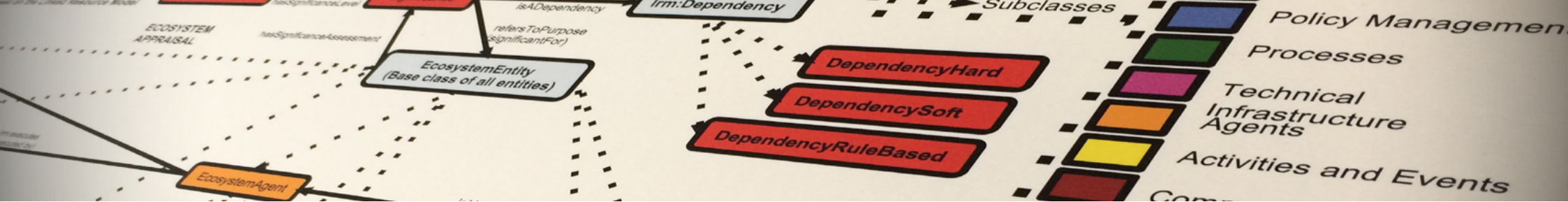# Taking a long-term perspective:

Maintaining digital assets by managing their evolving digital ecosystems

This article was written under the PERICLES FP7 project.

The Challenge of Change

As digital content and its related metadata are generated and used across different phases of the information lifecycle, and in a continually evolving digital environment, the concept of a fixed and stable 'final' version that needs to be preserved becomes less appropriate. As well as dealing with technological change and obsolescence, **long-term sustainability** requires us to **address changes in context**. This may include changes in policies, legal frameworks, professional practices, user expectations and behaviour, or semantics. This latter aspect has recently become a particular point of interest in information systems research, for example in investigations into the 'semantic drift' that arises from changes in language and meaning and its impact on the long-term reuse of digital information.

Such a changing environment necessitates a corresponding evolution of the strategies and approaches for preservation if stakeholder communities are to be able to continue to use and interpret content appropriately in the long-term. A key issue is the provision of sufficient contextual information to enable management and preservation of a digital object during its lifetime on the one hand,

and re-use or re-interpretation of content on the other, as well as the possibility to model and describe preservation processes, policies and infrastructures as they themselves evolve. Capturing and maintaining this information throughout an object's life, together with the complex relationships between the components of the preservation ecosystem as a whole, is key to an approach based on 'preservation by design', through models that capture intents and interpretative contexts associated with digital content, and enable content to remain relevant to new communities of users.

This approach is at the heart of **PERICLES**, a four-year integrated project (2013-2017) addressing the challenge of ensuring that digital content remains accessible in an environment that is subject to continual change. Funded by the European Union under its Seventh Framework Programme (ICT Call 9), PERICLES brings together a project consortium that unites academic institutions with expertise in various aspects of digital preservation and information analytics, commercial partners including SMEs and a multinational corporation, and non-academic public sector organisations with digital holdings in the creative arts, archives, and scientific data.

As change is an inevitable challenge in long-term digital preservation (LTDP), the question is how do we deal with it?

The research into a novel approach for managing the impact of change in digital environments carried out in PERICLES is supported and assessed for its relevance through case studies from the domains of space science and time-based media art, both of which deal with the creation and preservation of complex digital objects comprising multiple interdependent entities, such as files, audio-visual material, raw and processed data, hardware, software and documentation.

