Project no. 034762

DigitalPreservationEurope (DPE)

Instrument: Coordination Action

Thematic Priority: IST-2005-2.5.10

Access to and preservation of cultural and scientific resources

D3.2 Repository Planning Checklist and Guidance

Due date of deliverable: 29.02.2008

Actual submission date: 25.03.2008

Start Date of Project: 01 April 2006

Duration: 36 Months

Name of Organisation of Lead Contractor for this Deliverable:

SB

Names of Partners Engaged in this Deliverable:

HATII, MIBAC, NANETH, NKP, SB

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Dissemination Level: **PU (Public)**

Legal Notices

Catalogue Entry

Title	Repository Planning Checklist and Guidance
Creator	Statsbiblioteket, HATII
Subject	Information Technology; Science; Technology--Philosophy; Computer Science; Digital Preservation; Digital Records; Science and the Humanities.
Description	This document introduces the PLATTER toolkit which assists repositories in setting the necessary objectives and targets for achieving trustworthiness.
Publisher	HATII at the University of Glasgow, includes FernUniversität Hagen, Fondazione Rinascimento Digitale, Ministero Per I Beni E Le Attività Culturali, Národní knihovna Ceske republiky, Nationaal Archief van Nederland, Statsbiblioteket, Technische Universität Wien, Vilnius University Faculty of Communication.
Contributor	HATII
Contributor	MIBAC
Contributor	NANETH
Contributor	NKP
Contributor	SB
Date	25/03/2008
Type	Text
Format	Adobe Portable Document Format v.1.3 generated from OpenOffice.org original
Resource Identifier	ISSN
Resource Identifier	ISBN
Resource Identifier	DOI
Language	English

Title	Repository Planning Checklist and Guidance
Rights	© DigitalPreservationEurope Partners

Citation Guidelines
DigitalPreservationEurope, (April 2008), "DPE Repository Planning Checklist and Guidance DPE-D3.2", Retrieved from:
http://www.digitalpreservationeurope.eu/publications/reports/Repository_Planning_Checklist_and_Guidance.pdf

Document Version Control

Version	Date	Change Made (and if appropriate reason for change)	Initials of Commentator(s) or Author(s)
0.1	07/03/2008	Initial Version	CSR, ABR, JH, AM
0.2	14/03/2008	Corrections following QA by SS	CSR
1.0	25/03/2008	Amendments following HATII QA	CSR, EW, SR

Document Quality Control

Version QA'd	Date	Recommendations (and if appropriate reason for change)	Initials of Quality Assurance Person
0.1	12/03/2008	Addition of linkages between objectives (i.e. material in the Appendix). Other small corrections.	SS
0.2	17/03.08	QA and corrections	EW
0.3	23/03/2008	Final QA	SR

Document Change Commentator or Author and Reviewer List

Author Initials	Name of Author/Reviewer/Responsible QA Person	Institution
CSR	Colin Rosenthal	Statsbiblioteket
ABR	Asger Blekinge-Rasmussen	Statsbiblioteket
JH	Jan Hutař	Národní knihovna Ceske republiky
AM	Andrew McHugh	HATII
SS	Stephan Strodl	Vienna University of Technology
EW	Emily Witham	HATII
SR	Seamus Ross	HATII

Approved By with DPE (signature)	**Date**

Accepted by at European Commission (signature)	**Date**

1. EXECUTIVE SUMMARY AND INTRODUCTION TO PLATTER

The purpose of this document is to present a tool, the Planning Tool for Trusted Electronic Repositories (PLATTER) which provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime in a manner which will contribute to the repository establishing trusted status amongst its stakeholders. PLATTER is not in itself an audit or certification tool but is rather designed to complement existing audit and certification tools by providing a framework which will allow new repositories to incorporate the goal of achieving trust into their planning from an early stage. A repository planned using PLATTER will find itself in a strong position when it subsequently comes to apply one of the existing auditing tools to confirm the adequacy of its procedures for maintaining the long term usability of and access to its material.

In order to maintain the scope of the document at a reasonable level, we focus only on the process by which the repository organization sets and manages its *objectives*. The management of the process of *implementing* these objectives, encompassing such widely disparate areas as finance, human resource management, software and hardware planning, data warehousing etc. is too large a subject area to be covered by any single document and will typically require input from a range of subject experts.

Even the process of defining a generic tool for managing objectives and targets must deal with the considerable diversity amongst those organisations which may be included under the term “digital repository”. In PLATTER, this diversity is acknowledged and explicitly handled by requiring repositories as a first step in the planning process to answer a questionnaire which characterises the repository relative to other repositories and which can be used to determine how and whether the goals and objectives we have identified are to be realised in a given organisation.

The PLATTER process is centred around a group of Strategic Objective Plans (SOPs) through which a repository specifies its current objectives, targets, or key performance indicators in those areas which have been identified as central to the process of establishing trust. In the future, PLATTER can and should be used as the basis for an electronic tool in which repositories will be able to compare their targets with those adopted by other similar (suitably anonymised) repositories. The intention is that the SOPs should be living documents which evolve with the repository, and PLATTER therefore defines a planning cycle through which the SOPs can develop symbiotically with the repository organisation.

Table of Contents

2. THE TRUSTED REPOSITORY

The term "Digital Repository" is used to describe many distinct and overlapping types of system and organisation. Some usages restrict the term to digital collections implementing a particular model (such as OAIS[1]) or protocol (such as OAI-PMH[2]). In other contexts, "repository" is used very broadly to refer to any organisation with responsibility for managing digital material for a designated community of end-users. The term is also sometimes used somewhat abstractly to refer to a collection of services involved with the acquisition, management, and dissemination of digital material. This latter usage is of particular relevance when the services are managed by multiple institutions working in a federated structure.

The concept of "Trust" (generally used interchangeably, if ungrammatically, with "Trustedness" and "Trustworthiness") has a somewhat narrower, and therefore clearer, definition. A repository is Trusted if it can demonstrate its capacity to fulfil its specified functions, and if those specified functions satisfy an agreed set of minimal criteria which all Trusted Repositories are assumed to require. The requirement that compliance be *demonstrable* is critical, with the result that the acquisition of Trust is assumed to be largely synonymous with processes of audit and certification.

To this end, several initiatives have developed tools to enable repositories to be audited or self-assessed. These have been characterised by two complementary approaches. The TRAC[3] and nestor[4] groups have produced checklists of specific criteria which repositories are required to be able to fulfil and document in order to obtain certification. By contrast, the DRAMBORA[5] toolkit guides repositories through a risk-assessment exercise which enables them to evaluate (self-assess) their ability to fulfil their self-specified goals. Each method has its strengths and weaknesses. The checklist approach is more concrete and specific and is therefore well-suited to a certification process. On the other hand it is somewhat rigid and may be difficult to apply across the board to all possible digital repositories which might seek trusted status. By contrast, the DRAMBORA toolkit is extremely flexible because it assesses a repository relative to the repository's own self-defined goals, not to some externally defined standard. However this implies that the trustworthiness of a DRAMBORA-assessed repository can only be as good as the fitness of those self-defined goals.

A suitable compromise would be to allow repositories to identify their own goals within a broadly accepted framework of basic requirements relevant to all trusted repositories. Precisely such a framework is represented by the Ten Core Principles of Trust Repository Design which have been developed by The Center for Research Libraries (CRL), The Digital Curation Centre (DCC),

---

1 "Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1, Blue Book, January 2002"

http://public.ccsds.org/publications/archive/650x0b1.pdf

2 The Open Archives Initiative Protocol for Metadata Harvesting

http://www.openarchives.org/OAI/openarchivesprotocol.html

3 "Trustworthy Repositories Audit & Certification (TRAC) : Criteria and Checklist"

http://www.crl.edu/PDF/trac.pdf
http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

4 The nestor catalog of criteria for trusted digital repositories

http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

5 A McHugh, S Ross, R Ruusaleep & H Hofman, The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)', http://www.repositoryaudit.eu. 2007. ISBN: 978-1-906242-00-8

DigitalPresevationEurope (DPE), and The German Network of Expertise in Digital long-term preservation (nestor) in the field of audit and certification at a meeting hosted by CRL in Chicago in January 2007. The principles state that a repository:

1. Commits to continuing maintenance of digital objects for identified community/communities.
2. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfil its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfils responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as the relevant production, access support, and usage process contexts before preservation.
8. Fulfils requisite dissemination requirements.
9. Has a strategic program for preservation planning and action.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects. [6]

For the DRAMBORA toolkit, these principles determine a classification for the risks identified by the assessment process. For the checklists they represent an agreed classification scheme for the points to be checked. What remains open, and what this document is designed to address, is how these principles can be incorporated into the design and planning of a repository so that it is "trust-ready" from the start.

[6] http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

3. REPOSITORY CLASSIFICATION

One of the major stumbling blocks along the way to the development of a greater level of trust among repositories is the enormous diversity in the types of organisation covered by the very broad term "repository". In practice, experience with repository audits with DRAMBORA and TRAC demonstrates that interest in obtaining Trusted status is common to repositories of many types – for example national libraries and archives, institutional repositories, subject-based repositories and scientific data archives all recognise the value of an audit both as a tool to enable them to identify their weak and strong areas and as a source for external validation of their work.

Since no "one-size-fits-all" approach can hope to apply to all types of repository, it is vital for a repository planner to be able to classify their repository in order to be able to compare its policies and practices with other similar repositories. The first stage of the PLATTER analysis is a taxonomic classification which will enable a repository to be compared with other similar repositories. Many possible schemes for such a classification could be developed, and in PLATTER we have chosen to characterise a repository along a number of independent axes grouped into four major descriptive classes:

- Purpose and Function
- Scale
- Operation
- Implementation.

The classification axes have been chosen to be highly general in order to apply as widely as possible. It is nevertheless not improbable that some repositories in practice will find one or more of the axes to be overly restrictive or meaningless as a description of their operations. However it is expected that the taxonomy presented here will, when taken as a whole, provide a usable classification scheme for the vast majority of those repositories which are expected to be seeking trusted status now or in the near future.

