Modelling Organizational Preservation Goals to Guide Digital Preservation

Angela Dappert, Adam Farquhar

The British Library

29 September 2008

Motivation

Problem

- **Use preservation guiding documents**
- **as basis for preservation planning**

Solution

- **a data model**
- **a specific vocabulary**
  to
    - **describe**
    - **reuse**
    - **share**
  institutional preservation requirements
- **machine-interpretable models**

Methodology

Top-down approach:

- **Create a model from first principles**
- **and from literature analysis**

Bottom-up approaches:

- **Document analysis**
- **Interview decision makers**

$\Rightarrow$ **extract concepts, vocabulary**

$\Rightarrow$ **compile a requirements base**

$\Rightarrow$ **categories of requirements**

Contributions

- **Comprehensive model – everything you need to capture fits into the model.**

- **Risks and requirements are first class objects within the model.**

- **Different requirements categories play different roles in preservation planning**

- **The model lines up actions against the risks they mitigate.**

Comprehensive model – everything you need to capture fits into the model.
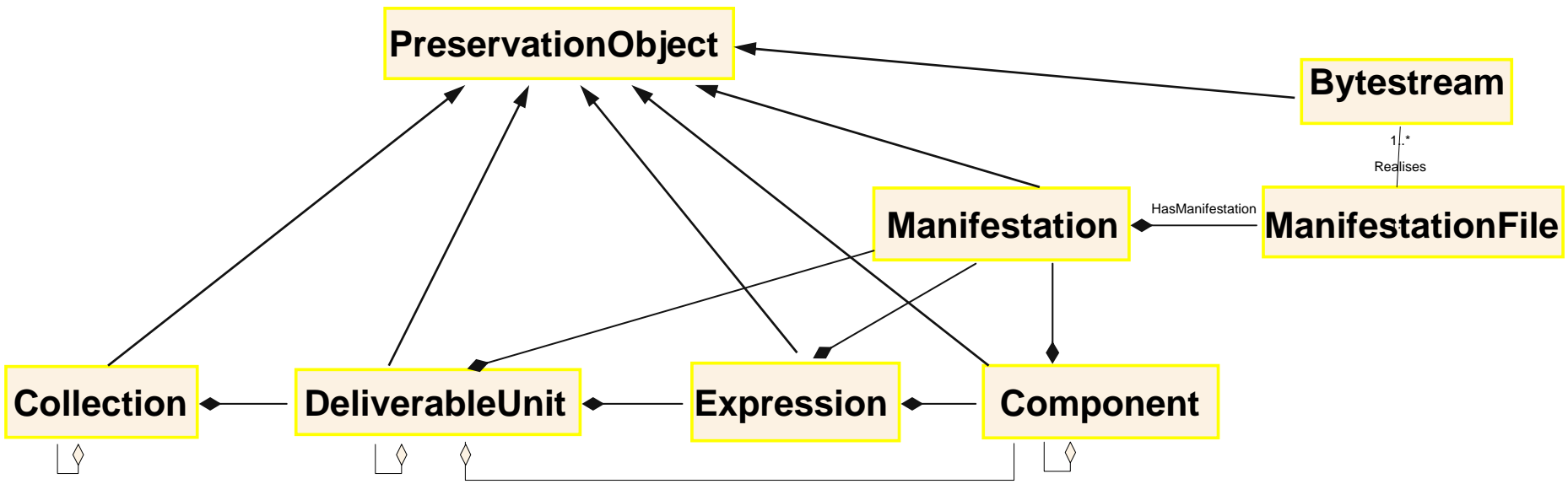
- **Comprehensive Model**
    - **full range of preservation processes**
    - **technical as well as organizational properties**
    - **full range of of preservation actions**
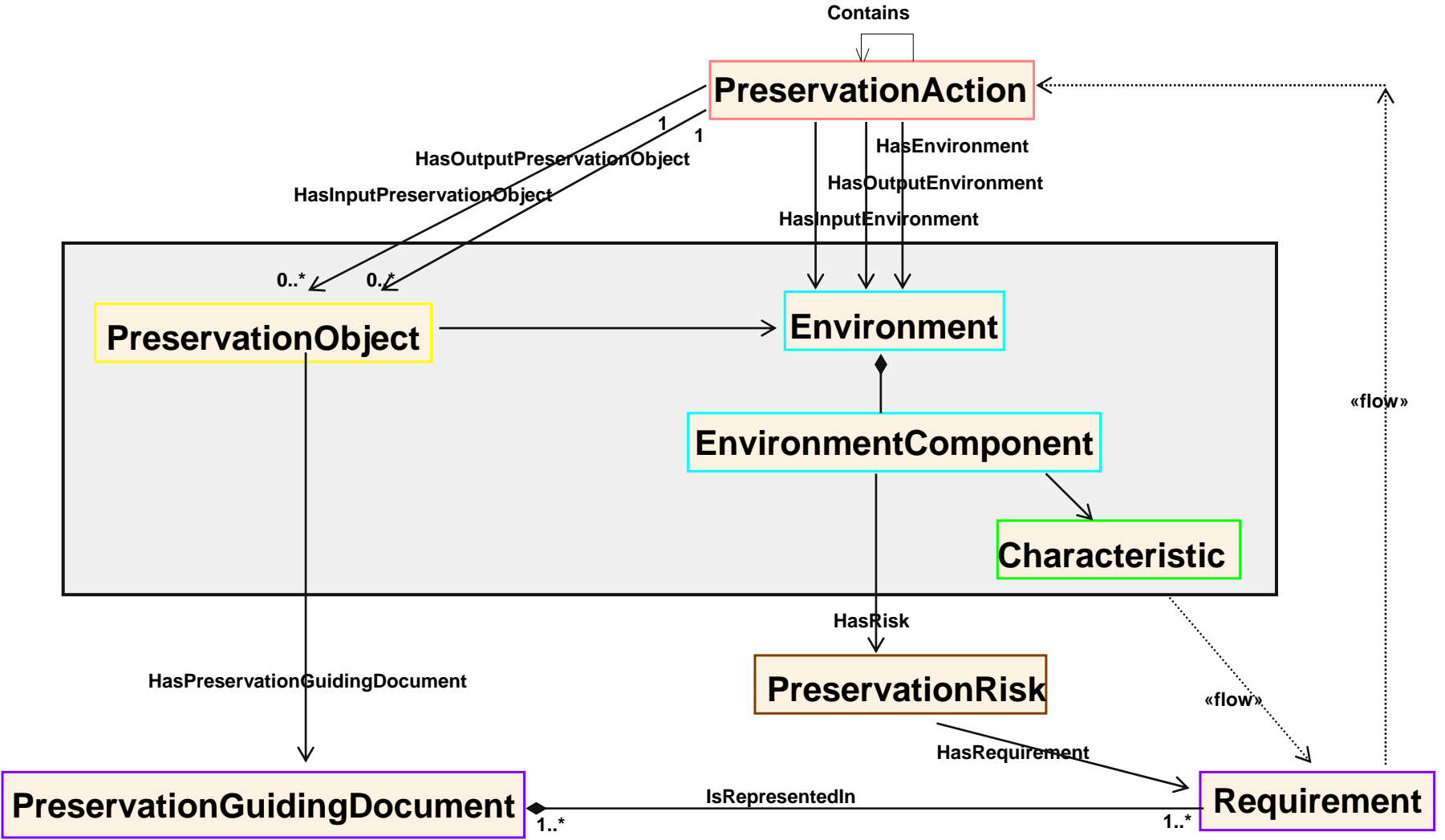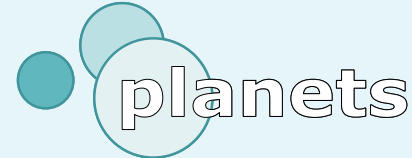   - **full range of entities**

A Preservation Object is any object that is directly or indirectly at risk and needs to be digitally preserved.

