

Planets: Preservation and Long-Term Access through Networked Services

VIII Workshop REBIUN sobre Proyectos Digitales
"La Preservación Digital : Memoria de Futuro"
Murcia, October 2008

Christoph Becker
Vienna University of Technology

www.ifs.tuwien.ac.at/~becker



partially based on slides by Adam Farquhar, Planets project lead



Why do we need Digital Preservation?

- . . .
 - . . .
 - . . .
 - . . .
- . . .
 - . . .
 - . . .
 - programs won't
 - . . .
- . . .

.....

Why do we need Digital Preservation?

- Digital Objects require specific environment to be accessible :
 - Files need specific programs
 - Programs need specific operating systems (-versions)
 - Operating systems need specific hardware components
 - SW/HW environment is not stable:
 - Files cannot be opened anymore
 - Embedded objects are no longer accessible/linked
 - Programs won't run
 - Information in digital form is lost
(usually total loss, no degradation)
 - Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.
-

The Longevity of Digital Objects

- Digital objects are the dominant way we exchange information
- Digital objects need technical environment to “function”
- Environments change
- Heterogeneity and complexity of file formats and speed of technological change make long-term access a challenge
- Digital preservation: Long-term access to digital objects
- Dominant types of preservation actions:
 - Migration
 - Emulation



Agenda

- Introduction to Planets
 - Who are we?
 - What are we doing?
 - Why are we doing it?
- The Planets architecture and components
- Preservation Planning in Planets
- Progress and next steps



The Planets project

- Addresses core digital preservation challenges
- Builds on strong digital archiving and preservation programmes
- 4-year research and technology development project co-funded by the European Union
- Started June 2006 with €15m budget
- Coordinated by the British Library
- 16 library, archive, research, technology partners



The Planets project

- 4-year research and technology development project co-funded by the European Union
- Address core digital preservation challenges
- Started June 2006 with €15m budget
- Coordinated by the British Library
- 16 partners
 - national libraries and archives
 - leading technology companies
 - research universities
- Builds on strong digital archiving and preservation programmes



Planets partners

BRITISH LIBRARY

KB

Koninklijke Bibliotheek



STATS BIBLIOTEKET

Österreichische

Nationalbibliothek



- The British Library
- National Library, Netherlands
- Austrian National Library
- State and University Library, Denmark
- Royal Library, Denmark



DET KONGELIGE BIBLIOTEK

NATIONALBIBLIOTEK OG KØBENHAVNS UNIVERSITETSBIBLIOTEK



the national archives



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

- National Archives, UK
- Swiss Federal Archives
- National Archives, Netherlands

nationaal archief



Planets partners



- Tessella Plc
- IBM Netherlands
- Microsoft Research
- Austrian Research Centers GmbH



rechenzentrum
universität freiburg



- Hatii at University of Glasgow
- University of Freiburg
- Vienna University of Technology
- University of Cologne



The Planets team



All Staff Meeting, February 2007



Planets goal

Build practical services and tools offering national libraries and archives an increased ability to ensure long-term access to their digital cultural and scientific assets



Basic assumptions

- Digital assets have real long-term value
- The importance, size, and heterogeneity of digital collections is growing
- Technology change makes digital assets increasingly difficult to access
- Digital assets age and rapidly become damaged

New digital preservation technology can reduce costs and unlock access to older digital material



Aims and objectives

- **Increase Europe's ability to ensure long-term access to its cultural and scientific heritage**
 - Improve decision-making
 - Control costs through increased automation and scalable infrastructure
 - Ensure wide adoption across the user community
 - Establish a market place for preservation services and tools
- **Build practical, comprehensive solutions**
 - Integrate existing expertise, designs and tools
 - Deliver tools and services for operational environments



Benefits for archives and libraries

Planets will enable you to

- Express your preservation policies
- Profile your digital collections
- Identify and diagnose problems in your digital collections
- Compare different treatment plans
- Select and implement treatments
- Verify that the treatment was successful
- Know which solutions work through empirical evidence
- Encourage vendors and service providers to provide these capabilities to you

*“Planets will provide the technology component of our digital preservation solution” Richard Boulderstone, BL Director,
15/06/07*