3.1. REPOSITORY PURPOSE AND FUNCTION

The purpose of this group of taxonomic axes is to determine the general functional type of the repository. The requirements of a national library may be quite different from those of an institutional or subject-based repository, a scientific data repository, or a national archive.

Q1.1 Source of Mandate

Repositories receive their mandates from various sources. National libraries and archives generally receive theirs from the government or the relevant ministry. Many repositories are part of a parent institution and receive a mandate from that institution as well as sharing the overall mandate of the parent body. Many repositories, especially subject-based repositories or smaller collections define their own mandate.

What is the source of the repository's mandate?	For example: government, parent organisation, self-defined

Q1.2 Commercial Status

The functioning of a repository is strongly constrained by its business status, and specifically whether it is has a responsibility to further the financial aims of itself or its parent body.

Is the Repository for profit or non-profit?	

Q1.3 Legal acquisition rights

A vital planning area for any repository is concerned with the acquisition of material. The issues raised vary radically depending on whether the repository has the legal right to acquire the relevant material or whether it needs to negotiate access with the relevant rights-holders. Typically, state and national archives acquire at least some material by legally mandated archival deposit, while national libraries obtain material deposited under legal deposit law. Other institutions usually have to make specific arrangements for voluntary or contractual acquisition of material.

Does the Repository receive a significant proportion of its material from a legally mandated source (e.g. archival deposit or legal deposit).	

Q1.4 Operational Maturity

The process of planning for trust varies substantially depending on whether one is planning a new repository or re-engineering an already fully functional repository. For a new repository, one has the relative freedom of designing the necessary policies and processes to support trust from scratch. However designing all the processes which support effective operation of the repository while supporting the goal of trust is a substantial challenge. With a mature repository one may hope to have in place already an efficient functioning system, so that the process of obtaining trusted status will require only relatively minor organisational changes. However even these may run up against organisational inertia, and there is the additional danger that one may uncover flaws in the repository procedures which require substantial and expensive organisational restructuring.

Operational maturity is not always well-modelled as a single parameter. It may be best to consider individual services implemented by a given repository and rate the operational maturity for each of these separately. As a general rule, the PLATTER toolkit is intended for use by new and non-mature repositories. Repositories which have been operational for some time and which wish to move towards trusted status will generally find that a risk-analysis based on the DRAMBORA toolkit is a better way forward than the PLATTER toolkit.

What is the operational status of the repository (not yet running, running but still under development, mature)	

3.2. SCALE OF REPOSITORY

In this group we consider the various factors which together define the overall scale of the repository, whether expressed in human, technical, or financial terms. The universality of the Ten Core Principles indicates that both small and large repositories have to address many of the same design issues but the solutions adopted are likely to be very scale-dependent. Repositories will generally find it easier to compare their organisational and technical structure to that of similarly sized repositories.

Q2.1 & Q2.2 Data Quantity

Two simple, perhaps somewhat crude, metrics of repository size are the quantity of digital material stored and the number of files or distinct objects. The quantity of data has a significant effect on the technical requirements and may affect the choice of architecture for the IT infrastructure. The

significance of the number of objects to be archived may not be so obvious, but in practice can be very important because the processes involved in managing data have an overhead which scales, or can scale, with the number of objects, not the total size. For example, the amount of metadata will typically scale with the number of objects, and this will have a knock-on effect on the specifications of any search system to be employed. The storage architecture may also need to be modified if the number of objects is very large, for example by adopting an aggregate file structure format such as arc[7]. More generally, repositories may wish to try to estimate their growth rate over the foreseeable future, including error bars to indicate the uncertainty in these estimates.

What is the amount of digital material you expect to archive per year (GB)?	
How many distinct digital objects do you expect to archive per year ?	

Q2.3 & Q2.4 Human Size

Another metric of repository size is the number of people involved in the repository either as staff or as end-users. Neither of these figures is necessarily easy to estimate. The staff size should be the actual fulltime-person equivalents working solely on the repository. This excludes any staff employed by sub-contractors such as data-centres. This is somewhat arbitrary as it makes it difficult to compare repositories which use large-scale outsourcing to those which work with in-house solutions. Nevertheless we recommend counting only in-house employees because a) estimating the number of people working in outsourced areas is difficult and b) repositories with and without substantial outsourcing will be identified (see section 3.4) as sufficiently different that they will in any case not be directly comparable in many other areas.

Measuring the size of the user community is also difficult, especially for new repositories. However given the primacy of the principle that the repository maintains its material for the benefit of an identified community, it follows that some form of market research is an essential part of trusted repository planning and this should lead to an estimate, however crude, of the size of the user community. This will have a very significant effect on decisions relating to the technical architecture.

How many fulltime-equivalent staff does the Repository expect to employ?	
How many distinct endusers are expected to access material in the Repository over the course of a calendar year?	

3.3. OPERATION

This group of axes is primarily concerned with how material enters into the repository, the kind of material stored, and the extent to which that material may be accessed by endusers. These parameters vary extremely widely amongst repositories, nor is there any *a priori* reason to suppose that they are correlated with each other. They are therefore grouped together here under “Operation” only because they are conceptually related.

Q3.1 Acquisition Method

This question is concerned with how material enters into the Repository. Classically, material can arrive from external sources in two ways – either the repository goes out and collects material, or

[7] http://www.archive.org/web/researcher/ArcFileFormat.php

material is delivered to the repository. These two forms of acquisition can be designated "pull" (or "harvested") and "push". A third way in which a repository may obtain digital material is by generating it in-house by digitisation. It will be clear that the acquisition method has profound consequences for the entire ingest workflow of the repository and all the essential associated functions such as data validation, generation of metadata, and quality assurance.

Which of the three acquisition strategies (push, pull, self-creation) account for a significant portion of the total material in the Repository?	

Q3.2 Data Complexity

Digital material comes in many forms. The challenges of long term preservation are clearly much greater for some types of material than others. Characterising data complexity is not easy and can be counterintuitive. For example, most video formats are well-described by self-contained specifications which are designed to be simple enough to allow the video to be played back in realtime. By contrast, some text-document formats (such as Word) are container formats which can contain complex information such as spreadsheets or embedded databases. Therefore it is not enough for a repository to define the formats it will accept. It must also know what level of complexity of content it can expect, or will allow, in its archive.

For simplicity, we have identified three degrees of complexity:

- Simple data: e.g. simple text formats, images, video
- Moderately complex data: e.g. composite material with multiple linkage
- Highly complex data: e.g. software, text with embedded spreadsheets

Is the majority of the material in the Repository simple, moderately complex, or highly complex?	

Q3.3 Data specialisation

Data specialisation relates to the degree of expert knowledge required to make use of and interpret the material in the repository. Evaluation of the level of data specialisation is essential to making successful decisions about preservation actions. Specialisation is distinct from complexity – for example personal photographs and medical images may have the same degree of data complexity if the software and hardware required to access them are the same. However the specialisation of the medical images is much higher because it requires highly specialised knowledge to determine, for example, what metadata needs to be recorded with a given image and which properties of a given image are significant in the sense that they must be preserved under any preservation action.

How specialised is the data in the Repository (low, medium or high)	

Q3.4 Data Sensitivity

Data sensitivity refers to the degree of significance of ethical and legal considerations regarding the acquisition, storage and dissemination of the material in the repository. Examples of material with high sensitivity would be material of high commercial value or medical data containing personal information about the research subjects. The critical question relates to the *most sensitive* material in the repository since it is this which will determine the level of security which the repository will need to implement.

How sensitive is the most sensitive material in the Repository (low, medium, high)	

Q3.5 Access Rights

Access to material in a repository can be open for all, open under restriction (for example to researchers only) or closed. Many repositories have material in more than one of these classes.

In which of the three access classes (open, restricted, closed) does the Repository have significant holdings?	

3.4. TECHNICAL SOLUTIONS AND IMPLEMENTATION CHOICES

This group of axes deals with the choices made in the implementation of the repository system.

Q4.1 Source of Metadata

The provision of adequate bibliographic and descriptive metadata is basic to repository functioning. Such metadata can come from various sources:

- Provided separately by the depositor (including, for example, http headers on a web page)
- Extracted "by hand" from the supplied data
- Extracted automatically from the supplied data
- Obtained from a third-party depositor

What are the main sources of bibliographic and descriptive metadata in the repository?	

Q4.2 Interoperability Standards

A crucial development in repository technology is the development of interoperability standards. These allow repositories to pool their resources at a technical level to share services, material, and metadata. Examples of the possible applications of interoperability are format identification and validation services, services for discovery (search) and access, and automatic replication services for preservation. Very few interoperability standards have become widely accepted but it is likely that the near future will see the emergence of a clearer picture.

What interoperability standards are implemented in the Repository?	

Q4.3 Storage Strategy

This question is concerned not with the details of storage architecture but with the basic strategy used, essentially whether the repository takes responsibility for running its own storage or uses an external storage provider. A third option is in-house storage but under an external maintenance and support contract.

What strategy is used for storage? (in-house, external, in-house under external support)	

Q4.4 Software Strategy

By software strategy we refer to the mode in which the repository obtains and manages the software used by the repository. This is not a question of Open Source versus proprietary software, but rather of the strategy used by the repository to ensure that its software is adequately supported. The most common strategies are

i) support by the software supplier

ii) support by a third party

iii) self-support (i.e. in-house) and

iv) support by a user and developer community

What strategy is used for software management?	

4. THE PLATTER PLANNING CYCLE

The PLATTER planning cycle describes a semi-formalised set of steps intended to facilitate the processes of definition and expression of organisational objectives, and implementation and evaluation of the measures intended to meet them.

The process is a cyclical one, and individual sections conform in many respects to parts of the DRAMBORA risk-analysis process. The following sections seek to describe in some more detail each stage of the PLATTER cycle, outlining their implicit parts and how they interrelate.