As change is an inevitable challenge in long-term digital preservation (LTDP), the question is how do we deal with it?
Existing approaches to preservation typically focus only on the technical environment necessary for using digital objects, and in particular addressing the issue of format change. The main task of digital preservation is thus often seen as managing this type of change to prevent technical obsolescence of the digital objects. In contrast PERICLES investigates how changes in any element of the environment of a digital object affect its usefulness and its interpretation, and how such change can be managed.
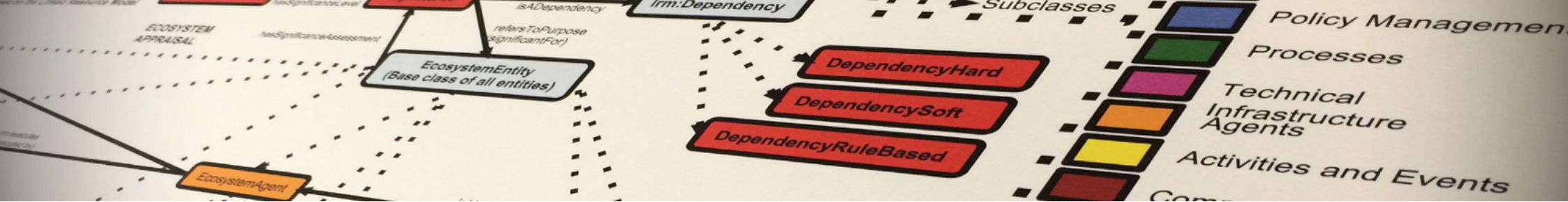
A consequence of approaches focussing on technical change is that a long-term archive is seen as the central component for digital preservation. While the key documents and standards do acknowledge that the task of such archives is not only technical, but also organisational, in practice institutions often overemphasise the technical aspects. The implicit and false assumption of many institutions seems to be that an archive as a technical system can solve their preservation challenge.

PERICLES tries to address a broader spectrum of aspects of change than just changes to the technical environment, such as:

- Change in user communities. This includes e.g. change in expectations, requirements, abilities and background knowledge which can all reduce the potential usefulness of a digital object.
- Changes in the institution. The aims that are reflected in the policies of an institution may change, and this may in turn have consequences for the retention of digital objects. The processes and workflows may change because of internal or external requirements, or because new technologies become available or old ones are no longer available.

- Relevant change may also occur in the larger social or cultural context and frameworks. Examples are changes in laws, disciplines or cultural norms, which require changes for the institutions and users.

Frequently, change is conceived as occurring at a specific point in time, but in reality change may be a slow, local and gradual evolution. The changes listed above do not need to be in themselves a threat to the usefulness of digital objects. It is primarily **unmanaged change** that may result in digital objects that are not usable or not as useful as they could be.

Digital Data and their Ecosystems

PERICLES is developing an approach that tries to extend the traditional focus of digital preservation practice on technological issues to include **dependencies** between all kinds of entities and change events of a variety of types. The central concept of this approach is that of a "**digital ecosystem**".

The usability of a digital object (DO) usually depends on access to different parts (entities) of its environment, whether that be metadata, software, policies, rights information, services, or the users themselves.
This is in particular valid for digital artworks, such as software-based installations, or space science data that relies on operation software for the extraction and interpretation of data.

It is through an understanding of these interactions and dependencies between entities that we are able to identify and deal with change so that the usefulness of the digital objects can be maintained in the long term.
We are not primarily examining ecosystems with a view to understanding the entities that are to be preserved, but rather we are
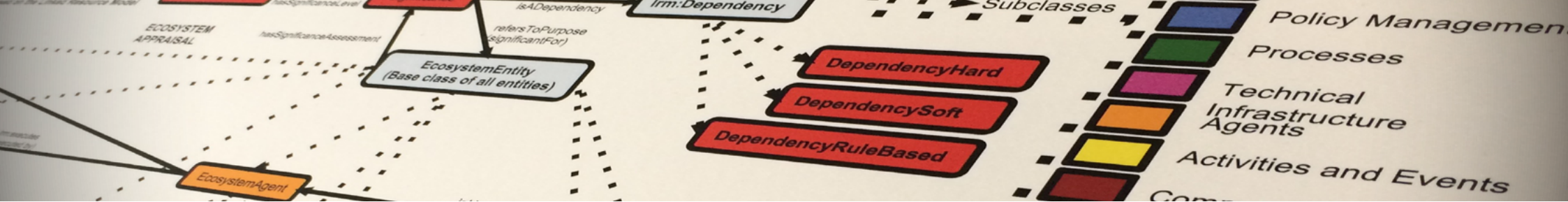
A digital ecosystem does not necessarily have to contain a dedicated preservation system or a digital archive at all

investigating how a digital ecosystem in principle should be managed so that the usefulness of the digital objects can be maintained.