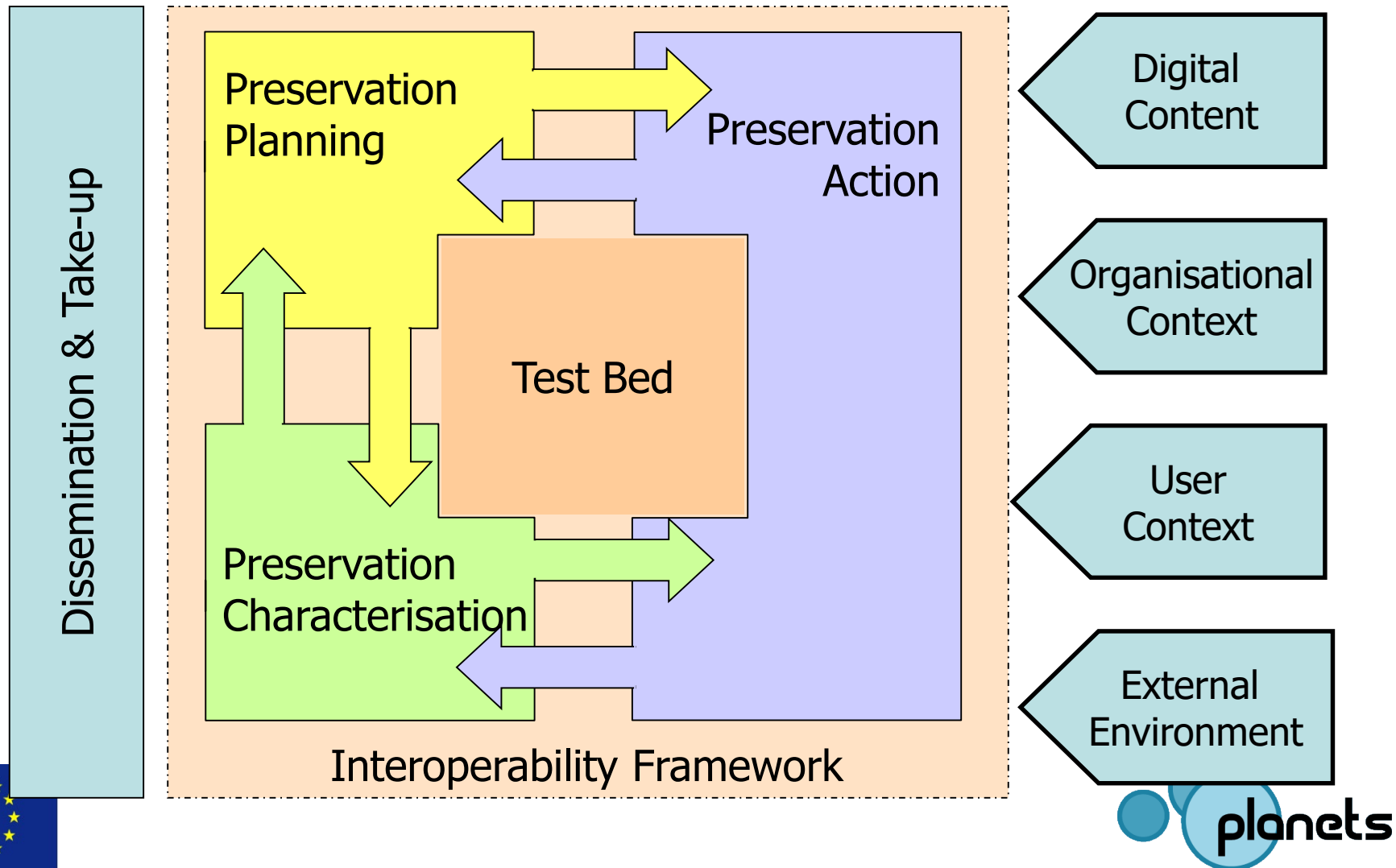


Project Structure

- **Preservation Planning** services that empower organisations to define, evaluate and execute preservation plans
- Methods, tools and services for the **Characterisation** of digital objects
- **Preservation Action** tools which to transform and emulate digital assets
- An **Interoperability Framework** to seamlessly integrate Planets tools and services in a distributed service network;
- A **Testbed** to provide a consistent and coherent evidence-base for the objective evaluation of protocols, tools, services and complete preservation plans
- A comprehensive **Dissemination and Take-up** programme to ensure stakeholder adoption and training, as well as feedback



Project structure



Preservation planning – what should I do?

- Empower organisations to define, evaluate, and execute preservation plans
- Establish a planning methodology
- Understand the needs of content holders
- Understand the needs of content users
- Provide support, automation with the planning tool **Plato**
 - Help organisations build, select, document, and execute plans
- Understand the profile of a collection
- Provide technology watch services



Preservation characterisation – what do I have?

- Characterise the relevant properties of digital objects
 - Identify formats
 - Extract properties
 - Assess risks
- Characterisation framework brings together tools to identify file formats and extract object properties
- Characterisation registry holds the format and property information
- eXtensible Characterisation Languages (XC*L) allows automated extraction and comparable descriptions
- Comparator measures differences between descriptions and the impact of actions undertaken



Preservation action – what can I do?

- Transform content: Migration
 - Web service infrastructure for third-party tools to handle many common formats
 - Database migration: SIARD
- Transform environments: Emulation
 - GRATE and Dioscuri emulate older hardware/software environments
 - Universal Virtual Computer (UVC) provides a layered durable approach to emulation
- Preservation Action registry
 - holds descriptions of tools and services so that they can be found, compared, and invoked depending on the needs of particular content



Testbed – what really works?

- Develop a scientific evidence-base to evaluate tools, services, plans, and approaches
 - Better data for better decisions
- Define and run experiments locally with the Testbed application
- View, define, and run experiments on the supported Testbed service
- Use Testbed data in the Planets registries and in preservation planning
- Benchmark outcomes against previous experiments



Interoperability framework

- The glue to hold the Planets tools and services together
- Enable third-parties to plug-in tools and services
- Enable vendors to integrate Planets preservation services
- Simple configurable installation
- Support for scalable distributed deployment
- Execute and manage workflows
- Provide shared services