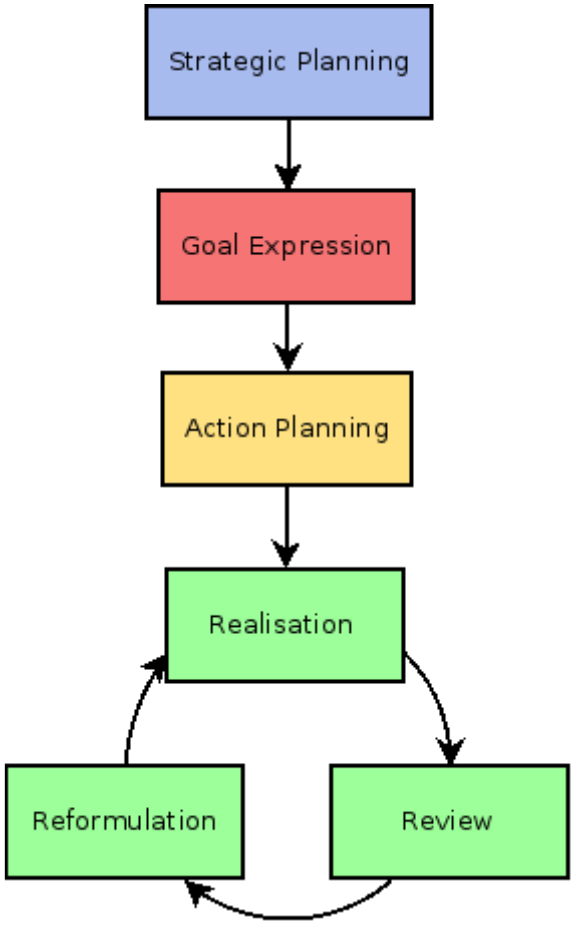

Illustration 1: The PLATTER Planning Cycle

4.1. STRATEGIC PLANNING

Strategic planning is an invaluable means for maintaining a sufficiently broad and forward-facing organisational perspective, even when individuals are focusing on much more immediate and specific aspects of business activity. Good strategic planning will provide leverage or justification for subsequent business decisions, a platform upon which to construct more detailed plans, a formal expression of purpose, legitimising the organisation's business objectives and greatly assist in the process of evaluation, maturity modelling and the pursuit of improvement. Strategic planning is related to, but distinct from operational planning, which is likely to consist of many shorter term, more explicit and more measurable objectives, in contrast with the widespread and more generally expressed strategic plan. It is similarly distinct from strategic management, and decision making, although implicitly tremendously influential with respect to that activity. Despite a focus of up to around four

years, and being subject to ongoing revision, strategic planning should nevertheless be an up front activity that steers and influences all other aspects of organisational development. Traditionally, one or more of three fundamental questions are posed during the process of strategic planning; an organisation's responses will influence to a considerable extent all subsequent decision-making:

1. What do we do?
2. Who do we do it for?
3. How can we excel?

Responses to these three questions organisations will encapsulate the repository's mandate (or reference a non-self imposed, e.g., legislative mandate), detail the identities and broad expectations of primary stakeholders and describe in general, but tangible terms, the circumstances and performance levels that will represent success.

A situation->target->path approach is commonly adopted to complete the strategic planning activities. Evaluation of the current situation, followed by an expression of goals and objectives, and finally the determination of possible means of their achievement provides a robust and evolutionary means for developmental, preproduction and production phase strategic planning. Variations on this approach exist, and these can be usefully combined, galvanising plans. A popular alternative approach requires organisations to consider an ideal organisational landscape, contrast it with contemporary business and environmental realities, and finally, through a process of comparative gap analysis, define objectives and plan for resource deployment that will improve the organisation, equipping it to deal with both internally and externally arising pressures.

4.2. DEFINITION OF GOAL OR PRINCIPLE – OPERATIONAL PLANNING

For any repository, indeed any formal organisation, the first expression of purpose is as a broadly stated definition of fundamental goals or objectives. These objectives, structured according to related business areas are implicitly flavoured by the strategic planning undertaken by the organisation. The most usefully expressed objectives are those that conform to the SMART requirements; they must be specific, measurable, assignable, realistic and time-related. Only SMART objectives can form the basis for subsequent evaluation. Similarly, the expression of SMART objectives will greatly facilitate planning and implementation of methods for their completion. Whereas strategic planning should be used as a vehicle to express the organisation's philosophy and fundamental purpose, objective, or operational planning is the expression of the practical achievements that will represent the consequence of the achievement of these more generally expressed goals.

Defined objectives must take into account the expectations and requirements of each major stakeholder. From the perspective of digital repositories, this may include management, funders, information creators, owners and depositors, and end users interested in accessing preserved content. Each must be related, either explicitly or implicitly with more fundamental strategic objectives, most immediately with the organisation's mission statement. In more commercially oriented environments objectives are likely to encompass aspirations for growth, profit, infrastructural development and markets. Preservation repositories may be subject to alternative motivators, but these must be explicitly defined. The quantification of preservation goals is challenging, but when roles and activities become sufficiently granularised, this is increasingly realisable.

The SMART criteria for objectives are mutually compatible, and to some extent consequential of each other. Measurable objectives demand specificity and temporal parameters; for objectives to be deemed realistic or otherwise, the responsibility for them must be assigned, and the quantitative demands for completion well understood. As well as the characteristics of individual objectives, one must also consider the relationships between multiple objectives. Inevitably, most organisations will be capable of distinguishing a number of objectives. In turn, it is likely that a hierarchical representation is feasible, whereby objectives are prioritised. Similarly, objective congruence can be determined to

evaluate to what extent objectives are compatible. Goals may also be defined as short, medium, or long term, although this will be implicit as a natural consequence of defining time-bounded objectives.

The definition of objectives is most usefully undertaken within a structured approach. Objectives will be associated with multiple organisational aspects, which may be determined or distinguished in either a horizontal, or vertical fashion. A horizontal approach structures according to distinct organisational characteristics, such as staffing, funding, technology, legal personality, data, stakeholders and policy. In contrast, a vertical approach is focused much more on procedural or functional realities; within a repository this may be management, acquisition, archival storage, preservation planning, preservation action, and dissemination, essentially the kinds of units described in the OAIS standard's functional model. In practice, a combination of both approaches, as favoured in the DRAMBORA audit process, is probably optimal.

Following the structured approach, a broad selection of individuals should be engaged in the process of objective definition (and assignment). Management, technical specialists, and stakeholder representatives should each contribute to the discussions. Management techniques such as mind mapping might usefully visualise the objectives, and their implicit, and wherever feasible, quantitative meaning. To ensure their measurability (a vital prerequisite for subsequent phases of the PLATTER process), defined objectives must have an associated unit, a means of determining their achievement. From a repository perspective the units will vary markedly depending on the context of the objective. Goals that relate to budgetary sustainability may use available currency as an indicative unit; those relating to access provisions may rely on numbers of page impressions on the repository's web site; objectives related to object acquisition may be expressed in terms of number of ingested objects. Specific preservation challenges are more difficult to express in such finely expressed quantitative terms, although considering preservation in terms of significant properties, and procedural and organisational demands, one can more meaningfully define success. An example may be the migration of image materials. Objectives of the process may specify minimal thresholds for colour reproduction, resolution, or platform support, detail requirements for process scalability or time per item, or place a ceiling on financial or other resource costs. Both binary and more metered measures may be implemented, and both are legitimate, although more illustrative assessment is possible when using a more granularised means of success determination. Where multiple achievement indicators are defined, they must be prioritised, in order to facilitate subsequent planning.

If not implicit, temporal parameters must be applied to each objective *('*we require *a* of *b* by *c')* and individuals or role holders assigned responsibility for all subsequent stages of planning, implementation and evaluation.

4.3. UNDERTAKE PLANNING

The planning stage is bridge-building; between the determination of what must be achieved, and the tangible realisation of such achievements. Related experiences in comparable environments can be considered and where appropriate absorbed into the planning cycle. Planning can also be supported by experimental evaluation, situational modelling, simulation and hypotheses, although must at all times remain structured according to fundamental strategic goals. Any planned solutions must implicitly support performance measurement.

Where possible, objectives should be grouped, in order to facilitate the planning process; the relationship between objectives and plans need not be 1:1, although every objective must be accounted for in the resulting plans. It perhaps needn't be said, but under no circumstances should orphaned plans exist – there ought to be no practical efforts, either proposed or already underway, that do not correspond to one or more existing organisational objectives.

Where planning results in incompatibilities or stresses, higher prioritised objectives or achievement indicators should be favoured. This may in some cases require additional hierarchical evaluation to determine the more desirable objective(s) and measures.

Action planning is only really feasible by adopting a global perspective of organisational constraints and influences. Many will have been formalised in the initial strategic planning stages, most notably those focused on establishing current situational awareness, and perceptions of any emerging contextual influences. Legislation, policy originating from parent organisations, stakeholder expectations and resource availability will all contribute to the success or otherwise of planned actions and must be given adequate consideration.

4.4. DELIVER, REVIEW AND REFORMULATE IMPLEMENTATION

An iterative cycle, that may extend beyond the planning and development of the repository into full production phases, these three interrelated activities are fundamental to the ongoing improvement and developing maturity of the repository. An agile approach to all three will benefit the organisation and the pursuit of its objectives; no period of implementation should become too prolonged prior to the initial phases of review and reformulation.

Assuming the successful completion of initial action planning, the first implementation phase should be straightforward. As soon as sufficient infrastructural maturity is reached, and as soon as performance evaluation measures can be effected, the initial period of review should follow. Exposing processes, procedures and policies to targeted scrutiny, based on defined measures should yield immediate insights into shortcomings and opportunities for improvement. Where temporal requirements or constraints are evident, reviewers should determine the extent to which targets are met based on the results of appropriate calculations.

The period of reformulation that follows review is akin to a more concentrated, and more informed repeat of the initial planning stage. Evident shortcomings are considered as additional constraints, and solutions redesigned to more adequately operate within the increasingly well understood problem space. Subsequent phases should see diminishing scales of re-engineering, until in the first instance objectives are extended, and then latterly the current strategic planning cycle is completed and another devised.

5. THE PLATTER STRATEGIC OBJECTIVE PLANS

In PLATTER, we associate a set of Strategic Objective Plans with the Ten Core Principles. These SOPs are closely related to the Ten Core Principles, although the correspondence we have adopted is not one-to-one. Repositories are free to choose their own set of SOPs although they will necessarily cover largely the same areas as those we recommend here. We note that the 4th Core Principle, that concerned with implementing a policy framework, is subsumed into all the SOPs which, taken as a whole, represent just such a framework.