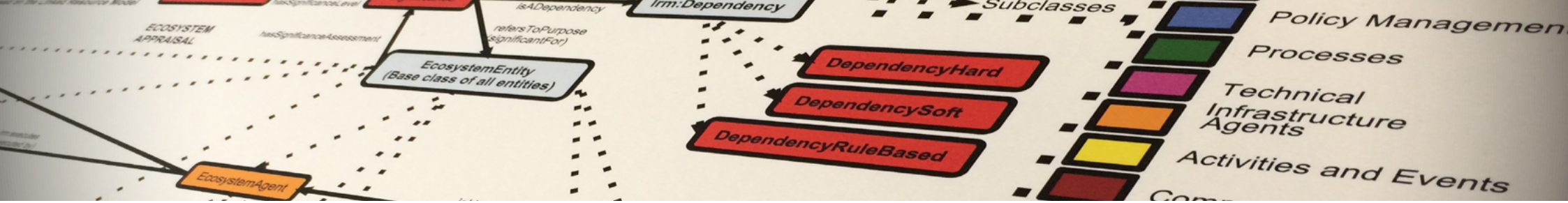In fact, the assumption in PERICLES is that a digital ecosystem does not necessarily have to contain a dedicated preservation system or a digital archive at all, if it is to maintain the usefulness of digital objects in the long term. To emphasise the long-term management aspect, we replaced the original term "preservation ecosystem" with preservation in "digital ecosystems".

We are using the term "digital ecosystem" instead of related terms such as "system", "environment" or "infrastructure" to emphasise the fact that we are talking about a diverse set of interacting entities that "evolve" over time often in an unplanned manner. In contrast an integrated system has been well planned in advance, and yet such a system may itself also "evolve" in an unplanned manner as changes gradually accumulate. Hence, in analogy to biological ecosystems, there is the notion of evolution, change and development. Related

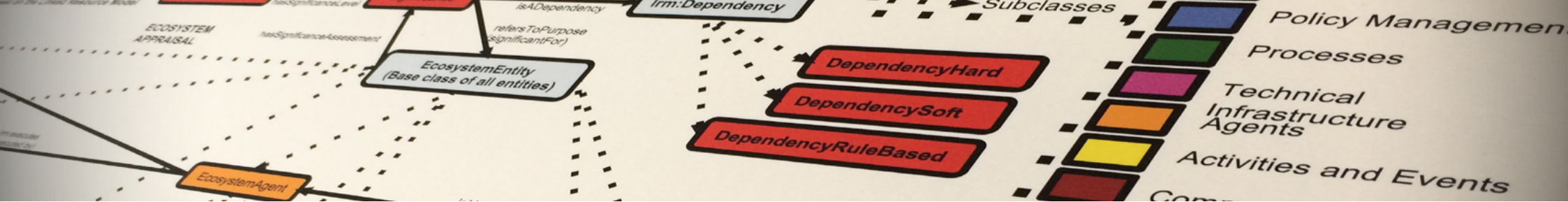to unintended and uncontrolled change is the observation that a digital ecosystem is not under the full control or authority of just a single actor or community. Individual institutions usually have limited influence on user communities, partner institutions, technology providers, legislation or society in general, all of which may affect the conditions of operation.
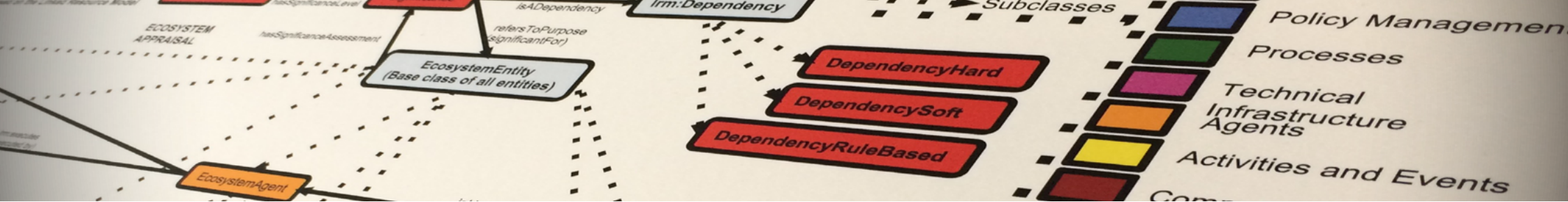
Model-driven Solution

The PERICLES focus on ecosystems that continually evolve in response to internal or external stimuli is underpinned by a **model-driven vision of digital preservation**. The use of models as abstract representations of reality is an established point of reference for existing approaches and practices in digital preservation. By developing a formal model of a digital ecosystem we can provide an even higher level of organisational planning and therefore help assess risks to sustainability and support tasks such as appraisal. In the context of PERICLES and LTDP, models play important role because they can also minimise the effort required to simulate change in a complex environment.