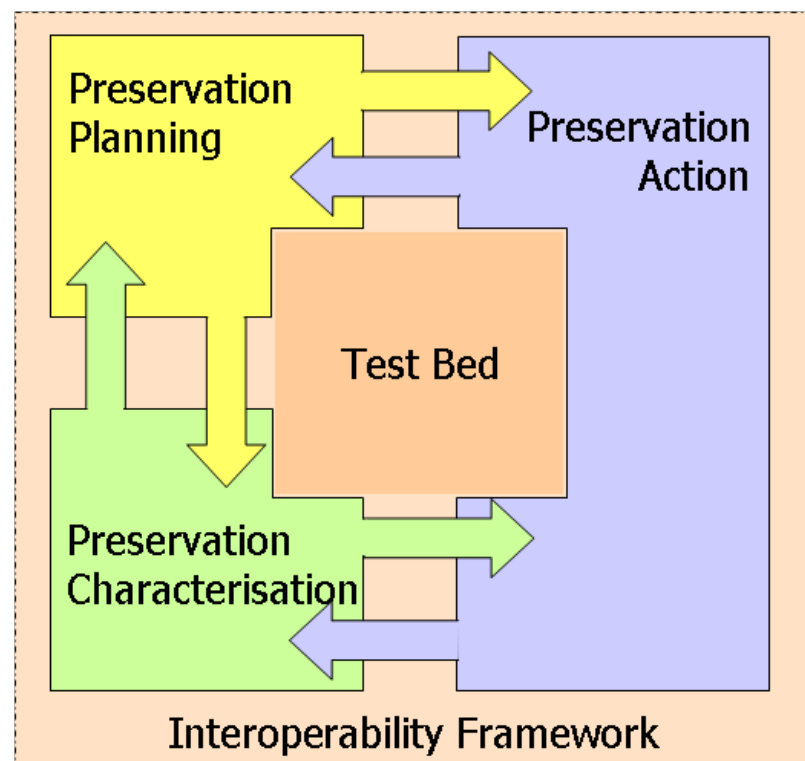
Dissemination and take-up

- Ensure communities are aware of Planets and its value in easing preservation activity
- Ensure that Planets learns about community needs
- Collaborate with European and national organisations
- Explore and stimulate opportunities for take-up beyond project closure
- Track progress and impact



Cross-project links

- Interaction with standards, projects, and organisations
- JHOVE, JHOVE2, GDFR
- OAIS
- PREMIS
- METS
- OOXML
- PDF
- CASPAR, DPE, SHAMAN



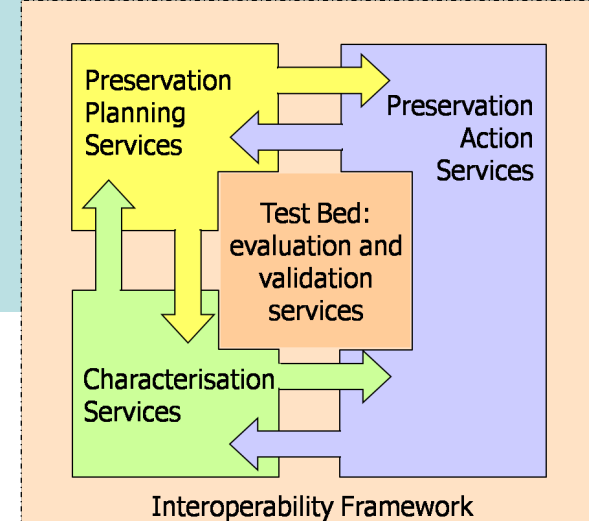
Agenda

- Introduction to Planets
 - Who are we?
 - What are we doing?
 - Why are we doing it?
- The Planets architecture and components
- **Preservation Planning in Planets**
- Progress and next steps



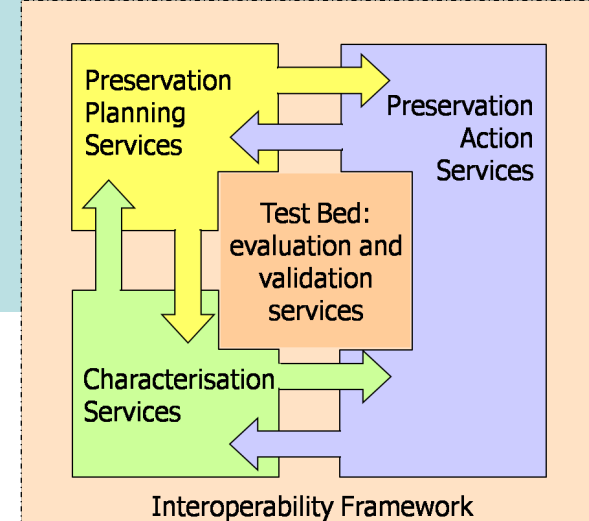
Preservation planning

- Collection profiling services
- Technology watch services
- Risk assessment of digital objects
- Preservation planning methodology
- Tool support: Plato, the Planning Tool



Preservation planning

- Collection profiling services
- Technology watch services
- Risk assessment of digital objects
- **Preservation planning methodology**
- **Tool support: Plato, the Planning Tool**



Evaluating preservation strategies

- Variety of solutions and tools exist
 - Each strategy has unique strengths and weaknesses
 - Requirements vary across settings
 - Decision on which solution to adopt is complex
 - Documentation and accountability is essential
-
- Preservation planning assists in decision making
 - Evaluation of strategies on representative sample content according to specific requirements



Decision support for preservation planning

- Systematic procedure for evaluating preservation strategies and building preservation plans
 - By conducting experiments on sample content
- Case studies
 - Electronic documents, interactive art, web archives...
 - Identify essential characteristics of objects and requirements for preservation strategies
 - Evaluate strategies and build plans
- Develop a decision support software
 - Plato – Planning Tool
 - Web application supporting the workflow



Scenario: Changes in user community

- Repository of electronic publications
- Policy: 90% of users can access all published reports
- Usage profile: 98% of users can not view dvi files
- Content profile: 5% of published reports in dvi format
- Mission: Build and execute a plan for preserving access to these documents for the designated user community



