Digital information governance: selection, organization, long term preservation

- module 3 -

Pierluigi Feliciati

University of Macerata (Italy)

Summer School “Digital Tools for Humanists”

Pisa, June 11th 2019

Laboratorio di Cultura Digitale

ASSOCIAZIONE per l'INFORMATICA UMANISTICA e la CULTURA DIGITALE

Università di Pisa

Content in modules

1. Why *Information Governance*?
2. Organizational issues: analysis and actions
3. IG program + Retention/Selection/Appraisal -» Preservation

premise

- Building and maintaining a good Records and Information program is for definition a long-term practice
- So, we should take as a premise:
    - The benefits of a good Records Management System
    - Principles for good recordkeeping and maturity model (GARP)
    - The importance of Retention, Appraisal and Disposition
    - The necessity of an ongoing maintenance program: compliance, continuous assessement and improvement

Maintaining a RIG program

- The RIG program should be an everyday part of an organization’s operations and communications.
- It requires vigilant and consistent monitoring and auditing to ensure that RIG policies and processes are effective and consistently followed and enforced.
- It also requires an ongoing training and communications program to keep employees aware of approved processes and behaviors.

Maintaining a RIG program

This goal can not be just a (formal) perspective:

- it requires that someone is accountable for continual monitoring and refinement of policies and tools
- the executive sponsor for the initial project could become the chief information governance officer or RIG coordinator, driving its active improvement
- the organization also may decide to form a standing IG board, steering committee, or team with specific responsibilities
- A RIG program must be ongoing, dynamic, and aggressive in its execution in order to remain effective.

Maintaining a RIG program

- The approach to an IG program is similar to that of a vital records program (those critical business records that an organization must have to continue operations). Backups of backups must be built in.
- The redundancies of accountability must be considered, tested, and implemented. This may mean that when the formal program manager is unable to execute his or her duties, an assistant or designated backup can carry out those duties.
- It is also a good idea to cross-train employees. With this approach, the legal team, for instance, will better understand the needs and requirements of the records management function, and vice versa.

Maintaining a RIG program

- Maintaining IG program effectives requires implementing principles of continuous process improvement (CPI). CPI is a “never-ending effort to discover and eliminate the main causes of problems. It accomplishes this by using small-steps improvements, rather than implementing one huge improvement.”
- Maintaining and improving the program will require monitoring tools, periodic audits, and regular meetings for discussion and approval of changes to improve the program.
- It will require a cross section of team leaders from IT, legal, records management, compliance, internal audit, and risk management as well as functional business units participating actively.

Why continuously improve IG program?

1. Changing technology. New technology capabilities need to be monitored and considered with an eye to improving, streamlining, or reducing the cost of IG. The IG program needs to anticipate new types of threats and also evaluate adding or replacing technologies to continue to improve it.
2. Changing laws and regulations. Compliance with new or updated laws and regulations must be maintained.
3. Internal IG requirements. As an organization updates and improves its overall IG, the program elements that concern critical information assets must be kept aligned and synchronized.

Why continuously improve IG program?

4. Changing business plans. As the enterprise develops new business strategies and enters new markets, it must reconsider and update its IG program.
5. Evolving industry best practices. Best practices change, and new best practices arise with the introduction of each successive wave of technology and with changes in the business environment. The program should consider and leverage new best practices.
6. Fixing program shortcomings. Addressing flaws in the IG program that are discovered through testing, monitoring, and auditing; or addressing an actual breach of confidential information; or a legal sanction imposed due to noncompliance are all reasons why a program must be revisited periodically and kept updated.

Long-term preservation and RIG

- Long-term continuity of digital information does not happen by accident— it takes information governance, planning, sustainable resources, and a keen awareness of the information technology (IT) and file formats in use by the organization, as well as evolving standards and computing trends.
- Digital preservation is the long-term, error-free storage of digital information, for access, retrieval and interpretation, for the entire time span the information was selected to be retained by a designed community (OAIS).
- Digital preservation applies to digital-born content as well as content converted to digital form.