Strategic Objective Plan	***Responsibilities***	***Corresponding Core Principle(s)***
Business Plan	Financial planning, monitoring, and reporting	2
Staffing Plan	Acquisition and maintenance of relevant skillset for managing repository	2
Data Plan	Specification of data and metadata objects, formats, and structures for ingest, storage, and dissemination, together with the relevant transformations and mappings.	5,6,7,8
Acquisition Plan	Management of the relationship with depositors and other data providers. Appraisal policy.	3,5
Access Plan	Management of relationship with end users. Access Policy.	1,8
Preservation Plan	Ensure that access and usability of material in repository is not adversely affected by technological change and obsolescence	9
Technical System Plan	Specifies goals for hardware, software and networking	10
Succession Plan	Manage obligation to ensure preservation of material beyond the lifetime of the repository	1
Disaster Plan	Respond to rapid changes to the repository environment	1,6

We now turn to the individual Strategic Objective Plans and the process by which the Ten Core Principles are concretised into a set of SMART objectives. The discussion of each SOP consists of a general introduction to the area covered by the SOP, followed by a series of Goals which the SOP is required to address. These goals are intended to be generic, which is to say that they represent a strategic level of planning which should be common to almost all repositories. For each goal we then list a number of specific examples, that is to say realisations of the generic goals which may be appropriate for a particular repository. There then follows a discussion in which we list the points which the realisation of the generic goal should address.

It is central to the PLATTER philosophy that none of the generic goals or issues raised in the discussion of them are ignorable. That is not to say that every point raised is relevant to every repository, but rather that every repository should seek to address every point raised. Where a repository chooses not to set any targets for a given issue, it should explicitly justify that decision. In some cases this will simply be because the specific nature of a given repository renders a particular issue irrelevant. In other cases, however, it may be that the cost or complexity of addressing the issue outweighs the benefits to be gained or the resources available. It is important that these two quite distinct cases are recognised and that the situation for the repository in hand is correctly identified, so that the decision to ignore a particular issue can be intelligently reviewed as part of the PLATTER planning cycle.

5.1. BUSINESS PLAN

Economic constraints have been identified as among the primary threats to long term preservation of cultural assets – digital or otherwise. The issue is not just the possibility of ultimate closure of a repository due to financial stringency, but also the ongoing threat that insufficient or inefficiently deployed funding results in a failure to secure material against threats to its usability. For these reasons, a digital repository must implement sound business practices.

It is difficult to discuss financial planning for a generic repository because the status and funding arrangements for repositories vary enormously. It is unrealistic to require that a repository's status as trusted should be dependent on its being able to demonstrate guaranteed long-term funding, because few, if any, repositories can realistically be certain of the stability of their funding streams for more than a few years. Instead the focus of the Business Plan SOP is on sound financial planning and monitoring and the provision of contingency plans for dealing with financial stringencies. (The issue of what actions to take in the event of complete closure of the repository due to loss of funding or insolvency belong more properly in the Succession Plan.)

Goal 1.1: Monitor and review business plan on a regular basis

Examples:

- Financial statement and budget to be produced by date ddmmyy each year and reviewed and approved by repository steering committee

Discussion:

In order for the business plan to remain current, regular review is needed. As sources of income cannot normally be guaranteed for more than a few years, monitoring is needed to identify such shortages before they develop into a budget deficiency.

Goal 1.2: Maintain financial support at a level suitable for routine functioning of repository

Examples:

- Obtain funding at 100% of estimated budget necessary for adequate operation

Discussion:

There are inherent costs in running a digital repository, which cannot be avoided. Budget shortages will lead to commitments being broken which is detrimental to the trustworthiness of the repository. As such, the repository must achieve an income sufficient for routine functioning. If not possible, commitments will have to be adjusted (rather than broken), to what can be achieved with the available budget.

Goal 1.3: Ensure contingency plans for financial cutbacks or emergencies are adequate to protect vital data

Examples:

- Have an agreement with another repository about housing vital data in case of a foreseeable financial problem
- Prioritise services to be retained in the case of financial stringency

Discussion:

This is one of the most sensitive areas for repository planners. Many are understandably reluctant to develop contingency plans for severe financial cutbacks in case these are treated by their funding agencies as an admission that the current level of funding is unnecessarily generous. Nevertheless, trust requires that every repository must have plans to protect the most vital data in the eventuality of substantial funding loss. This protection might take many forms, and will be detailed in the Succession Plan and the Disaster Plan.

Goal 1.4: Define and maintain marketing and outreach plans suitable for the repository's needs

Examples:

- Create an outreach plan by date ddmmyy and review it every two years

Discussion:

Outreach and marketing address several key areas of repository viability - specifically communication with depositors, users, funders and external collaborators. Targeting of the repository's outreach activities will depend on the relative priority it gives to these areas, which will depend on the perceived adequacy of the repository's current profile. For example, a repository with a well-defined and satisfactory source of new material, or one which relies on harvesting of publicly available sources (e.g. a webarchive), may not need to seek new depositors. On the other hand an institutional repository which relies on researchers to voluntarily upload material will require an effective strategy for communicating to the researchers the benefits they will receive from taking the time to deposit material. Repositories must also be aware of the synergistic effects between different areas of outreach. For example, an increased user base and strong international contacts may be used as arguments to obtain extra resources from funding agencies.

5.2. ACQUISITION PLAN

There are three primary components to an Acquisition Plan: specification of the desired material, negotiation of the relevant agreements to obtain the material, and development of procedures for acquiring the material.

Goal 2.1: Acquire relevant material

Examples:

- Archive 90% of national internet
- Archive 75% of all articles published in house
- Ingest at least 10000 new images per year

Discussion:

The purpose of this goal is to specify quantitative targets for the material to be acquired by the repository. Specification of the desired material will generally refer back directly to the repository

mandate which specifies the type of repository involved – for example an archive, eprint collection, webarchive etc. Further specifying the exact extent/scope of the material to be collected may require extensive analysis – for example, a market analysis to determine what is available, a stakeholder analysis to determine the wishes of depositors and end users, and a cost-benefit analysis to determine economic constraints. In other cases, specification of material to be collected may be simple or trivial – for example all papers published by the parent institution. But even here, setting a numerical performance target may require careful analysis or a pilot project to obtain realistic numbers.

Goal 2.2: Negotiate deposit agreements

Examples:

- Negotiate deposit agreement with company X guaranteeing access to all their eprints for at least five years
- Negotiate upload arrangements for digital archival material

Discussion:

Obtaining access to material depends on the legal status of the relationship between the repository and the depositor. In some cases, such as national libraries and archives, the deposit of material may be legally mandated. In other cases material will be obtained by free negotiation with the provider. There are also cases where the material is generated by the repository institution itself, for example by digitisation of its existing analogue holdings.

Where negotiation of a deposit agreement is necessary, nestor and TRAC checklists provide considerable inspiration regarding the points which such agreements should cover:

- Scope of material
- Delivery form (e.g. ftp download)
- File formats
- Accompanying metadata
- Right to take preservation action on delivered material (e.g. migration, creation of multiple copies)
- Usage/distribution rights
- Obligations on depositor to notify repository of any changes to file formats, delivery form etc.
- Transfer of legal rights, if any
- Appropriate duration of agreement

It should be noted that even in the case of material acquired by legal or archival deposit, some of the above points may still require to be negotiated or clarified by seeking legal advice. For example, a repository may need to determine whether the relevant legal deposit law allows the taking of copies and/or migration for preservation purposes. Technical issues associated with the transfer will almost always require negotiation. Even in the case where the material is generated within the repository the issue of legal rights still needs to be addressed – for example does the repository have the right to digitise holdings for preservation purposes and what restrictions are there on distribution of the digitised material.

As a general rule, the issues to be dealt with in the acquisition of descriptive and bibliographic metadata from external sources are identical to those for primary material[8]. As an example, a television

[8] This is trivially true since “one person's metadata is another person's data”.

archive obtained descriptive and technical metadata about broadcast programs from a market-research specialist whose main business was selling viewer figures and demographics to potential advertisers. The major items to be negotiated between the metadata provider and the repository included all the elements discussed above.

Goal 2.3: Obtain Physical Control of Materials

Examples:

- Purchase and install harvesting software
- Develop workflow for monthly ftp download of new material

Discussion:

This goal refers only to the process needed to place the actual data material under the control of the repository. The remaining tasks involved in ingestion of the material into the repository are covered in the Data Plan.

Goal 2.4: Monitor acquisition

Examples:

- Requested material to be physically downloadable from company X's website.
- TV broadcasts to be recorded with at least 95% time coverage

Discussion:

It is essential that a repository have in place a monitoring system to determine that the required material is actually being made available by the producer or depositors. The details of this system will vary widely depending on the repository type. An institutional research repository, for example, might employ procedures to compare published works in journals and other sources with the list of deposited materials, while an internet archive could use statistical analysis of its harvested material to estimate its percentage coverage. A national archive might audit its depositors to ensure compliance with their legal archive deposit obligations.

Goal 2.5: Maintain Relevance of Deposit Agreements

Examples:

- Annual review of deposit agreements

Discussion:

Deposit agreements must continue to be relevant to the overriding goals of the repository. The repository should institute procedures to monitor the relevance of any deposit agreement, taking into account the same issues considered in the initial development of the agreement. These procedures should also be concerned with the possible need to negotiate further deposit agreements with new providers.

5.3. STAFFING PLAN

The relatively rapid growth of the repository sector has resulted in a difficulty in hiring appropriately qualified staff. The problem is not simply that suitably qualified people are unavailable but that the rapid development of repository infrastructures has created uncertainty as to exactly what qualifications and experience are necessary or relevant to repository work. There are not yet any qualifications which are specifically related to repository work so that repository staff tend to be drawn from many different backgrounds – archivists, librarians, it-specialists, administrators etc. A related problem is that the relative novelty of repository work results in a lack of clear career-development

paths, which hinders the retention of experienced staff. It is vital that repository organizations participate in efforts to institutionalise repository work by creating nationally and internationally accepted accreditations and career structures. However, at the present time, when such standards do not yet exist, individual repositories must themselves take responsibility for defining roles and career paths which will enable them to hire and retain talented and competent staff.