Scenario (2)

- Create a preservation plan with Plato
 - Define requirements
 - Identify possible actions using registries
 - Convert to PDF with Tool A
 - Convert to TIFF with Tool B
 - Provide users with a viewer plug-in
 - Evaluate actions on sample content
 - Build a preservation plan
- Convert content (using data registry)
- QA results (using comparison services)
- Ingest results into repository (using adaptor)

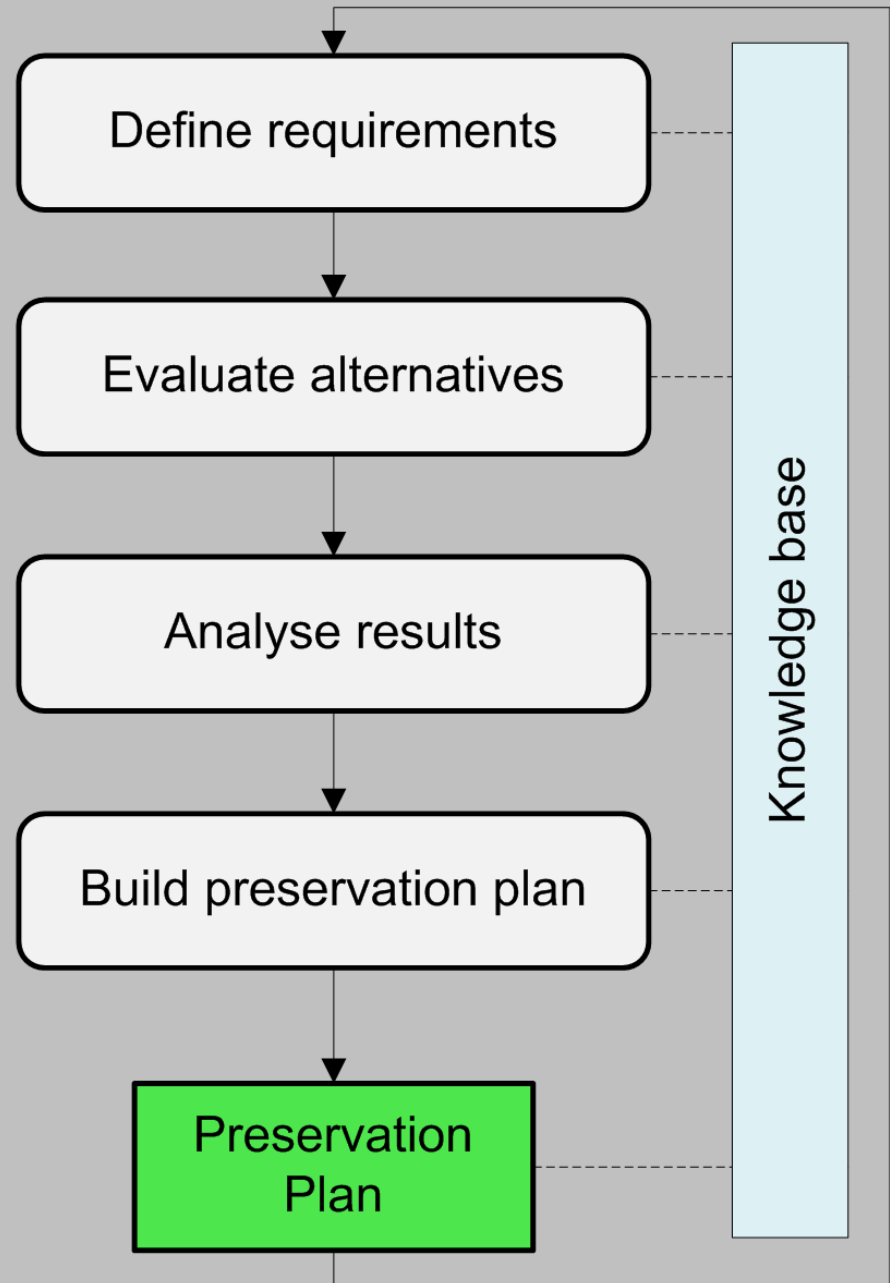


Workflow

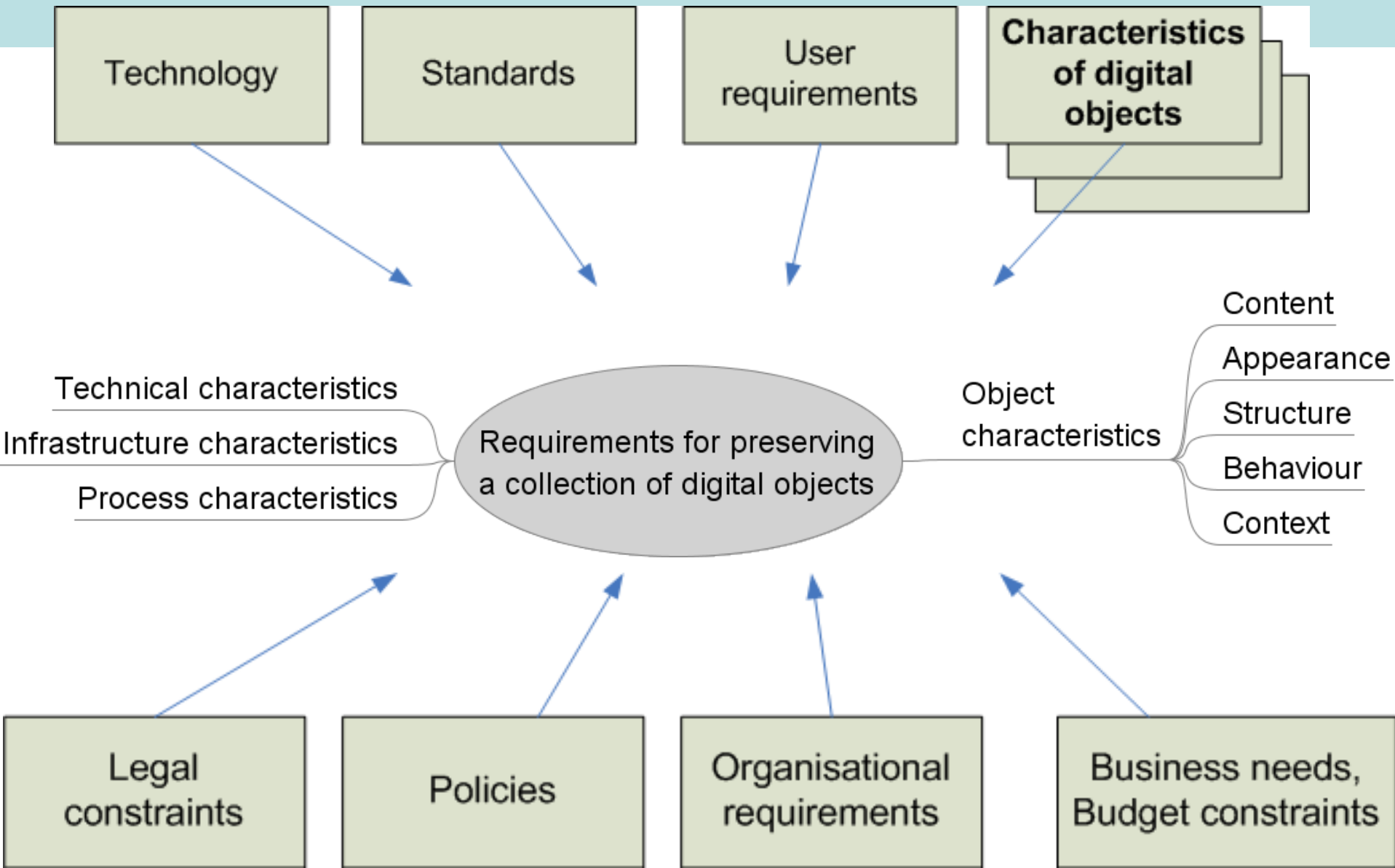
1. Define requirements
2. Evaluate potential actions
3. Analyse the results
4. Build a preservation plan



