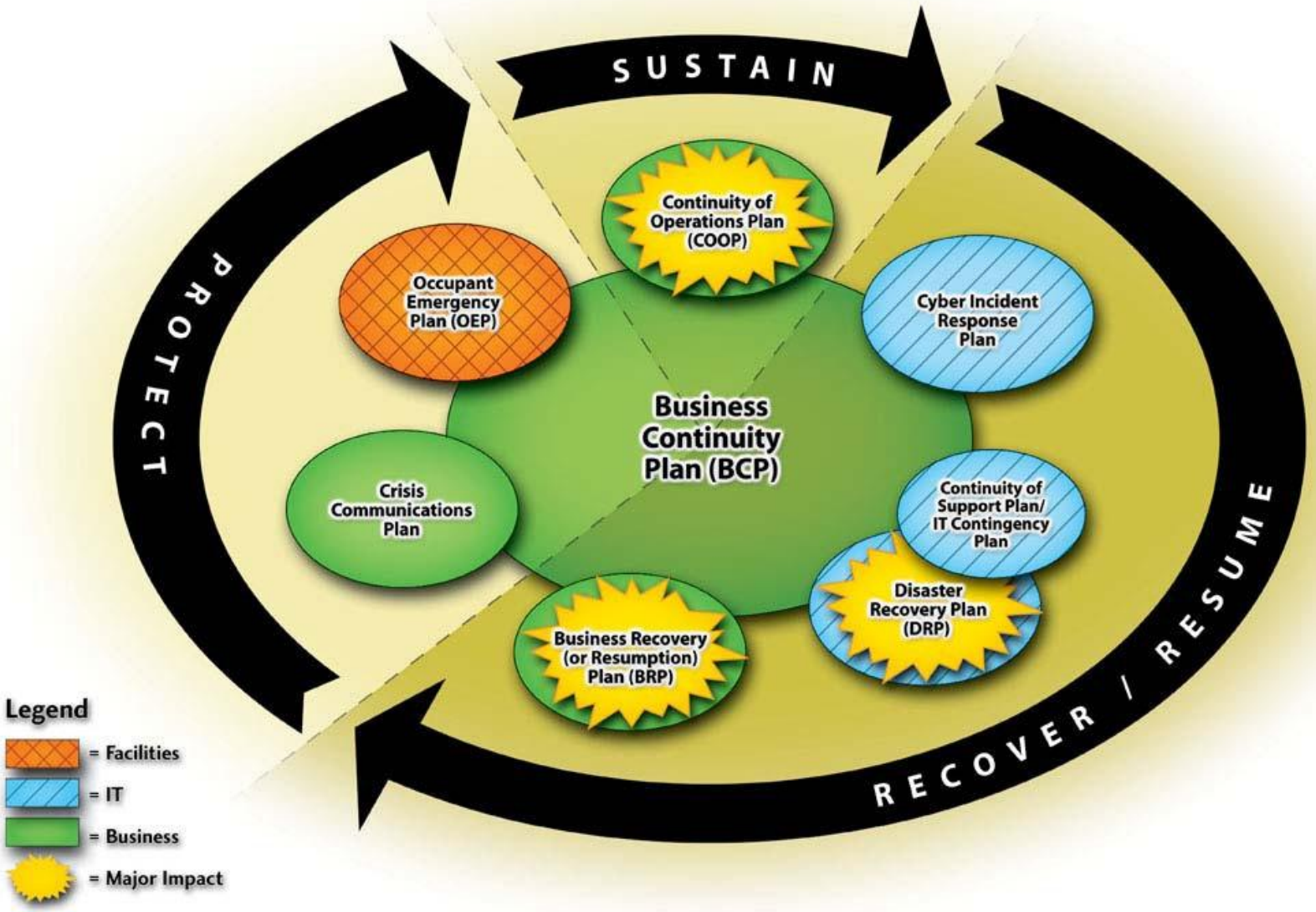
NIST Planning Components

- Business Continuity Plan (BCP)
- Business Recovery Plan (BRP)
- Continuity of Operations Plan (COOP)
- IT Contingency Plan
- Crisis Communication Plan
- Cyber Incident Response Plan
- Disaster Recovery Plan (DRP)
- Occupant Emergency Plan (OEP)

NIST Contingency Planning Guide for Information Technology Systems, pg. 11.

LIBRARY OF CONGRESS

ICPN



NIST Contingency Planning Guide for Information Technology Systems (2002), pg. 11.

LIBRARY OF CONGRESS

ICPN

Priorities in Emergencies

- Safety of employees and guests comes first
  - Can't store a backup copy of those
- What needs to be available soonest?
  - Identify core functions as part of planning
  - Determine allowable downtime for each
  - Consider steps to re-establish each function
  - Develop relevant documents
  - Make sure planning documents are accessible

Disaster Planning Resources

Ready
Prepare. Plan. Stay Informed.

FEMA

NIST
National Institute of
Standards and Technology

CoSA

EMERGENCY PREPAREDNESS INITIATIVE
Securing Our Nation's Essential Records

NORTHEAST
DOCUMENT
CONSERVATION
CENTER

The LIBRARY of CONGRESS
PRESERVATION

LIBRARY OF
CONGRESS

ICPN

Outcomes of Disaster Planning

- Practices in place to manage day-to-day protection – an implemented security plan
  - System security
  - Facilities maintenance
- Procedures in place to predict, prevent, detect, respond, repair
  - Knowing your risk environment
  - Knowing your collections
  - Diagnostic tools in place
  - Prioritizing recovery efforts

Questions?

Kris Stenson
Electronic Records Archivist
Illinois State Archives
217-557-1085
kstenson@ilsos.net

LIBRARY OF CONGRESS

ICPN