Long-term preservation and IG

- Some digital information assets must be preserved permanently as part of an organization’s documentary heritage, documenting its history and reputation.
- Dedicated repositories for historical and cultural memory, such as libraries, archives, and museums, act as trustworthy digital repositories that can match the security controls, and wealth of management and descriptive metadata similar to those that have been created for analog assets (such as books and paper records).

Long-term preservation and RIG

- The definition of "long term" comes from the International Organization for Standardization (ISO) standard 14721 (OAIS), which defines long-term as "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely."
- There is the requirement to retain the metadata associated with records even longer than the records themselves.
- A record may have been destroyed according to its scheduled disposition at the end of its life cycle, but the organization still may need its metadata to identify the record, its life cycle dates, and the authority or person who authorized its destruction.

Long-term preservation and RIG

- Planning for the proper care of the electronic records selected to be preserved, protected, and monitored over long periods of time to ensure they remain authentic, complete, and unaltered and available into the future is a component of a records management program and should be integrated into the organization’s RIG policies and technology portfolio as well as its privacy and security protocols.
- The capability for ensuring proper access to authentic electronic records over time (in addition to the challenges of technological obsolescence), is a sophisticated combination of policies, strategies, processes, specialized resources, and adoption of standards.

Threats to Preserving Records

- Failure of storage media.
- Failure of computer systems.
- Systems and network communications failures.
- Component obsolescence.
- Human error.
- Natural disaster.
- Extermal or internal Attacks.
- Financial shortfall.
- Business viability.
- Organisational changes

Digital Preservation Models

There are 2 broad categories of digital preservation standards.

- The first category involves systems infrastructure capabilities and services that support a trustworthy repository.
- The second category relates to open standard technology-neutral file formats.
- The international standard ISO 14721:2003 , 2012, Space Data and Information Transfer Systems — Open Archival Information System (OAIS) is a key standard applicable to LTDP (Long Term Data Preservation)