Preservation Planning in Plato

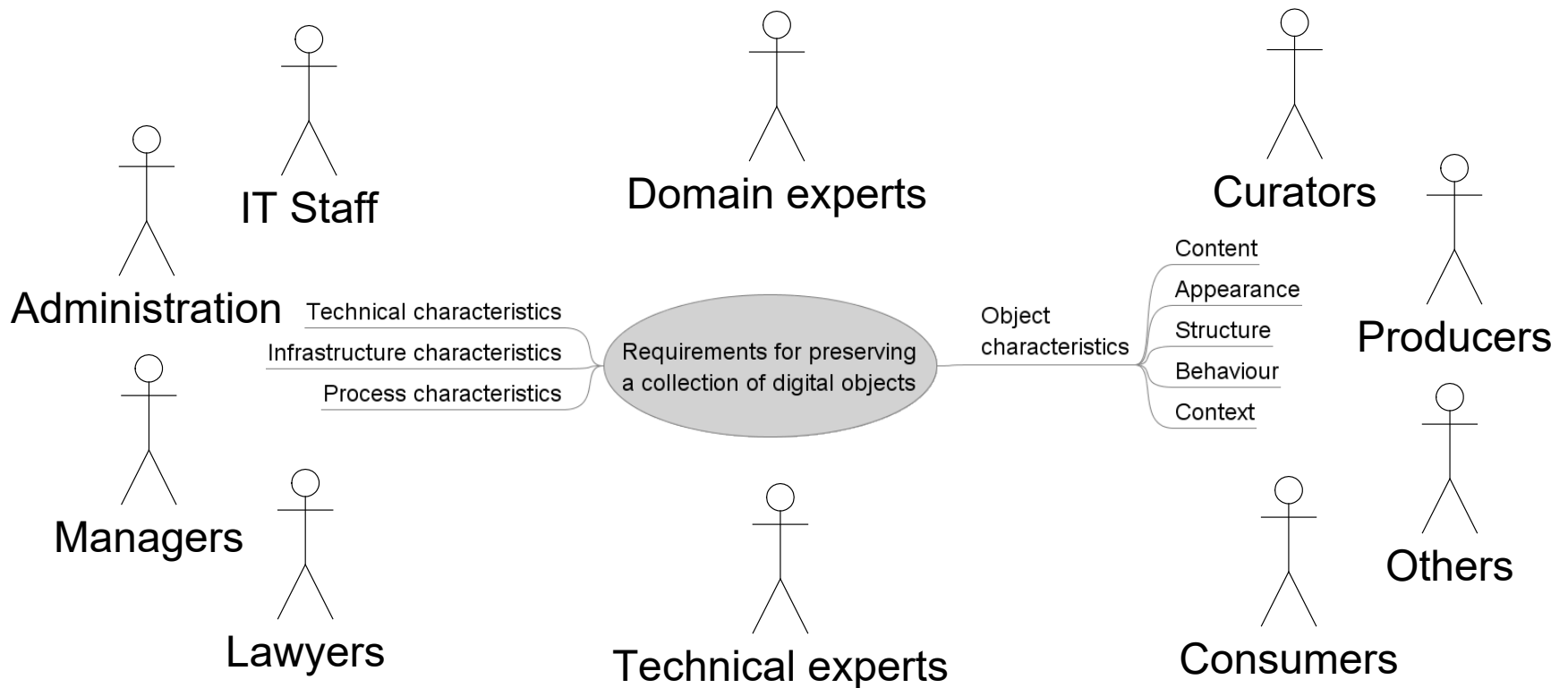


Influence Factors

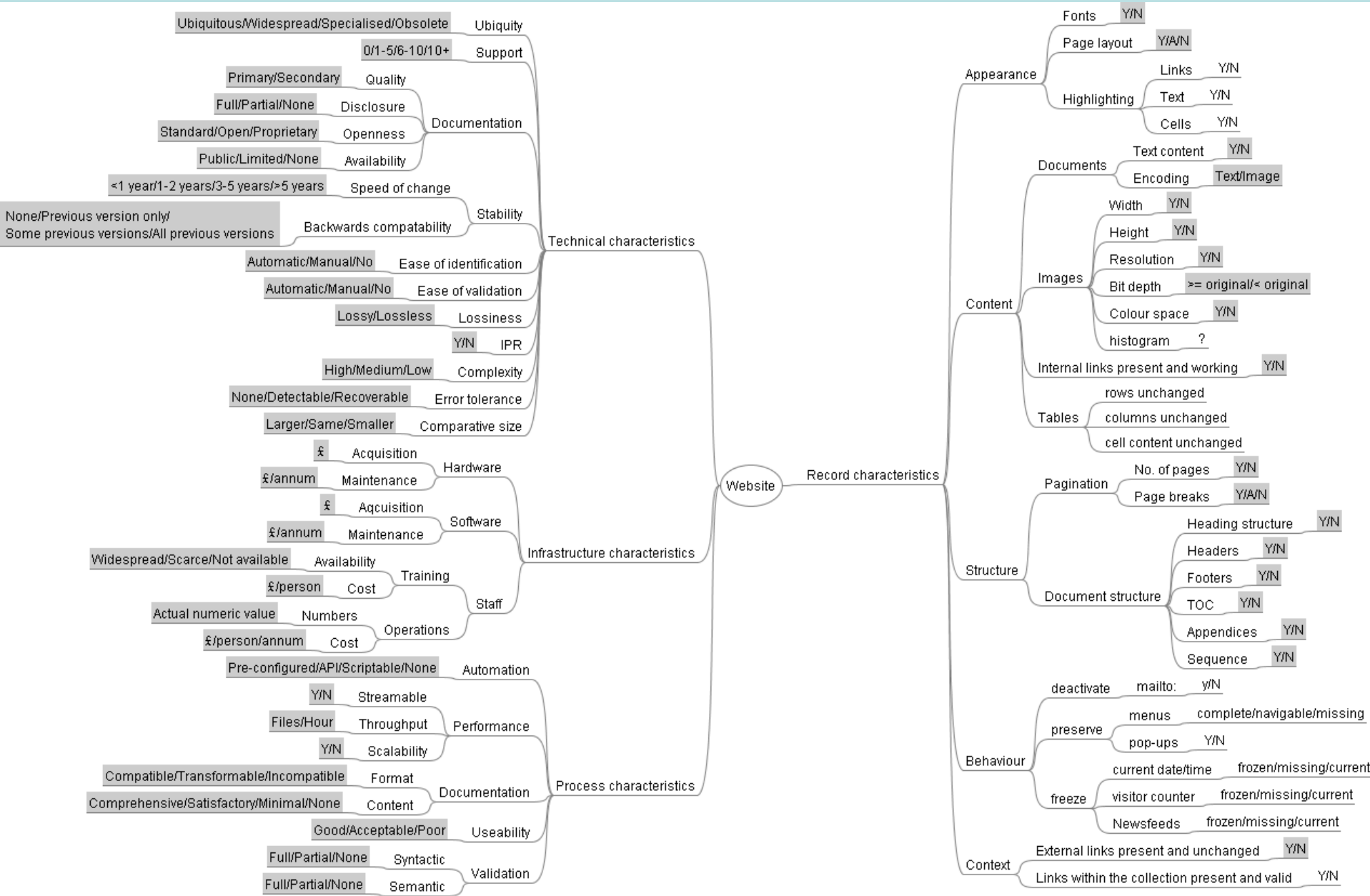


Stakeholders

- Input from a wide range of persons, depending on the institutional context and the collection



The Objective Tree



The Objective Tree



PLANETS Preservation Planning Tool (*Plato*)



[\[logout admin\]](#) [\[help\]](#)

Project | Define Requirements | Evaluate Requirements | Consider Results |

PP4 workshop - The National Archive



Identify Requirements

[Objective Tree](#)

[Descriptive Information](#)