Goal 3.1: Delineate roles, responsibilities and authorizations of repository staff

Examples:

- Produce detailed breakdown of staff roles (by date ddmmyy)
- Review staffing roles biennially

Discussion:

Running a repository is a very cross disciplinary endeavour. Delineating responsibilities will ideally allow staff with vastly different training to concentrate on their field of expertise, and thus increase the quality of the work.

Having clearly defined responsibilities also makes for a simpler organizational structure. Having just one role, that handles server maintenance for example means everybody knows who to talk to about that.

Lastly, having defined responsibilities and budgets to go with these responsibilities will lead to more empowered staff, which in turn can help to build repository career paths. In addition, the repository's overall leaders will be able to focus more on the general strategy of the repository, rather than approving expenses for minor tasks.

Goal 3.2: Acquire and maintain adequate staffing to fulfill specified roles

Examples:

- Repository should hire and retain a Manager and Administrator by date ddmmyy

Discussion:

Once the repository has defined the roles and responsibilities of its organization, it must ensure that these roles are filled with qualified personnel. Lack of staffing will, over time, lead to commitments being broken, which will in turn adversely affect the trustworthiness of the repository.

In many cases, acquiring staff will also include negotiating the use of currently employed human resources within the parent organisation.

Goal 3.3: Maintain staff skills

Examples:

- All staff to participate in an annual personal development review to determine and review individual training goals
- Funding should prioritise sending all staff to at least one relevant international workshop or conference per year

Discussion:

Maintaining an up-to-date staff skill set is paramount. The rapid growth of the repository sector has the effect of quickly making staff skills redundant. In order for the staff to best fulfil their functions, the repository should host or pay for further training.

Best practices for repositories and standard ways of handling threats, along with ways of disseminating the data for the users, are still being developed. In order for the repository to stay current with what other repositories are doing, participation in workshops, user groups and conferences should be prioritized.

Repositories should be wary of too much compartmentalization. Having indispensable experts will place the repository at risk, if these people leave the repository. The sharing of knowledge among staff should be encouraged to lessen this impact.

5.4. ACCESS PLAN

Dissemination of the contents of the repository is one of the most important and most visible outcomes for the vast majority of repositories. We may consider dissemination and access, beside archival function, to be one of the main goals for which repositories are actually established (universities, e-prints archives etc.) in general and that this scenario, with some public access, will be the case of for the majority of the audience of this document.

Of course there are some repositories called “dark archives”, which do not offer any access and a primary function only to store data safely. Such repositories need not consider all issues about dissemination and access in depth. However, even these repositories have a designated community. In the case of ‘dark archives’, the “designated community” of end users may, for example, be hypothetical generations of future historians. Even so, it is still essential for the repository planners to consider at least the minimum needs of such future users such as the descriptive metadata which will allow them to navigate the archived material and understand its context.

The primary questions we need to ask when thinking about dissemination are; to whom do we want or need to offer data from our repository, what are the conditions and restrictions under which we will allow access and what technical background, equipment and pre-arrangement do we need to have and undertake.

Goal 4.1: Create, Maintain and Review a Mission Statement which reflects the Repository's Mandate

As the primary function of any repository must be to preserve material for future use, we believe that discussion of the repository Mission Statement also properly belongs under discussion of its dissemination planning.

Examples:

- Have mission statement approved by repository steering committee by date ddmmyy
- Review compliance of repository to its mission statement every two years
- Review relevance of mission statement every five years

Discussion:

The mission statement of a repository will generally be closely derived from its mandate. The distinction between the two is that, in most cases, the mandate will be generated externally by the body responsible for creating the repository – for example a government ministry or a parent organisation. The mission statement is defined by the repository itself, and reflects its commitment to preserving a body of data, material or knowledge for the benefit of a particular “Designated Community” of end users.

A repositories mission statement is very important as it helps to fulfil their mandate and clarifies its relationships to external agencies. A mission statement should be a “living” document and be committed to formal, periodic review and assessment to ensure continued development.

Some areas to be considered in a Mission Statement:

- It should define general goals of the repository
- It should reflect a commitment to the long-term retention of, management of, and access to digital information on behalf of a designated community
- The mission statement should refer to the repository's goal of establishing Trust with its stakeholders through a process of audit and certification.

Goal 4.2: Develop and maintain a definition and understanding of your Designated Community/ies

Examples:

- Define designated community of end users by date ddmmyy
- Review the needs of the designated community every second year
- Set up a contact group with the designated community to meet every six months

Discussion

The definition of the designated community should be available on the repository website and the community should be very well aware of available delivery and access options. It is essential to monitor and reflect all changes inside the community over the time. To avoid failure to meet their needs.

Some of the issues to be addressed in developing definitions of a Designated Community are:

- Who is the target community?
- How large is the community and how is it expected to grow?
- How diverse is the community (age, professions, different backgrounds etc.)?
- What is the knowledge base of the community?
- What level of service do they expect?

It is important to have procedures in place for monitoring or receiving notifications about changes in the needs of the Designated Communities. The Access Plan should specify goals for monitoring changes through, for example, surveys, formal reviews, workshops, individual interactions etc.

Goal 4.3: Create and Implement A Repository Access Policy

Examples:

- All material to be freely available (Open Access)
- All material available to registered users
- Material available to all users on-site but only to *bona fide* researchers off site
- Review compliance of repository with Access Policy every year
- Review Access Policy every second year
- Develop, by date ddmmyy, procedure for dealing with suggestions or complaints regarding Access Policy

Discussion:

Access policies can vary enormously. Most repositories are interested in disseminating their holdings as widely as possible. The subtlety lies in determining the limits of the possible. Some sources of restriction are

- Copyright law
- Law on dissemination of personal or commercially-sensitive data
- National security
- Libel, obscenity, hate-speech and blasphemy laws
- Specific contractual restrictions imposed in deposit agreements

Moreover some of these restrictions may affect different classes of users in different degrees. For example some material can be made available to researchers but not to the general public. Access policies should also determine the level of access allowed to repository and support staff. In determining an access policy it is essential to be able to specify the metadata necessary to support the policy – for example the metadata schema may need to specify that certain material can only legally be distributed to adults.

Access policies will also specify how the repository deals with issues of authorization, authentication and access-logging. The implementation of an access control system which implements the access policy is a matter for the Technical Plan. However the Access Plan should also set goals for reviewing the implementation for compliance with the access goals.

Goal 4.4: Specify and fulfill technical requirements for dissemination and access

Examples:

- Establish and implement minimum metadata requirements to enable the Designated Community to discover and identify material of interest (by date ddmmyy)
- Implement search tool capable of finding and retrieving documents by title or author (by date ddmmyy)
- All material to be available for access within three days of acquisition

Discussion:

Whereas Goal 4.3 was concerned with *restrictions* on access to material, we are here concerned with specifying goals in *enabling* access. Of critical importance here is the descriptive and bibliographic metadata to allow the users to find and obtain the requested material. The repository should specify goals in search and discovery as well as for electronic distribution of the actual digital objects, taking into account the expected request-load on its systems. (The technical form of the material disseminated, or the 'Dissemination Information Package', is the responsibility of the Data Plan.) Other areas to be considered are logging of access and the use of DRM (Data Rights Management) in distributed material.

5.5. TECHNICAL PLAN

When preserving digital data, the IT infrastructure is of course of prime importance. The issues threatening preservation are manifold. The technical plan is meant to deal with the threats to the data, and the systems on which they reside, as well as the systems that provide services for endusers. Even though protecting against all classes of problems is unachievable, it is better that the technical plan mentions as many as possible, rather than ignoring the ones where the repository does not have a well thought out strategy. It is common to approach technical requirements and goals by focussing on the

distinct areas of software, hardware, and networking requirements. However for the purpose of objective-management it is better to consider major constraints on repository operation across these categories. In PLATTER, we identify the three major areas to be taken into account in setting goals for the technical infrastructure as scale, security, and services.

Goal 5.1: IT Infrastructure must be capable of coping with the scale of data storage, processing and transport appropriate for the repository

Examples:

- Repository must have external network connection with minimum speed xMbps
- Internal network speeds must be at Gigabit speed
- Repository must have sufficient processor power to transcode xGb of video data to mpeg file format per day
- Repository must use scalable server-farm solution for external services
- Ability of systems to cope with scale of repository operation to be reviewed annually

Discussion:

Overall, the IT infrastructure must be able to maintain the level of service detailed by its commitments, taking into account both changes to the repository itself and changes in the external environment, such as increased web traffic or new security threats. The exact layout of the IT infrastructure is heavily dependent on the commitments, but for all repositories these points must be addressed:

- *Has hardware, software and networking systems appropriate to the load it receives*

  The repository must establish an IT infrastructure that is suited to the requests and load it will receive. Redundancy, in terms of other systems sharing the load, should be used to improve uptime.

- *Functions on well-supported operating systems and other core infrastructural software*

  The issue here is not one of Open Source vs. Closed Source software. Rather what is essential is adequate support, which can come in the form of a formal support agreement, contact with user and developer groups, or in-house development and maintenance. Monitoring the adequacy of the level of support is essential.

  Using the same systems as other repositories, and participating in user groups, as well as ensuring a high level of staff training, can greatly increase the ease and decrease the cost of support.

- *Identifying which software or hardware systems are no longer adequately supported, and must be migrated.*

  The repository must have procedures for monitoring obsolescence of hardware and software, including withdrawal of adequate support.

- *Maintaining a record of all changes to its IT infrastructure*

  Keeping a record of all changes to the IT infrastructure can go a long way to establish trustworthiness. Inadequate documentation coupled with high staff turnover represents a serious threat to security and data integrity. Documentation of all changes to the systems will also allow for quicker changes in staffing, as new employees can refer to the log, rather than the memory of other staff.