To describe the dependencies between resources, PERICLES has developed a computer-processable model called the **Linked Resource Model** (LRM). The LRM is an OWL ontology that is a semantic mark-up language for publishing and sharing **ontologies** on the World Wide Web. The LRM is also defined as an **upper-level ontology** because it is a formal description of relations between objects that can be

> By developing a formal model of a digital ecosystem we can provide an even higher level of organisational planning

applied across multiple knowledge domains.

Elaborating a model based on OWL ontologies presents two main advantages at very different conceptual levels. It is a way to organise, make explicit and share knowledge about ecosystems in general (schema level) and also a way to capture and reflect the structure and the evolution of a particular ecosystem (the data level).

Having a relevant representation of the target ecosystem offers many interesting perspectives: one can observe the evolving system through a well-organised and structured analysis grid, understand what happened in the past (provenance, in the broadest sense) and what may occur in the future. But, more interestingly, it brings some level of control for directing the evolution of the ecosystem, since we expect a preservation system not only to mirror the evolution of resources, but also to react to (and pre-empt) this evolution, to fix issues that may arise when some fundamental operational properties can no longer be ensured, and to
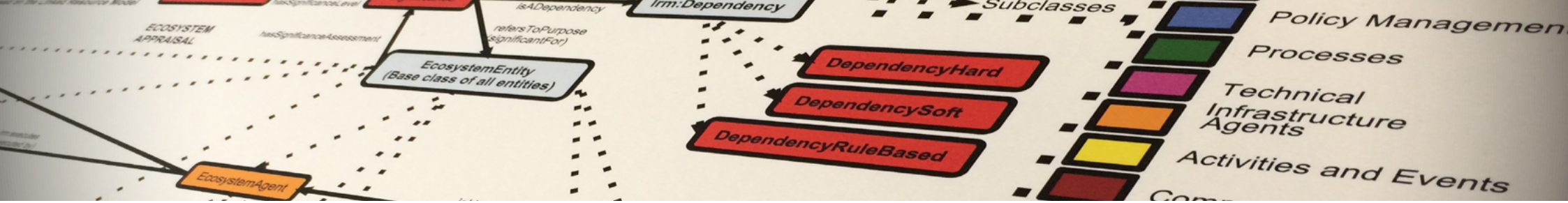
maintain a level of internal consistency as insurance for the future exploitation of the resources.

The working assumption is that based on the LRM it should be possible to develop a **formal model of a digital ecosystem**. Such a formal model would include the dependencies and interactions between all entities as a dependency graph and would allow simulation of the consequences of changes of different entities. Based on certain properties, we should be able to evaluate how suitable a digital ecosystem is from a long-term perspective with respect to a given institutional remit.

Another strand of PERICLES work has been focused on the production of two **domain-ontologies**, which are formal representations of concepts that relate to a specific domain. Following the two case studies we are developing a partial ontology for the Art and Media domain and a partial one for the Space Science domain based on the LRM. The main benefit of developing a domain-specific ontology lies in the ability to model digital preservation risks that
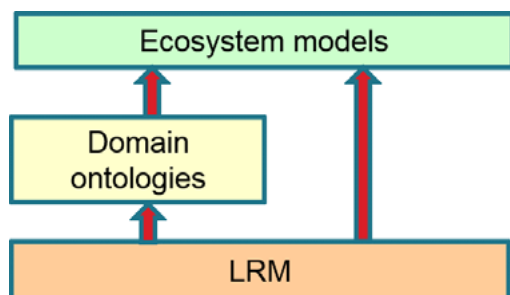
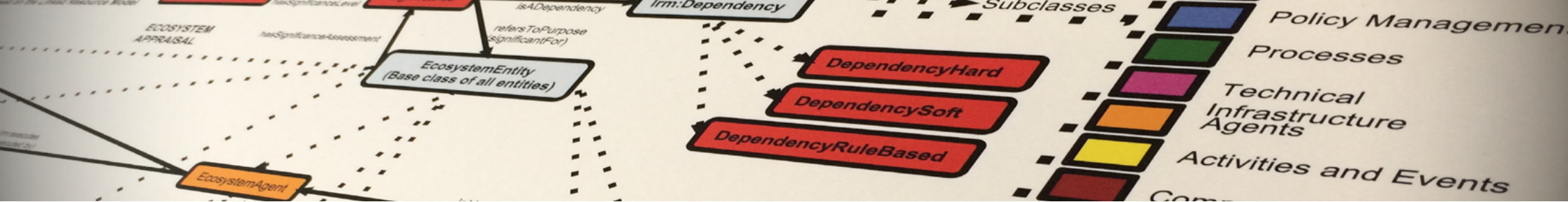are distinct to a particular sector, and hence offering more accurate insights for making decisions related to carrying out or reacting to change.