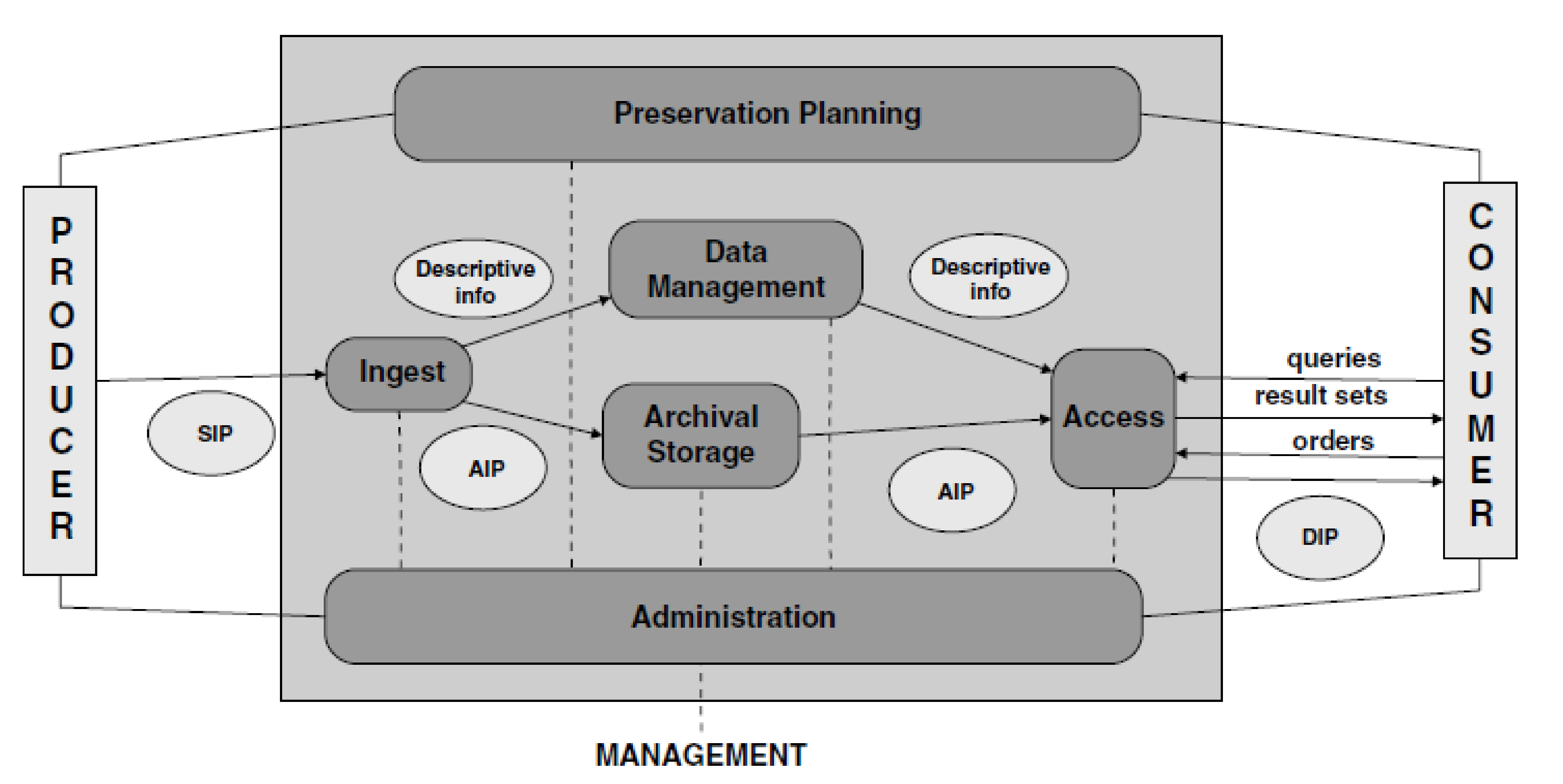
OAIS – Open Archives Information System

- The OAIS Reference Model defines an information system as an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available and understandable for a designated community, who should be able to understand the information.
- An OAIS-conforming strategy is the best way to preserve an organisation’s digital heritage.
- OAIS encapsulates digital objects into information packages, including the digital object content (a sequence of bits) and representation information that enables rendering of an object into human usable information along with preservation description information (PDI) such as provenance, context, and fixity.

OAIS

- The OAIS Information Model employs 3 types of information packages: a submission information package (SIP), an archival information package (AIP), and a dissemination information package (DIP).
- The OAIS functional model consists of six entities:
  - Ingest of SIPs into an archival system
  - Archival storage, i.e. storage of AIPs.
  - Data management, tracking the use of storage media.
  - Administration, technical and human processes like include audit, policy making, strategy, and provider and customer service
  - Preservation planning, a technology watch program for sustainable standards, file formats, and software
  - Access, i.e. creation of DIPS after the query of designated community

OAIS


Preservation Planning
PRODUCER
Descriptive info
Data Management
Descriptive info
Ingest
SIP
AIP
Archival Storage
AIP
Access
queries
result sets
orders
DIP
CONSUMER
Administration
MANAGEMENT

Other standards

- ISO TR 18492 (2005), Long-Term Preservation of Electronic Document-Based Information provides practical methodological guidance for the long-term preservation and retrieval of authentic electronic document-based information, and describes the necessary strategy to ensure the usability and trustworthiness of electronic documents, made of media renewal, software dependence, migration, open standard technology-neutral formats, authenticity protection, and security.
- ISO 16363 (2012)— Audit and Certification of Trustworthy of Digital Repositories is the certification standard with the functional specifications for records producers, records users, ingest of digital content into a trusted repository, its archival storage, and digital preserving planning and administration.