Goal 5.2: IT infrastructure must be able to guarantee the integrity and security of the stored data

Examples:

- Data are stored on a RAID system, to ease recovery
- Weekly tape backups of all data are stored offsite
- Checksums of all data are generated every 3 months, and compared with previous values, to detect changes
- Communication with the servers goes through a restrictive firewall, and the physical access to the servers are restricted to select personnel
- Repository to undertake security audit every X years

Discussion:

Security and integrity, in all their aspects, are essential to the proper functioning of a digital repository. Data must not be altered without authorization and data fixity must be demonstrable. The repository must document the threats against which it tries to protect, and the measures taken to do so.

There are four main aspects to consider:

- *Protect the digital data from being read by unauthorized users*

  Many repositories preserve data that is either commercially, political or personally sensitive. The repercussions can sometimes be major, if unapproved users got access to the data. This includes staff taking the data home, without authorization, where it can be accessed or stolen outside the safeguards of the repository. The question of access policy is discussed under the access plan. The technical implementation must be capable of supporting the repository's access policies.

- *Protect the digital data from being altered by unauthorized users*

  For the digital data to remain trustworthy every change must be documented. If users have the ability to alter data, especially without leaving a record trail, the data can no longer be relied upon. Identifying which data has been altered, even if where it is not possible to undo the alteration, goes a long way to establish trustworthiness for the repository.

- *Prevent the digital data from being destroyed by events outside the repository's control*

  Environmental disasters can easily destroy IT systems not built to withstand them. Likewise political instability, war and the ensuing looting can deprive a repository of its systems, and worse, its staff. Even something as simple as a blackout can damage sensitive systems. These threats are addressed in the Disaster Plan.

- *Prevent the digital data from being destroyed by events within the repository's control*

  Unsecured IT systems can easily be compromised and damaged by virus and trojan programs. Especially ransom-ware, software that encrypt your data and tries to ransom it back to you, could prove problematic.

  Disgruntled staff might also introduce these threats or perform the same actions. Safeguards should take this into account, for example by ensuring that no member of staff has access to both the online data and all backup copies.

Goal 5.3: The IT infrastructure must be able to guarantee that certain services remain available to the users

Examples:

- Review usage statistics for all services every 3 months
- Service uptime for search and retrieval of 99%, averaged quarterly
- Store information about users and logs in a separate secure system, not in the system providing the service
- Have services distributed between multiple virtual machines, to dynamically scale up and down the resources available, and ease the process of restoring a failed service
- Have another facility with backups of the data, that can, in case of a local disaster, continue the services
- Have agreement with Internet Service Provider about actions in case of loss of connectivity or DDoS attacks

Discussion:

The services correspond to the commitments of the repository, so failure parting this area reflects on the trustworthiness of the repository.

The services made available to the users are, for many users, the only way they will interact with the repository. If the systems are unavailable, or compromised, this will negatively affect their view of the whole repository.

To achieve this goal, these points should be considered:

- *Identifying which services are no longer required and can be closed and which are in great demand*

  As the repository grows and the user community changes, certain services might no longer be required, and should not be maintained. This could include old search systems, arcane display formats and the ability to request the data on outdated media.

  New ways to present the data to users might also be in demand.

- *Prevent information about the users being made available without proper authorization*

  The information about what data a given user has accessed or the personal information entered, when the user signed up, will often be quite sensitive. While the repository might store such information, it must make sure that it is not made available to other users, through security breaches in its services.

- *Prevent the services being disrupted by events inside the repository's control*

  Failure of single systems inside the repository should not affect the services from the users' viewpoint. General redundancy, such as hot-swappable systems, should be employed.

- *Prevent the services being disrupted by random events outside the repository's control*

  Environmental effects should not disrupt the services, from the users' viewpoint. This point covers such thing as power loss, or loss of internet connectivity, as well as more serious effects such as flooding or fire. The possible strategies against such threats include, but are not limited to, having offsite facilities, that can continue the services, or having multiple internet connections and an emergency power supply. Individual repositories will need to assess the costs and benefits of such backup services assessed against the risk of their occurrence.

Denial of Service attacks, which might take the form of hacking attacks meant to destroy, can negatively affect the uptime or trustworthiness of the services. The repository must protect its systems against such attacks.

5.6. DATA PLAN

The data plan describe the data and metadata formats used by the repository, the transformations used during ingest and access, and the strategy used by the repository to monitor the suitability of these data formats. In this section we use the common terminology provided by OAIS[9]:

SIP: Submission Information Package

AIP: Archival Information Package

DIP: Dissemination Information Package

to describe the stages of transformation which data and metadata undergo as they are cycled through a repository. The path of the data through the repository, as envisaged in this plan, can be described like this:

1. A data provider encodes his data in a package format (SIP) acceptable to the repository
2. The repository receives these SIPs, and repackages them for storage (AIPs)
3. A repository user requests the data, and the repository repackages it in a format appropriate for the user (DIP)

Data and metadata formats for each step must be defined, as well as ways of converting between them.

Goal 6.1: Specify the digital object formats the repository will accept (SIP)

Examples:

- Internet archive repository downloads and stores web pages "as-is" with http headers as additional source of metadata
- E-print archive accepts upload of pdf files with accompanying metadata provided by a web form

Discussion:

Specifying the format of the digital objects the repository will accept is crucial for digital preservation. While it would be easy to build a very generic repository, that takes care of preserving digital bits, such stored data risks quickly becoming useless. Rather, the repository needs to specify which data file formats are acceptable, and what required metadata (and the file format thereof) should accompany each digital object.

Concerning file formats, the devil is occasionally in the detail. Formats (potentially) containing encryption, or embedded objects and files, as well as commercially protected formats can sometimes prove impossible to convert. The files can be stored, but once technological progress renders them outdated, the content will be lost, and so preservation cannot be guaranteed.

The repository should have some description for content providers about how to package data and metadata or representation information in SIPs . The repository should have a policy regarding the

---

9 CCSDS 650.0-B-1: Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1. January 2002. 148 s. Available from:

http://public.ccsds.org/publications/archive/650x0b1.pdf

completeness and correctness of new SIPs and what action to take regarding invalid or incomplete SIPs. In some cases (e.g. a webarchive) the appropriate policy may simply be "take everything" whereas in other cases much stricter criteria may be relevant.

The repository can, and probably will, handle several kinds of digital objects. There should be defined SIPs and verification methods for each kind.

Goal 6.1.1: Specify sources and formats for bibliographic and descriptive metadata in the SIP

Examples:

- Author/title/keyword metadata extracted by hand from eprints
- Broadcast archive obtains programme metadata from press agency
- Medical imaging descriptive metadata entered by physician during upload

Discussion:

Bibliographic and descriptive metadata specifies what an object is, what it contains, and the context in which it was created. This is the basic information required to enable the object to be discovered (e.g. by a search engine) and interpreted. The source of this metadata varies widely depending on the nature of the individual repository so that it is difficult to generalise. For most repositories, the actual metadata generated will be a compromise between the needs and wishes of the end-user community and the costs associated with obtaining the metadata and ensuring its quality.

Goal 6.1.2: Specify technical metadata in the SIP

Examples:

- File size and checksum only to be stored
- Format information to be extracted by standard tools

Discussion:

Technical metadata describes the relation between the digital object and the content source (if the data was not born digitized) and the form of the digital object itself, such as data formats and which conversions have been made. A minimum requirement for bitpreservation is generally that checksums are stored for all objects, but technical metadata can also be much richer. For example technical metadata can include detailed format information.

Technical metadata and information about the formats can be automatically extracted during the ingest process by certain tools, including JHOVE and the New Zealand Metadata Extractor[10].

Goal 6.2: Specify the data format and metadata content for archiving digital objects (AIP)

Examples:

- Video files are stored as DVD-compliant mpeg
- Text files are stored as pdf/a

Discussion:

The repository must define the file format(s) and the necessary metadata for archived digital objects.

Where the SIPs defined in the previous goal should strike some balance between commonly used formats, and formats useful for archiving, the AIPs do not need to. Rather, it should be a specification

[10] More tools on the Library of Congress' page: http://www.loc.gov/standards/premis/tools.html

of the data and metadata that the repository will take responsibility for preserving. Decisions about formats are also likely to be the result of a cost benefit analysis, perhaps involving a choice between "raw" format, lossless compression and lossy compression. Other factors to be considered include the additional risks associated with proprietary formats.

It is difficult to make mathematical proofs about conversions, even simple ones, in computer science, and the repository should not rely always being able to avoid any conversion methods. For that reason, the repository should have a method for verifying the completeness and correctness for newly generated AIPs, as with SIPs.

Goal 6.2.1: Specify the metadata in the AIP

Examples:

- Dublin Core XML to be stored along with source data

Discussion:

The metadata from the SIP should be conserved, and elaborated, rather than being changed. A file format, that strikes a balance between being human readable and being machine parseable, such as XML could be used to encode the metadata.

The AIPs should at least, in addition to the metadata from the SIPs include metadata about:

- Preservation measures and other actions taken on the digital objects.
- Legal and administrative rights of the digital object
- A unique identifier

The AIP needs to encode which organizational bodies have responsibility for the digital object, and which legal rights apply to the object. This information should be contained in the deposit agreement under which the digital object was obtained (see the Acquisition Plan).

Other metadata, like the structural relationship between AIPs will also be relevant for certain types of repositories, and should be included.

The unique identifier must, as a minimum, be unique within the repository. Repositories should also consider using a service which provides globally unique identifiers.

Goal 6.3: Specify the data formats used for disseminating digital objects (DIP)

Examples:

- Audio files are converted on demand from .wav to .mp3 format

Discussion:

The repository must, in conjunction with its users, define useful digital object formats for dissemination. Depending on the nature of the users, such DIPs can vary tremendously. They could be as simple as the metadata encoded in with OAI or Dublin core along with links to the raw data files from the AIPs, to a web page detailing the metadata along with the datafiles converted to modern (lossy) formats, or even more extreme examples.

For each AIP scheme there must be a number of DIP schemes. The same AIP can easily have more than one DIP scheme attached, depending on the context it is disseminated in. And some AIPs will have no DIP schemes, which effectively means that there are no ways of disseminating these digital objects to users.