Integration Framework and Testbeds

Rather than attempting to provide a complete system, PERICLES is proposing a management approach detailed in an architectural framework for demonstrating workflows and components that would allow the introduction of a **"change management" layer** into existing repository systems. In support of this approach, the project partners are producing a variety of components (models, tools, ontologies etc.) that can be used independently in different combinations with available tools and components.

PERICLES will also deliver **testbeds** where different test scenarios will be carried out to validate the functionality of the various components and the overall framework for the model-driven management approach.

Two new subsystems have been designed to couple the workflow execution layer to the abstract models developed through the research activities of the project: the Entity Registry-Model Repository (ERMR) and the Process Compiler (PC). The ERMR will be used to coordinate the management of externally stored data and information about entities that are referenced in the models. The
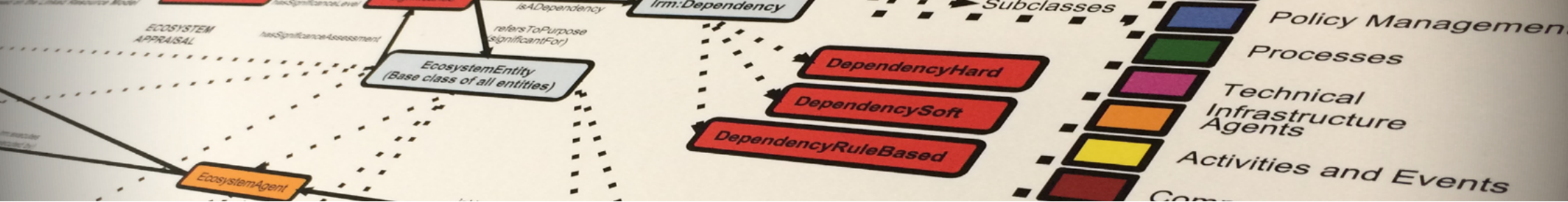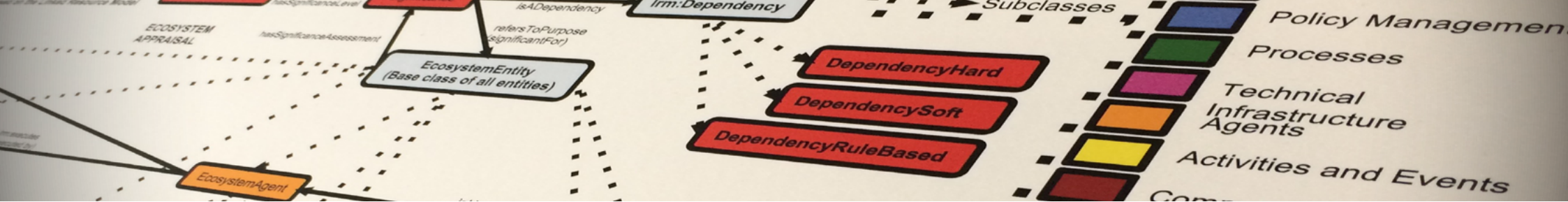
PC is designed to take one representation of a process model and transform it into another form that can be executed by the workflow engine. As part of this the PC will transfer information to the ERMR, which is used to update the process descriptions.