PREMIS

- OAIS specifies that preservation metadata associated with all archival storage activities should be captured and stored in PDI. This high-level guidance requirement demands greater specificity in an operational environment.
- A reference standard for preservation metadata creation and management is PREMIS, an international standard supported by the Library of Congress and the Research Library Group.
- PREMIS enables designers and managers of digital repositories to have a clear understanding of the (meta)information required to support the functions of viability, renderability, understandability, authenticity, and identity in a preservation context.

PREMIS version 3

- PREMIS is made of a Framework (=model) and a Data Dictionary, i.e. a set of semantic units. Each semantic unit is mapped to an entity that is organized within a simple data model. A semantic unit can therefore be understood as a property of an entity.
- The 4 entities to consider for digital preservation activities in PREMIS 3 are:
- Objects (+ Environments),
- Events,
- Agents
- Rights

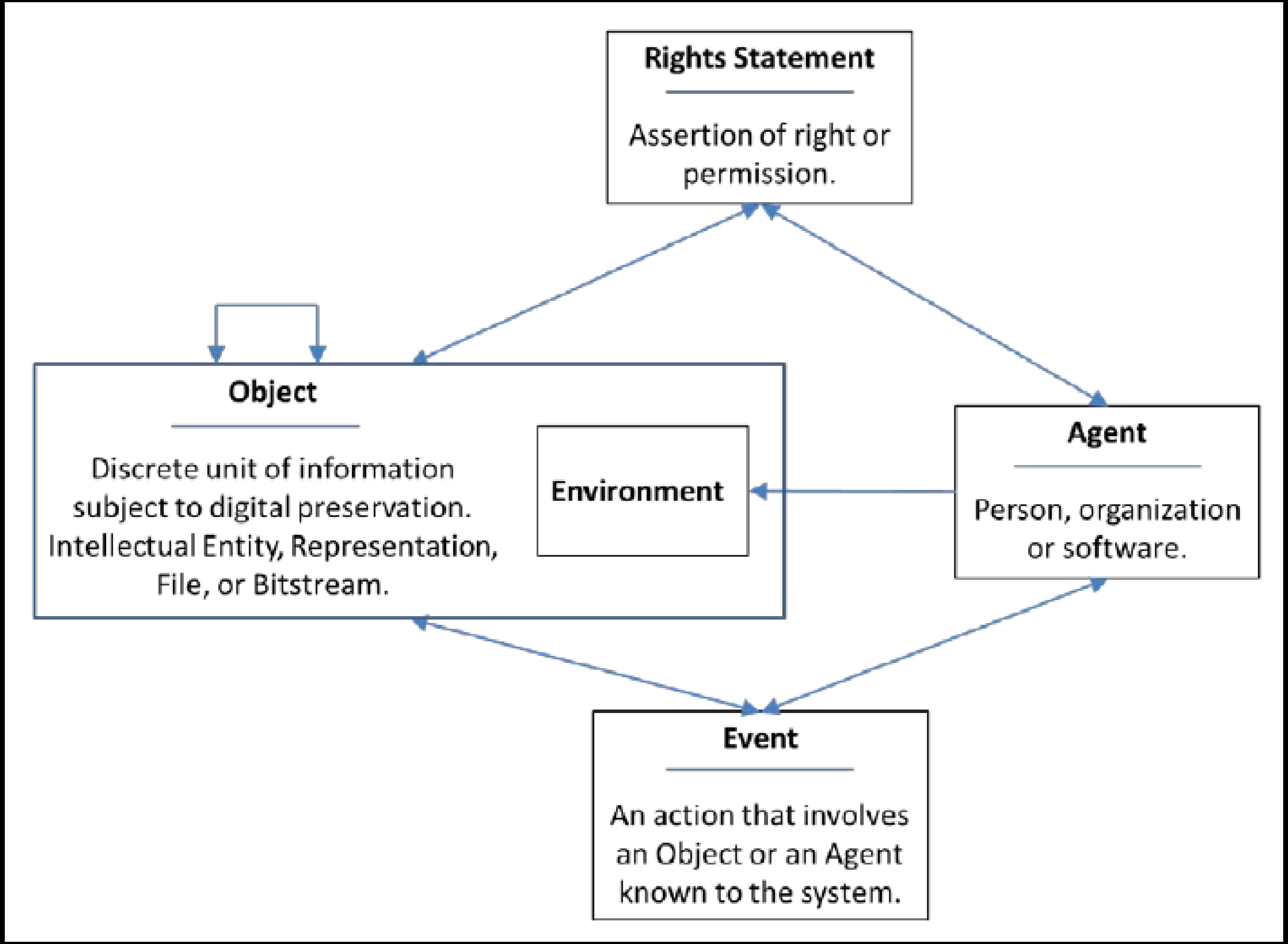
PREMIS 3 data model

Object (or Digital Object): a discrete unit of information subject to digital preservation. This can be an Environment (Technology supporting a Digital Object in some way).

Event: an action that involves or affects at least one Object or Agent associated with or known by the preservation repository.

Agent: person, organization, or software program/system associated with Events in the life of an Object, or with Rights attached to an Object.

Rights Statement: assertion of one or more Rights or permissions pertaining to an Object and/or Agent.



Rights Statement
Assertion of right or permission.
Object
Discrete unit of information subject to digital preservation. Intellectual Entity, Representation, File, or Bitstream.
Environment
Agent
Person, organization or software.
Event
An action that involves an Object or an Agent known to the system.

PREMIS version 3 – types of Objects

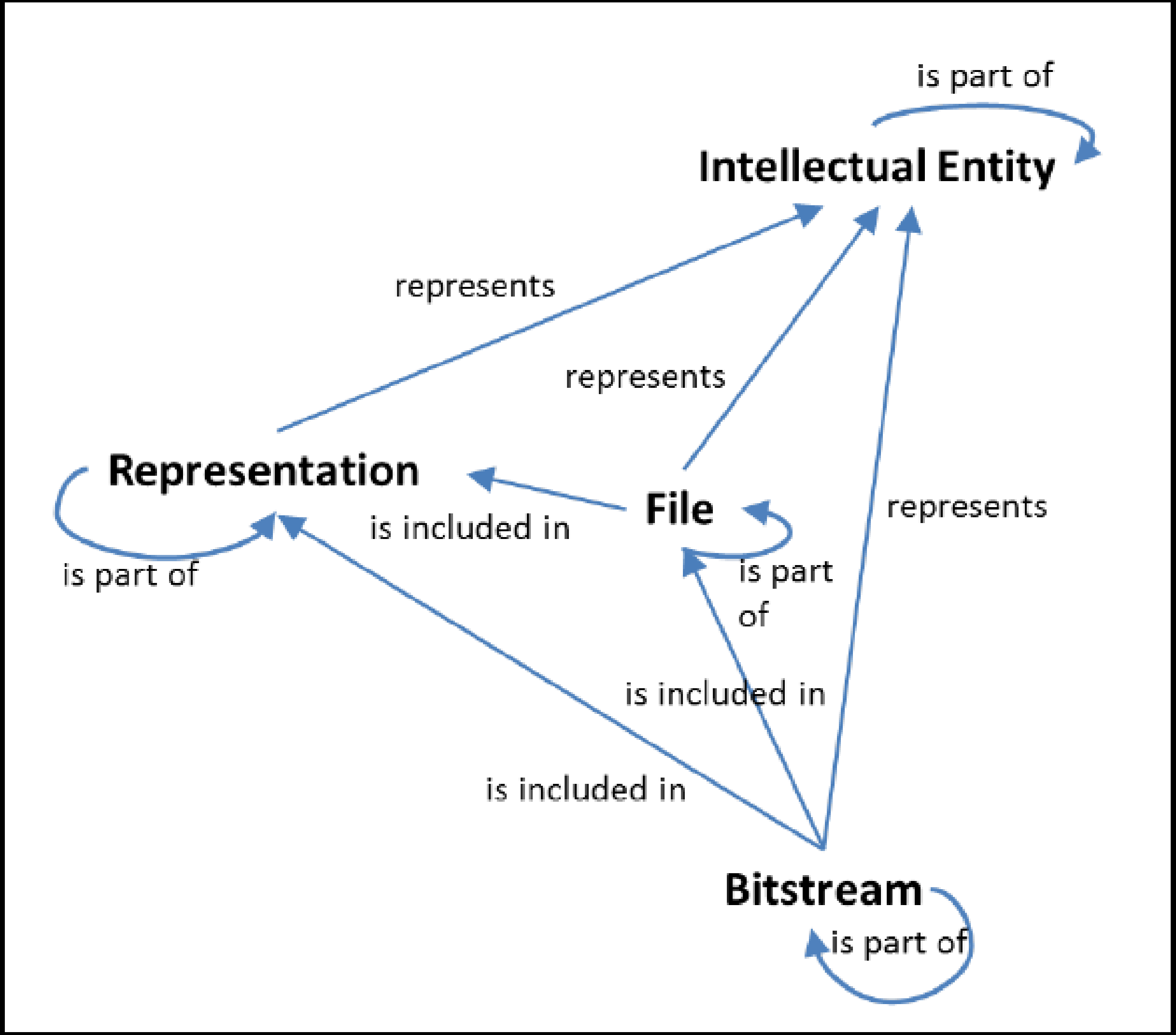
The Object entity has four subcategories: Intellectual Entity, Representation, File, and Bitstream.