Data Model –
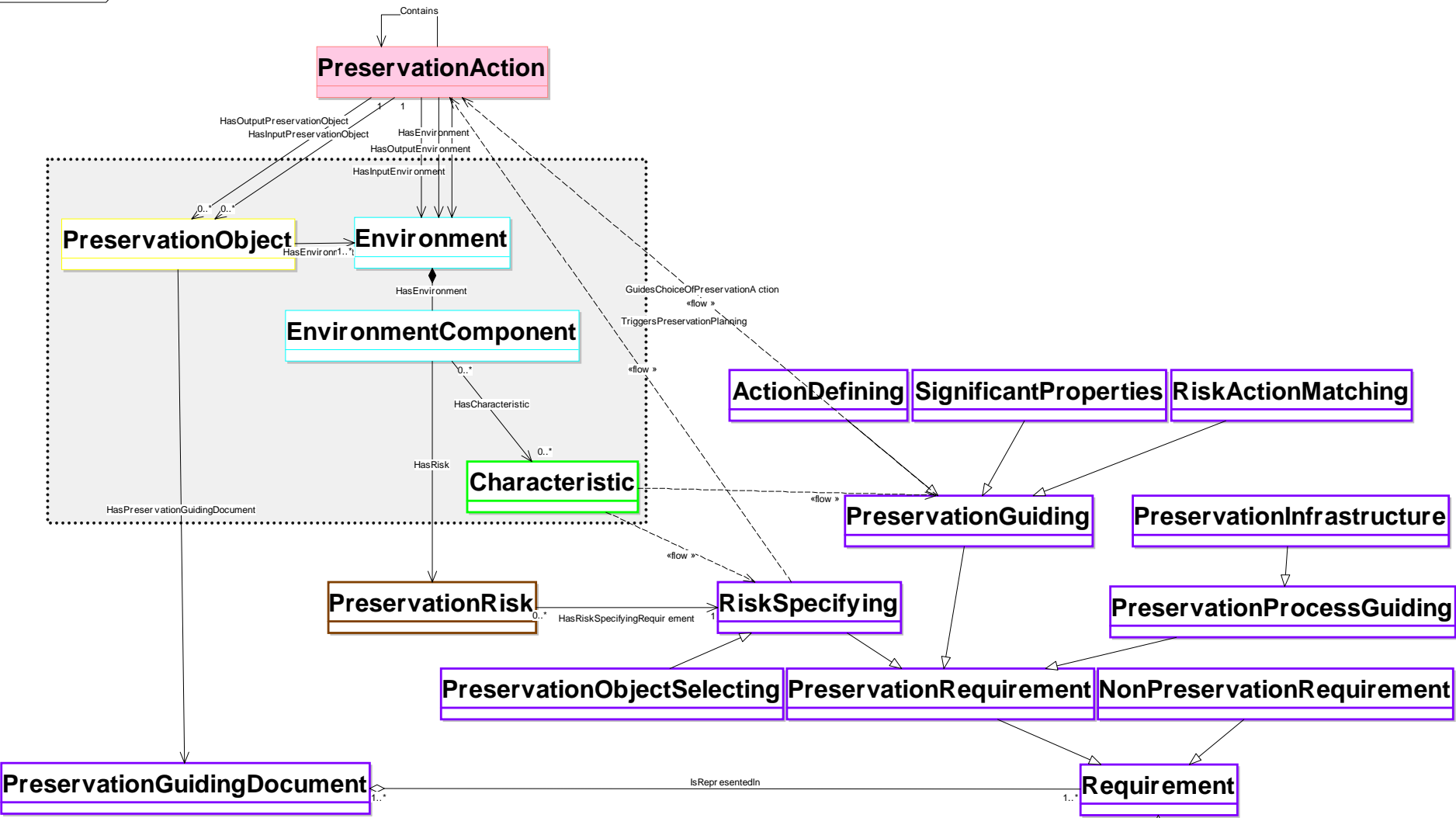## Mitigating Risks through Action

planets

Different requirements categories play different roles in preservation planning.

planets

class RequirementsTypes

Contains

PreservationAction

HasOutputPreservationObject
HasInputPreservationObject
HasEnvironment
HasOutputEnvironment
HasInputEnvironment

PreservationObject

Environment

HasEnvironn 1..*

HasEnvironment

EnvironmentComponent

HasCharacteristic

HasRisk

GuidesChoiceOfPreservationAction
«flow »
TriggersPreservationPlanning

«flow »