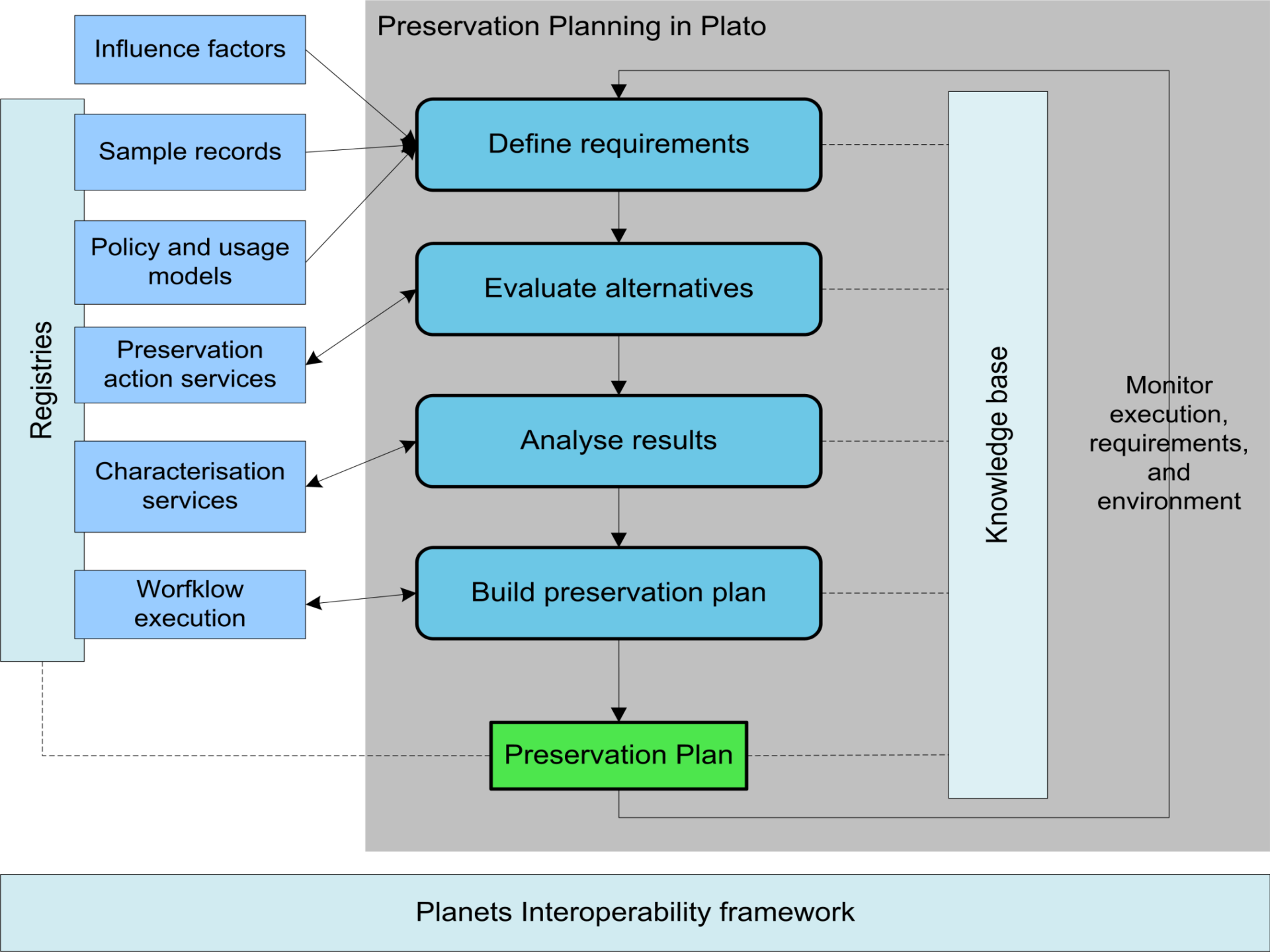
[How can I define the objective tree?](#)

[1] Objective Tree

[Expand All](#) | [Collapse All](#)

[Website](#)

Focus	Node	+	+	-	Single	Scale	Restriction	Unit
	Website	+	+	-				
X	Record characteristics	+	+	-				
X	Appearance	+	+	-				
X	Content	+	+	-				
X	Structure	+	+	-				
X	Behaviour	+	+	-				
X	deactivate	+	+	-				
X	mailto:	+	+	-	☐	Boolean	Yes/No	
X	preserve	+	+	-				
X	menus	+	+	-	☐	Ordinal	complete/navigable/missing	
X	pop-ups	+	+	-	☐	Boolean	Yes/No	
X	freeze	+	+	-				
X	current date/time	+	+	-	☐	Ordinal	frozen/missing/current	
X	visitor counter	+	+	-	☐	Ordinal	frozen/missing/current	
X	Newsfeeds	+	+	-	☐	Ordinal	frozen/missing/current	
X	Context	+	+	-				
X	External links present and und	+	+	-	☐	Boolean	Yes/No	
X	Links within the collection pres	+	+	-	☐	Boolean	Yes/No	
X	Technical characteristics	+	+	-				
X	Ubiquity	+	+	-	☐	Ordinal	Ubiquitous/Widespread/Specia	
X	Support	+	+	-	☐	Ordinal	0/1-5/6-10/10+	
X	Documentation	+	+	-				



Empirical evaluation of actions

- ❑ Define representative sample content
- ❑ Discover applicable actions in service registries
- ❑ Apply potential actions to sample content
- ❑ Evaluate outcomes
- ❑ Select most suitable action(s)
- ❑ Define an (executable) preservation plan



Analysing results

Results: Weighted multiplication

Result-Tree with all Alternatives, Aggregation method: Weighted multiplication

[Expand All](#) | [Collapse All](#)