The DIP schemes could, and probably should, change over time, as the capabilities of the users change. This might enable, but should not require a change to the AIPs.

Having a way to guarantee the completeness and correctness of the DIPs before they are presented to the users could be nice to have, but is much less of a requirement than for AIPs and SIPs.

Goal 6.3.1: Specify the metadata in the DIP

Examples:

- The mp3 files (from the example in Goal 6.3) have the metadata encoded in id3 tags

Discussion:

The metadata can, at times, be even more interesting to the users, than the data itself, so take special care when presenting it. All the metadata from the AIP, which includes the metadata from the SIP, should be available, so the real choice lies in selecting which portions will be relevant, and which will be clutter. Many search engines will only be able to find a DIP based on the metadata in the DIP, as they will not be able to read the data files, and as such the metadata can be very important for the users searching for data.

Goal 6.4: Specify the transformation from SIP to AIP

Examples:

- The SIP is an audio CD. The conversion copies the tracks as WAV files, looks up the CDDB information, and stores it in a XML document, and the cover is scanned and stored in TIFF format

Discussion:

The repository must have clearly defined methods for creating AIPs from SIPs. These processes will collect the metadata from the SIP, and include any metadata derived from the deposit agreement and other relevant specifications, and encode it in the metadata storage format chosen for the AIP. The digital data in the SIP might have to be converted to another format more suitable for archiving, and metadata about such changes should also be stored in the AIP.

The unique identifier of the AIP will have to be generated and any other metadata not covered here should also be encoded.

The AIP must be verified for correctness and completeness, and the repository must include checks to ensure that each SIP is either used or disposed of in a recorded fashion.

Goal 6.5: Specify the transformation from AIP to DIP

Examples:

- WAV files from a ripped CD are converted to MP3 format. The metadata is encoded as ID3 tags in each file. The scanned cover file is converted to JPEG, and packaged with the MP3 files in a zip archive for the user to download

Discussion:

The repository should define the transformation methods that are used to convert an AIP into a DIP.

One of the fundamental requirements of such a transformation is that the DIPs should be authentic copies of the (contents of the) original SIPs or objects traceable to originals, irrespective of encoding formats.

While this process is simple to explain, it might be technically complex. It could potentially involve file format conversion, and re-encoding of metadata, possibly required in realtime upon a dissemination request.

Ideally the repository should be able to change this process when new conversion technology arrives, without having to change either the AIP or the DIP format.

5.7. SUCCESSION PLAN

Repositories are organizations, and like all organizations, they will have a finite and probably short lifetime. The quote from Deep Time, by Gregory Benford helps to illustrate this, and the challenges repositories face:

"We foresee the future by reviewing the past, seeking long-term trends. But this can tell us little about the deep future beyond a thousand years.

A bit over two centuries ago, what is now the Eastern United States was in the late English colonial period. At least in the European world, there were some resemblances to the current world---in fact, some countries have survived this long. For this period, extrapolation is useful in predicting at least the range and direction of what might happen.

Going back 1,000 years takes us to the middle of the Middle Ages in Europe. Virtually no political institutions from this era survive, although the continuity of the Catholic Church suggests that religious institutions may enjoy longer lifetimes. Most history beyond 1,000 years is hazy, especially on a regional scale. Prior to the Norman invasion in 1066, English history is sketchy. Beyond 3000 years lie vast unknowns; nine thousand years exceeds the span of present human history"[11]

Goal 7.1: The preservation tasks is ensured even beyond the existence of the digital repository

Examples:

- Repository establishes an agreement with another institution willing to act as successor if necessary
- Repository involves itself actively in partnerships with other similar repositories

Discussion:

It is hopefully realized by now that the proposition of retaining the usability of digital materials beyond the lifetime of the repository housing them is frighteningly difficult. A digital collection is, with the current technology, never maintenance-free. Formats need to be migrated, storage media needs to be renewed, integrity need to be checked, and so on. Failure to do so will result in the loss of usability of the collection.

As such, the only feasible strategy known to the authors remain to have other repositories take over the collections from a dying repository. The very purpose of the succession plan(s) is to detail agreement(s) about who will inherit the digital data if the repository ceases to function. Such plans should ideally include:

- Details of the inheriting repository
- The license under which it will accept the material
- The commitments to which it will pledge itself, concerning the inherited collections
- The format of data and metadata it is willing to accept the collections in
- What compensation it will receive for taking this burden upon itself

[11] Benford, Gregory, Deep Time, Harper-Collins, 2000

In practice, such detailed succession agreements may be difficult to negotiate or maintain. It is therefore vital for repositories to maintain strong links and partnerships with other similar institutions, both at home and abroad, who might stand as potential successors in case of need.

While the repository might have some choice in who will inherit its collections, the control stops there. Who will inherit the collections once this repository passes on, will be difficult to guarantee. Thus, a repository cannot effectively control or determine who will maintain its collections a generation or two from now, but it does have a great deal of control over how easy their jobs will be. The repository must not forget that one of the major reasons for following international data standards is exactly to ease the workload when other repositories take over its collections.

5.8. DISASTER PLAN

The Disaster Plan has a unique position among the SOPs. Where the other SOPs talk about how to run and ensure the continuing running of the repository, the Disaster Plan deals with handling threats to the very existence of the repository.

The focus in Disaster Plans are often on what to do when the repository shuts down for good, which is most certainly an area that needs focus, but is actually the province of the Succession Plan. One must not overlook that another worthwhile focus exists; about how to handle threats, not to the data or services, but to the repository as a whole. Such threats do not need to be hostile, they can be the natural cause of technological advances, or economic changes.

Goal 8.1: The digital repository reacts in timely fashion to substantial changes in its environment

Examples:

- Repository implements Risk Analysis and Management Strategy
- Repository details procedures for dealing with foreseeable disasters

Discussion:

The environment in which a repository operates can roughly be classified into these points, or at the very least must include these points.

- *Economic upheaval*

  Repositories with only a few major sources of income, especially those functioning as part of a larger organization, will be vulnerable to financial disruption. Very common to repositories is irregular grants of money, which could prove difficult to obtain year on year.

  The repository, or at least mature repositories, must have strategies to deal with budget shortfalls. Common strategies include willing creditors to boost the repository for a period, or having savings. The repository should also have a prioritized list of the services it provides, beyond what is absolutely required of it, to ease in selecting which could be terminated, if income shortages arrive.

- *Political upheaval*

  As the collapse of the Soviet Union and much of the Eastern Block demonstrated, political upheaval can come to seemingly stable societies. Repositories caught in such circumstances could face opposing demands to both open up their collections, and to close them down. Political upheaval could also lead to the sacking or looting of the repository, as stories about the museums in Baghdad have shown.

- *Loss of purpose/mandate*

Many repositories are not self-governed, but part of a larger organization, be it a ministry, a corporation or a university. These parent organizations could be forced to make budget cuts, or change in focus, or new leadership (or government) could have problems seeing the purpose of the repository. The repository cannot just be a passive participant, but must actively seek to demonstrate its value to those with executive power over it.

- *Technological upheaval*

  Technological changes can greatly affect the business model of a repository, and at times come about very quickly. The rise of the commonly available mp3 player could affect the business of repositories specializing in music, for example, in the form of online music shops having much of the same content. The same of course applies for technological trends in hardware, software and networking. Repositories need to watch the technological trends, evaluate which could be potentially harmful, and adapt.

- *Environmental upheaval*

  Major environmental changes can affect the community housing the repository. While many environmental disasters are unforeseeable, there are locations that are more likely to weather special occurrence. Hurricanes, and the resulting flooding are more likely to happen in certain parts of the world, and earthquakes are likewise much less rare in certain areas. The institution running the repository will be expected to have current plans to handle the more likely natural disasters and common disasters like fire, pipe leaking or just electrical blackout, but having up-to-date plans for every environmental eventuality is unrealistic. The repository should, so far as possible, detail procedures for dealing with environmental disasters in a timely manner. It is a feature of disasters that they tend to occur quickly, so preparedness is essential. For each foreseen disaster scenario, the disaster plan should aim to[12]:

  1. provide for the personal safety of all individuals who are present at the Repository
  2. maximize order, efficiency, and speed in responding to a disaster.
  3. mobilize all appropriate staff to participate in assigned functions in a disaster and in recovery operations.
  4. minimize damage or loss of repository data and to minimize the length of time in which the repository's services are unavailable to users after a disaster.

- *Loss of users and/or the arrival of competition*

  As the dissemination process is fundamental to being a repository, loss of users can affect the repository gravely.

  The repository has users to which it provides a service. As such, it is a business of sorts, and can therefore suffer competition. Even governmental backed repositories risk being challenged by other part of the government creating digital repositories. Examples of this could include the ministry of culture (having control of the public libraries) and the ministry of education (having control of the libraries connected to educational facilities) both having repository ambitions.

  Repositories could also lose users simply from being regarded as untrustworthy. Users and depositors might be worried about what the repository uses the data for, or the services the repository provides might be perceived as unstable. Whether or not such judgements are true, the repository should have a clear communication strategy to deal with them.

---

[12] These priorities are drawn from those of the Center for Southwest Studies, Fort Lewis College, http://swcenter.fortlewis.edu/Forms/DisasterPlan.htm

In any case, the repository should periodically evaluate usage statistics, and watch for services going out of favour with the users. Giving the users a way to request new features could also help the repository staying current.

- *Loss of educated key staff*

  While the Staffing plan deals with how to prevent the loss of key staff members, there should also exist plans for what to do should these preventions fail.

  There are two major consequences associated with loss of staff; the internal workings and trade secrets of the repository could be revealed to outsiders, and the services of the repository could cease to function.

  The first point is only really a threat to the trustworthiness of repository if the internal workings include untrustworthy procedures, that have not previously been revealed to the users. If this is the case, the repository should prepare press statements and other communications for when the secrets break.