ActionDefining **SignificantProperties** **RiskActionMatching**

Characteristic

«flow »

PreservationGuiding **PreservationInfrastructure**

HasPreservationGuidingDocument

«flow »

PreservationRisk

HasRiskSpecifyingRequirement

RiskSpecifying

PreservationProcessGuiding

PreservationObjectSelecting **PreservationRequirement** **NonPreservationRequirement**

PreservationGuidingDocument
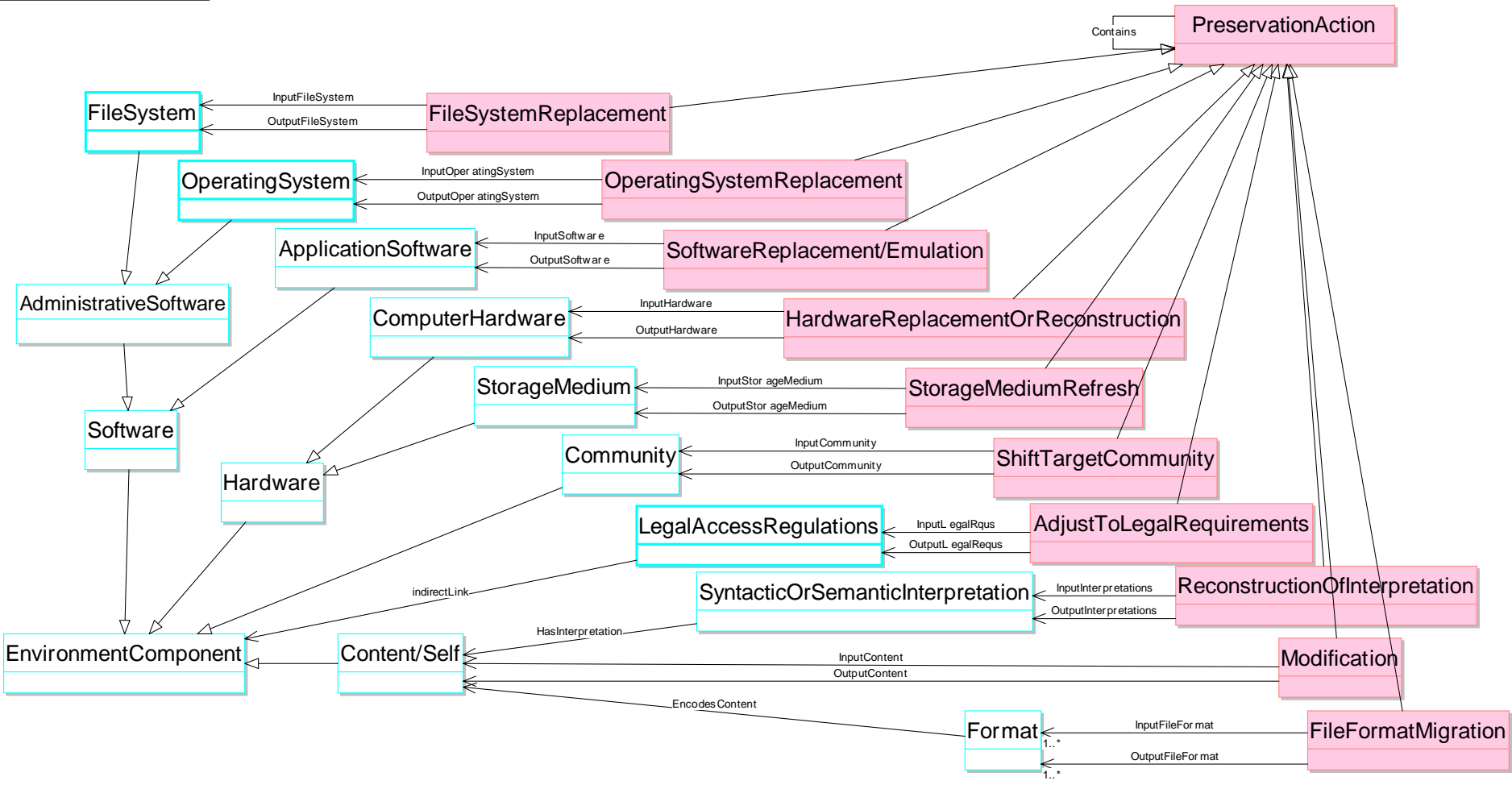
IsRepresentedIn