National Library Publications

Focus	Name	Result
	▼ National Library Publications	Adobe Acrobat->DOC: 0,00 Convert Doc->DOC: 3,44 Adobe Acrobat->HTML:3,18
X	► Object characteristics	Adobe Acrobat->DOC: 1,55 Convert Doc->DOC: 1,63 Adobe Acrobat->HTML:1,52
X	► Technical characteristics	Adobe Acrobat->DOC: 1,14 Convert Doc->DOC: 1,14 Adobe Acrobat->HTML:1,16
X	▼ Process Characteristics	Adobe Acrobat->DOC: 0,00 Convert Doc->DOC: 1,14 Adobe Acrobat->HTML:1,08
X	► Duration	Adobe Acrobat->DOC: 0,00 Convert Doc->DOC: 1,23 Adobe Acrobat->HTML:1,06
X	► Automation of the process	Adobe Acrobat->DOC: 1,55 Convert Doc->DOC: 1,90 Adobe Acrobat->HTML:1,55
X	► Integrity	Adobe Acrobat->DOC: 1,00 Convert Doc->DOC: 1,00 Adobe Acrobat->HTML:1,00
X	► Costs	Adobe Acrobat->DOC: 1,67 Convert Doc->DOC: 1,63 Adobe Acrobat->HTML:1,67

Summary

- Repeatable, decision making process
- Evaluate potential actions empirically
- Analyse results, take well-documented decision and build a preservation plan
- The planning tool Plato

PLANETS Preservation Planning Tool (Plato)

Project Define Requirements Evaluate Requirements Consider Results PP4 workshop - The National Archive [logout admin] [help]

Identify Requirements

Objective Tree Descriptive Information

How can I define the objective tree?

Objective Tree

Expand All Collapse All Website

Focus	Node	Single	Scale	Restriction	Unit
X	Website				
X	Record characteristics				
X	Appearance				
X	Content				
X	Structure				
X	Behaviour				
X	deactivate				
X	mailto:		Boolean	Yes/no	
X	preserve				
X	menus		Ordinal	complete/navigable/missing	
X	pop-ups		Boolean	Yes/no	
X	freeze				
X	current date/time		Ordinal	frozen/missing/current	
X	visitor counter		Ordinal	frozen/missing/current	
X	Newsfeeds		Ordinal	frozen/missing/current	
X	Context				
X	External links present and und		Boolean	Yes/no	
X	Links within the collection pres		Boolean	Yes/no	
X	Technical characteristics				
X	Ubiquity		Ordinal	Ubiquitous/Widespread/Special	
X	Support		Ordinal	0/1-5/6-10/10+	
X	Documentation				

Release 1.1 - Institute of Software Technology and Interactive Systems: «off-ice bears» Quick Access

Results: Weighted multiplication

Result-Tree with all Alternatives, Aggregation method: Weighted multiplication
Expand All Collapse All
National Library Publications

Focus	Name	Result
	▼ National Library Publications	Adobe Acrobat->DOC: 0,00 Convert Doc->DOC: 3,44 Adobe Acrobat->HTML:3,18
X	► Object characteristics	Adobe Acrobat->DOC: 1,55 Convert Doc->DOC: 1,63 Adobe Acrobat->HTML:1,52
X	► Technical characteristics	Adobe Acrobat->DOC: 1,14 Convert Doc->DOC: 1,14 Adobe Acrobat->HTML:1,16
X	▼ Process Characteristics	Adobe Acrobat->DOC: 0,00 Convert Doc->DOC: 1,14 Adobe Acrobat->HTML:1,08
X	► Duration	Adobe Acrobat->DOC: 0,00 Convert Doc->DOC: 1,23 Adobe Acrobat->HTML:1,06
X	► Automation of the process	Adobe Acrobat->DOC: 1,55 Convert Doc->DOC: 1,90 Adobe Acrobat->HTML:1,55
X	► Integrity	Adobe Acrobat->DOC: 1,00 Convert Doc->DOC: 1,00 Adobe Acrobat->HTML:1,00
X	► Costs	Adobe Acrobat->DOC: 1,67 Convert Doc->DOC: 1,63 Adobe Acrobat->HTML:1,67

What's next in Planets? (1/2)

- Preservation Planning tool: Plato 2.0
 - Version 1.3 available
 - Version 2.0 will be published in November
 - Integration of registries and services
- Integrated preservation planning services
 - risk assessment
 - automated collection profiling
 - technology watch
- Planets-compliant migration tools for digital objects
- Database migration: SIARD
- Emulation tools for specific environments



What's next in Planets? (2/2)

- Characterisation tools
 - extract significant properties from digital objects
 - compare different objects
- A characterisation description and extraction language (XCL)
- Characterisation and preservation action registries
- A controlled environment for the empirical assessment of services (Planets Testbed)
- Planets Interoperability Framework
 - as downloadable “click-and-install” software package



Dissemination and Take-up programme

- Workshops and training events
- Courses for example in
 - Vilnius, October 2007
 - Vienna, April 2008
 - Pisa, June 2008
 - London, July 2008
 - Prague, October 2008
 - **...Valencia, Spring 2009**
- Scientific publications
- Newsletter and web page



Conclusion

- Planets methods, tools, and services help organisations diagnose and treat problems with their digital objects
- High levels of automation and scalable components reduce costs and improve quality
- Empirical data enables improved decision making
- Find out more: <http://www.planets-project.eu>



Thank you very much for your attention.

www.planets-project.eu

becker@ifs.tuwien.ac.at
www.ifs.tuwien.ac.at/~becker