The resulting middleware is intended to support services based on the PERICLES models and interaction with the Linked Resource Model service (which provides semantic reasoning over LRM-based models), applying the project's change management functions to the testbed data and metadata. Together, these two subsystems provide the means to couple powerful semantic reasoning and policy-driven models to a "live" digital preservation system.
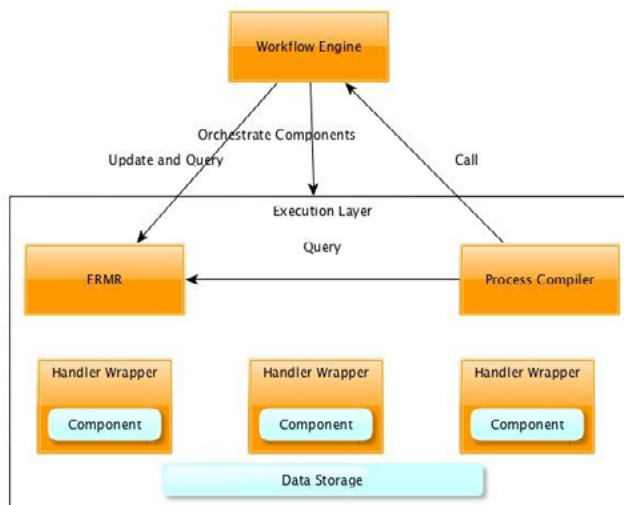
An example of an LRM Service is the following: A client asks for a change impact evaluation from the ERMR, which informs the LRM Service about changes occurring in the ecosystem (creation, deletion, update). It will

signal these as events, which may trigger reactions from the LRM Service. Depending on the configuration of the LRM, these events may lead to internal modifications and/or trigger calls to external services in order to perform some transformative actions (e.g., launching a command to verify the validity of an XML file, computing a digest, etc.).

The PERICLES integration framework is designed for the flexible execution of varied processing and control components in typical preservation workflows, while itself being controllable by abstract models of the overall preservation system. It is the project's focal point for connecting tools, models and application use cases for demonstrating the potential of model-driven digital preservation.

PERICLES is not building a complete preservation system – it is developing an approach that can be adopted by existing systems or used in new systems. The testbeds will not at any one time represent a complete implementation of the full PERICLES approach. How then will this enable the concepts and technologies from PERICLES to be applied in a real-world system? Can it be integrated into such
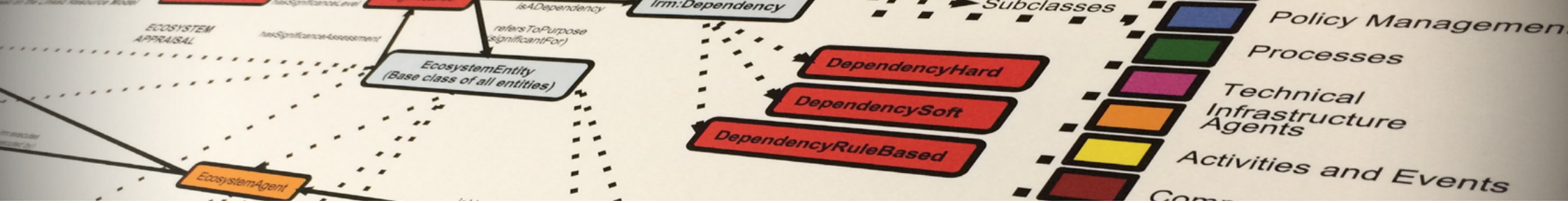
environments, or is PERICLES just building some nice demonstrations of no practical use? To answer these questions we look to our testbed.
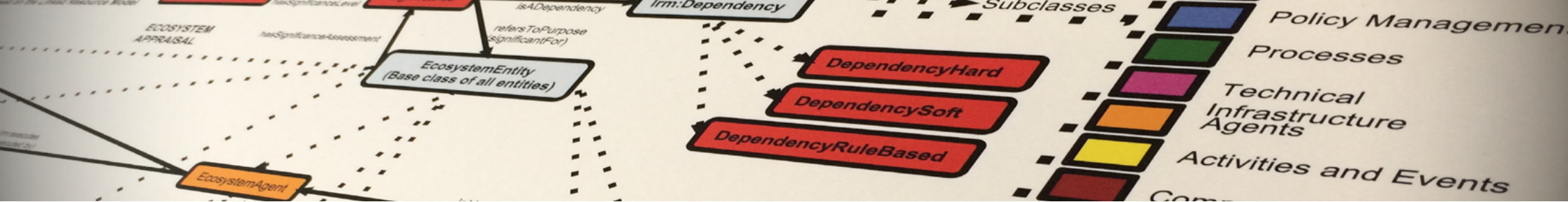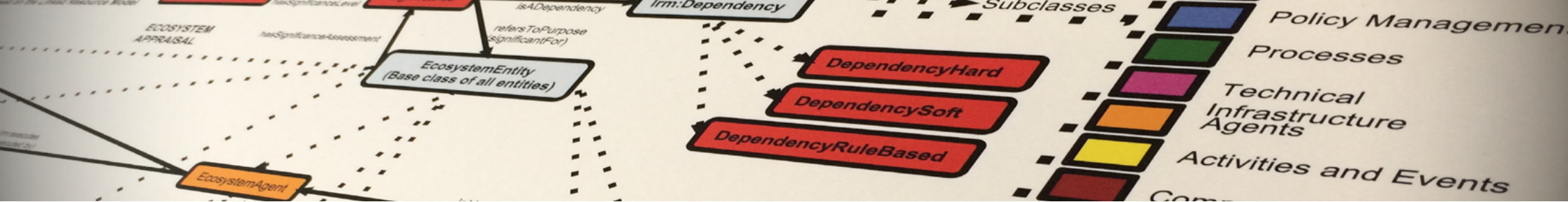
From Testbed to Reality