Requirement

8

The model lines up actions against the risks they mitigate.

planets

class 2a1 Preservation Action - examples

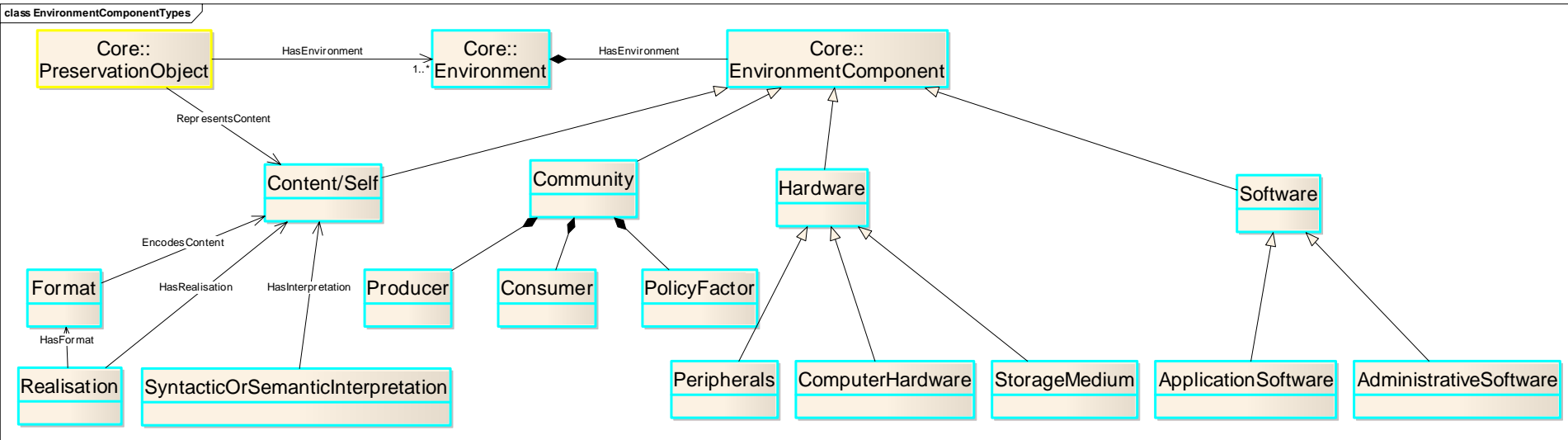# Key Findings from the Analysis of Preservation Guiding Documents