An Intellectual Entity is a distinct intellectual or artistic creation that is considered relevant to a designated community in the context of digital preservation: for example, a particular book, map, photograph, database, or hardware or software.

A Representation is the set of files, including structural metadata, needed for a complete rendition of an Intellectual Entity (es. HTML page)

A File is a named and ordered sequence of bytes that is known to an operating system.

A Bitstream is contiguous or non-contiguous data within a file that has meaningful common properties for preservation purposes.



PaaST - Preservation as a Service for Trust

- Preservation as a Service for Trust (PaaST) (2018) presents functional and data requirements for digital preservation.
- The requirements reflect the findings of the first and second InterPARES collaborations that in the digital realm, preservation of information necessarily extends either to the output of an information object intended for human use in a form suitable for human consumption or, in the case of something intended to be run on a computer, to reloading it on a computational platform.
- PaaST goes over InterPARES findings: 1. its requirements are applicable to the preservation of virtually any type of digital information, not just records. 2. the PaaST requirements could support implementation and even the production of software for preservation.

PaaST - Preservation as a Service for Trust

(PaaST) defines a comprehensive set of functional and data requirements that support preservation of digital information regardless of the technologies used or who uses them.

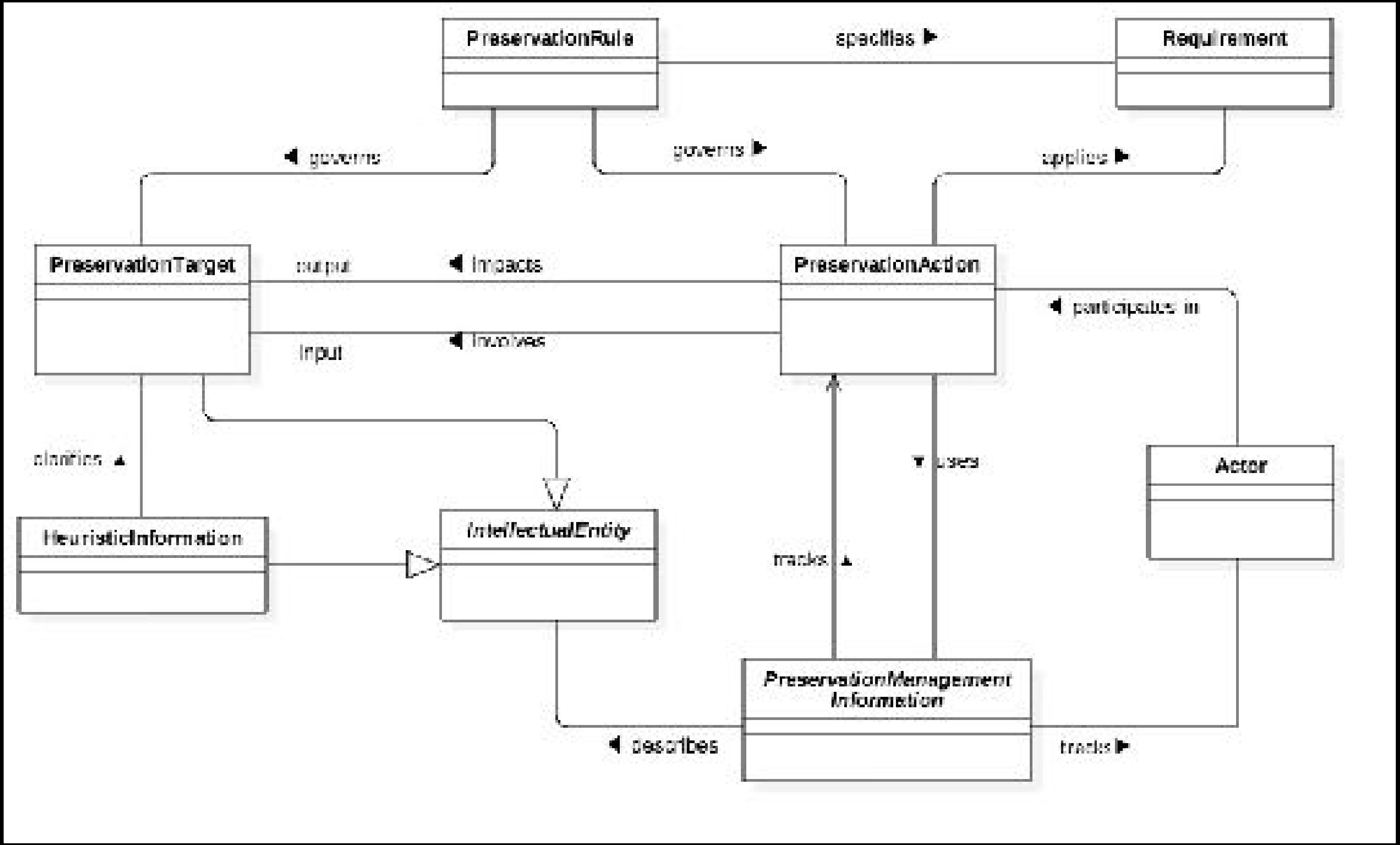
The requirements are intended to enable authentic digital preservation in the Cloud; nevertheless, the requirements are valid in other scenarios as well, including in-house preservation and situations where digital preservation includes both in-house and contracted services.

The requirements are formulated with sufficient flexibility to enable adaptation of the criteria for success to cases where information objects are not preserved as records.

PaaST - Preservation as a Service for Trust

- The PaaST requirements supplement the Open Archival Information System (OAIS) Reference Model. As a reference model, OAIS neither specifies a design or an implementation nor prescribes or even recommends any specific technology for preservation. PaaST requirements supplement OAIS in that they are intended to be directly implementable in software.
- PaaST requirements are empirically oriented, aiming to accommodate cases that are far from ideal, as well as supporting best practices.
- The PaaST requirements are neutral with respect to preservation policies and methods.

PaaST - Preservation as a Service for Trust


PreservationRule
specifies ▶
Requirement
◀ governs
governs ▶
applies ▶
PreservationTarget
output
◀ impacts
PreservationAction
◀ participates in
◀ involves
input
clarifies ▲
▼ uses
Actor
HeuristicInformation
IntellectualEntity
tracks ▲
PreservationManagement Information
◀ describes
tracks ▶