Although the PERICLES concepts could be used to create a new preservation system from the ground up for either replacing or starting where no such system existed, it should be noted that the concepts and technology could also be used to augment and upgrade existing systems. While superficially these are two very different tasks, the actual processes involved would be quite similar which becomes obvious through the testbeds.

Naturally a testbed is not the real world system in that it creates for each test run a new "pristine" installation, can use "Controlled Usage Patterns" in contrast to the unpredictable reality, and while in the real world errors occur unexpectedly, in tests you induce known faults at predetermined times.

However, if there were not key similarities between a testbed and the real world system, then it wouldn't make much sense to do test runs. The key points of similarity are that technology used in the testbed is the same as the one to be used in the real world system, real world process

> While in the real world errors occur unexpectedly, in tests you induce known faults at predetermined times

PERICLES is proposing a compelling approach to the communities engaged in managing and preserving their digital asset

patterns which are created in concert with the domain experts are created and the same data types shared.
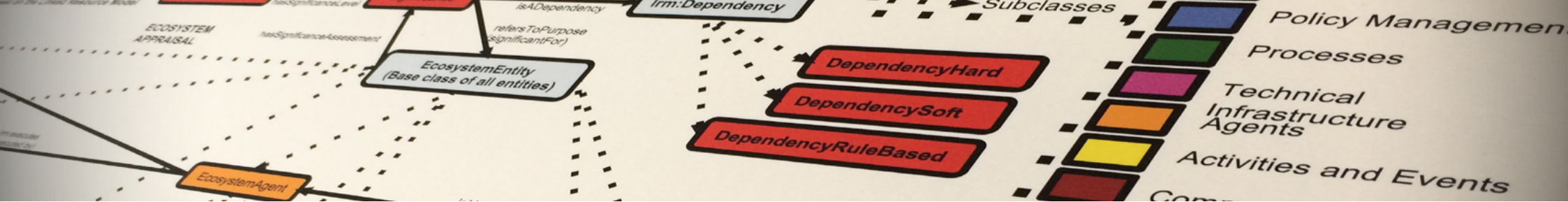
Our testbeds and test runs are designed to be applicable for both situation: creating a new system or complement an existing system.

Our guiding principles are:

- Incremental start-up: not everything at once
- Use existing systems: bridging
- Map processes: don't throw out current processes
- Differences: a chance to improve

PERICLES through its model-driven and runtime service instantiation could lead to different approaches to process development – initially many processes will be directly mapped but as the level of knowledge about the approach increases, so too does the possibilities of changes to how things work.

PERICLES is proposing a compelling approach to the communities engaged in managing and preserving their digital asset. With their support, we are looking at attracting IT industries to take up the next steps to facilitate implementation of the model-driven preservation into real-world management. ■

>> Further readings:

- Deliverables

- Publications

- Videos

>> Contact:

info@pericles-project.eu