- **Data carrier refresh has become an urgent priority.**
- **Most current preservation polices specify preventive actions during ingest.**
- **Most institutions currently hold fairly homogeneous digital collections.**
- **There is a lack of consensus on digital preservation terms, goals, practice.**
- **Preservation documents do not provide specific practical guidance.**
- **Policies may not accurately reflect the institution's actual preservation goals.**
- **Some institutions mandate a specific "technical preservation strategy".**
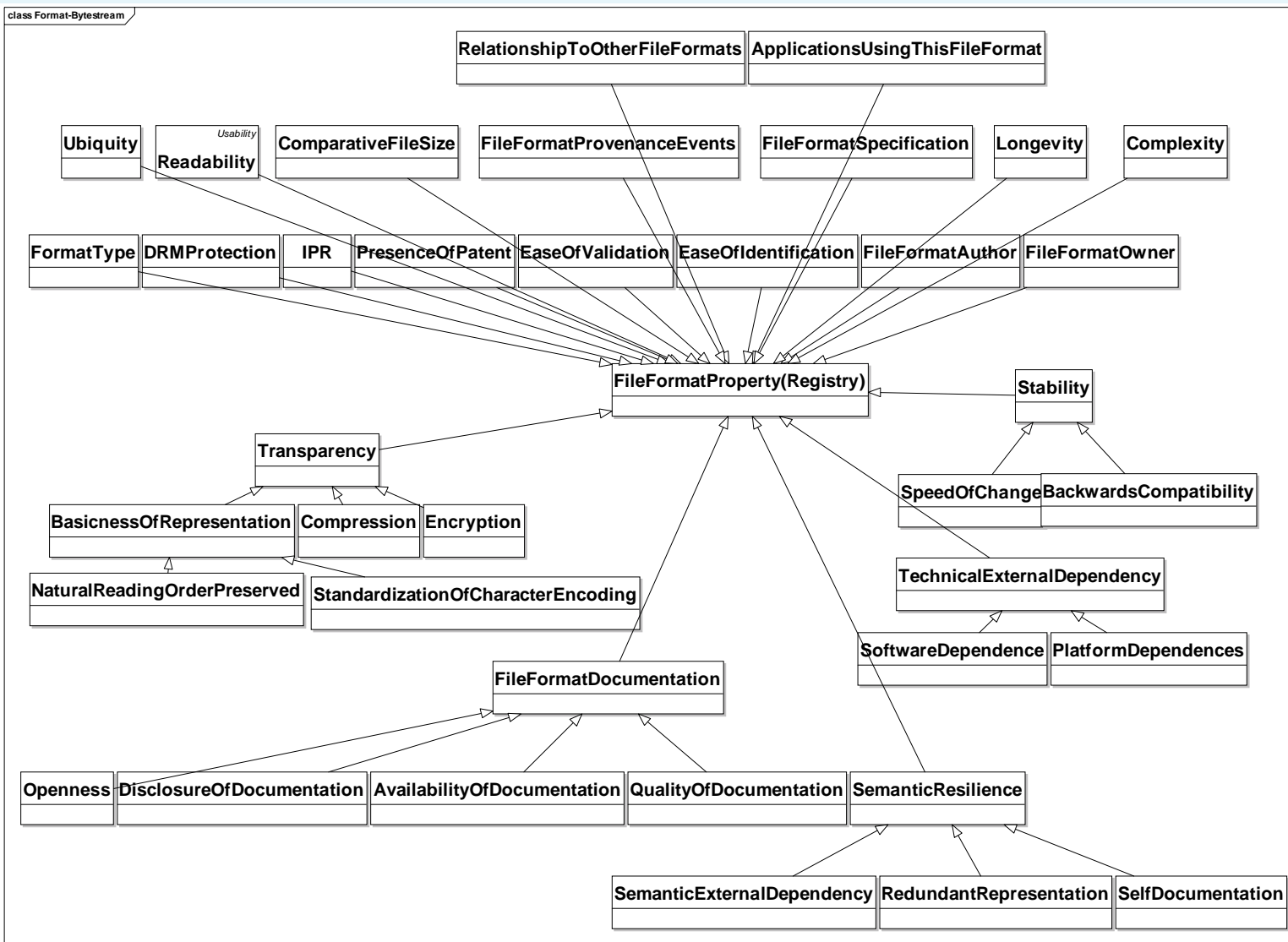- **Non-technical aspects, need to be more detailed.**

A factor which constrains a *Preservation Object* and that is necessary to interpret it.

- **Representation Information (OAIS)**

- **Environment Components at all Preservation Object levels**

Characteristics – Properties - Values

planets

class Format-Bytestream

RelationshipToOtherFileFormats

ApplicationsUsingThisFileFormat

Ubiquity

Usability
Readability

ComparativeFileSize

FileFormatProvenanceEvents

FileFormatSpecification

Longevity

Complexity

FormatType | DRMProtection | IPR | PresenceOfPatent | EaseOfValidation | EaseOfIdentification | FileFormatAuthor | FileFormatOwner

FileFormatProperty(Registry)

Stability

Transparency

SpeedOfChange | BackwardsCompatibility

BasicnessOfRepresentation | Compression | Encryption

NaturalReadingOrderPreserved | StandardizationOfCharacterEncoding

TechnicalExternalDependency

SoftwareDependence | PlatformDependences

FileFormatDocumentation

Openness | DisclosureOfDocumentation | AvailabilityOfDocumentation | QualityOfDocumentation | SemanticResilience

SemanticExternalDependency | RedundantRepresentation | SelfDocumentation

12