  The second point could prove more problematic. The Staffing Plan deals with ways to mitigate this threat. However if it occurs despite such plans the best course of action would be to shut down unmaintainable services until new staff have been hired or trained. While this will cause a loss of trust in the repository, continuing to run unstable services will also have this effect. Nevertheless, sometimes, it will not be politically feasible to shut down services. In such cases the repository must refer to the prioritised list of service referred to under 'Economic Upheaval' and sacrifice lower priority services.

- *Breach of Security*

  The physical security of the repository can also be threatened. Depending on the nature of the materials housed, and the nature of the building housing the repository, such threats can take many forms. A repository housed in a library could lose users if the users perceived the building to be a likely terror target, for example, while a repository housing sensitive data could be regarded as untrustworthy if no measures are taken to prevent physical theft of data or hardware containing data.

  The repository should identify likely threats, and address them, in addition to having plans in the event of these threats materializing.

5.9. PRESERVATION PLAN

A repository's preservation planning objectives are likely to correspond to a time-scale far beyond that of the immediate short, medium or even long term strategic management of the organisation. Implicit preservation agendas will in many cases demand the continued availability and understandability of digital resources for many, many years, if not in perpetuity. It is therefore difficult to conceive of broadly expressed preservation goals that can be feasibly and meaningfully measured. An obvious implicit goal of most repositories will be to preserve one or more varieties of content, from one or more sources, to facilitate long term access and use by one or more sets of users. However, given the highly temporal context within which success or otherwise is determined, this presents challenges of expression. It may be realistic within a shorter term period to identify where preservation is *not* being achieved successfully, but proof of successful preservation is only really evident at an unknown point in the future where information is needed, and utilised successfully. Unless preservation practitioners can guarantee that their own lives, and that of their erected preservation infrastructures will extend to the point where data is no longer of value, an alternative means of objective definition and evaluation is required.

Preservation is often described as being akin to ensuring interoperability with the future; this notion offers opportunities for exploration. Interoperability is a challenge even within the contemporary; a

useful, and measurable starting point for repositories is to demand that specific information assets, or classes of information, are maintained sufficiently to support their usability on all of the currently available platforms by a sufficiently wide and diverse range of user communities. This is realistic, measurable, and can be made sufficiently specific by detailing explicitly the platforms (hardware and software) upon which information can be satisfactorily accessed and understood, and by defining the communities (and their knowledge bases) that must be capable of doing so. As discussed in other sections, there are considerable difficulties in repositories who seek to *prove* their own sustainability. However, irrespective of unknown or unpredictable issues that can threaten the repository itself, preservation objectives can be largely met by ensuring that information is maintained, with sufficient contextual and representation information to facilitate its use by identified user communities using contemporary and emerging technological platforms.

The introduction of time constraints must be both considered and realistic; it is of little value for repositories to formalise an objective to preserve content in perpetuity or until its value is lost. Instead, preservation aims should be broken into smaller, regular and more predictable temporal chunks, interspersed with information appraisal to determine that which must continue into the next phase of preservation. Approaching preservation aims as a series of short to medium term goals emphasises the information's contemporary value and opportunities for adding value (a central tenet of digital *curation*) and also makes more explicit the active, and not reactive, nature of preservation activities.

Goal 9.1: Repository must maintain understanding of contemporary and emerging hardware, software and storage technologies

Examples:

- Maintain and document technical, social and legal analyses of Microsoft Windows XP, Linux, Mac OS X, Sun Solaris and Novell and other contemporary operating systems
- Maintain and document technical, social and legal analyses of x86, AMD64, PowerPC, SPARC and other contemporary architectures
- Maintain and document technical, social and legal analyses of optical disc, LTO tape, optical tape, solid state, hard disk and other contemporary media storage devices;
- Maintain technology watch aimed at identifying emerging hardware, software and storage technologies appropriate for subsequent analysis

Discussion:

In order to confidently seek to preserve digitally encoded content, repository practitioners must maintain a detailed and widespread understanding of current and emerging trends within the technological domain. The required intimacy of understanding is not limited to strictly technical concerns, although of course these are of tremendous importance. As well as developing expertise with issues of software, hardware and media architecture, performance and vulnerabilities, repositories must similarly familiarise themselves with legal contextual issues associated with particular devices and applications (such as intellectual property rights, often described within end user license agreements), and more social issues such as ubiquity of use, vendor stability, and any identifiable and commonly implemented combinations of technology or modifications.

Goal 9.2: Repository must maintain understanding of all structural (e.g. file encoding) standards and formats

Examples:

- Formally document technological, social and legal characteristics of each accepted or potentially acceptable file format in a format-specific action plan

- Formally document technological, social and legal characteristics of each archival file format used or planned for use within a format specific preservation plan

Discussion:

File format analysis is an invaluable part of the repository's preservation activities, although by no means should it be the sole focus for determining appropriate preservation solutions. Nevertheless, formats provide a technical structure that lends information a physical form. Again, there are numerous technical, legal and social considerations that must be fully explored to determine both opportunities and vulnerabilities implicit in particular formats. However, preservation strategies based solely on format are insufficient. Considerable information value is attributable not to specifics of structure, which the format analysis is capable of conveying, but to the semantic characteristics of each information object, and the expectations of users and their anticipated uses, which are much more fundamental and meaningful concepts.

Goal 9.3: Repository must maintain understanding of identified user communities and their associated competences and knowledge base

Examples:

- Formally identify and document each user community with reference to their available expertise, knowledge and technological aptitudes as well as their usage expectations
- Continue to monitor these communities, amending documentation and adding, removing, splitting or merging communities where appropriate

Discussion:

As described above, an understanding of the structural context that surrounds each information object is insufficient to facilitate preservation for the potentially diverse communities that will ultimately demand information understandability. It is vital that repositories maintain relationships with each of their identified user communities, and where appropriate, pursue new ones. Information describing expectations and capabilities of user communities should be maintained on a continual basis, with every aspect of preservation activities ultimately influenced and informed by the end user needs. It is by analysing such information that many threshold standards for preservation strategies can be meaningfully defined. Irrespective of the time-scales over which preservation will take place, the repository must continue to acknowledge that its ultimate purpose is to facilitate useful access to their user base, which itself is constantly evolving, expanding and diversifying.

Goal 9.4: Repository must maintain understanding of preservation requirements for each stored information asset or class of information

Examples:

- Objectives here must make explicit:
  - Record requirements relating to content, behaviour, appearance, context, interpretability, and interoperability must be made explicit defining maximum or minimum thresholds that any acceptable strategy must preserve;
  - Technical requirements relating to, for example, ubiquity of formats, lossy/lossless-ness, error tolerance or documentation must be made explicit, defining maximum or minimum thresholds that any acceptable preservation solution must satisfy;
  - Infrastructure requirements relating to hardware, software, or staff requirements must be made explicit, defining maximum or minimum thresholds that any acceptable preservation strategy must satisfy;

- Process requirements relating to, for example, the extent to which strategies can be automated, validated or scaled must be made explicit defining maximum or minimum thresholds that any acceptable preservation strategy must satisfy;

Discussion:

The success or otherwise of preservation activities can be measured in a meaningful way, but only if preservation requirements are strictly and thoughtfully defined. Resources such as the *Plato*[13] preservation planning tool developed by the PLANETS project can greatly facilitate the processes of defining and prioritising preservation targets, which may be associated with objects themselves, or various contextual factors that surround them or the preservation process. It is essential to consider that the appropriateness of particular preservation strategies will be determined only by considering specific characteristics of individual information objects and the implicit challenges in exercising their preservation. Strategies determined at the level of technology or structural characteristics are insufficiently focused; for example, a repository may elect to preserve all Microsoft Word encoded content as plain text. This is only a viable strategy if the value of every one of these objects is wholly associated with the textual content within. If for example user communities are interested in issues of formatting, pagination, layout, embedded objects or file metadata, then clearly these are not being preserved within such a simplistic and generically applied approach.

Goal 9.5: Repository must maintain, exercise and evaluate preservation strategies capable of meeting specific preservation targets

Examples:

- Develop and evaluate preservation strategies with reference to preservation requirements
- Deploy preservation strategies when technical, legal, social and community monitoring activities identify vulnerabilities with existing preservation information infrastructures

Discussion:

The contextual awareness maintained by the repository should provide sufficient information to trigger the execution of preservation strategies when they become necessary. Numerous developments may be sufficient to provoke preservation. Technology may appear to lose its currency, vendors may announce a cessation of trading or discontinuation of support for products, legislative developments may introduce additional vulnerabilities associated with preserved content or user communities may demonstrate shifting expectations, requirements or capabilities. When currently preserved content begins to exhibit fragility, the repository should have at its disposal one or more preservation strategies capable of protecting the understandability and value of its information from such pressures. Irrespective of whether the employed strategy is emulation, migration or any other emerging approach, it should be thoroughly evaluated, in an appropriate test-bed environment, with results considered in direct comparison with preservation requirements defined for specific information objects, or classes of information.

Goal 9.6: Repository must maintain and exercise appropriate appraisal policies to determine which information must continue to be preserved

Examples:

- Develop and maintain criteria capable of expressing the extent to which preserving specific content is worthwhile, and aligned with repository's purpose. This may include coverage of:
  - its relevance to organizational mission

[13] http://olymp.ifs.tuwien.ac.at:8080/plato

- the extent to which its preservation adheres to organizational policy
- its perceived level of authenticity
- its perceived integrity and usability
- the perceived strength of its provenance
- its condition or completeness
- availability of contextual information (e.g. metadata)
- its accuracy

- Deploy criteria on a frequent and regular basis, removing or redeploying content as is necessary to maintain a focused preserved collection

Discussion:

In some respects, the need to continually appraise the value of information throughout the lifetime of its usefulness is not strictly a component of the preservation process. Nevertheless, it will be instrumental in determining the extent of information objects, or classes of information, that will continue to be preserved. Building appraisal into the preservation process provides natural punctuation between each discrete preservation step, validating and justifying subsequent preservation activities. It is a means to effect control over the preservation process and ensure that ongoing awareness is maintained of the value of individual digital assets, and how this corresponds with more widespread goals and priorities of the organisation.

6. FROM PLATTER TO TRUST

The PLATTER tool is concerned exclusively with management of the objectives and targets of repository. It is not itself a tool for establishing trust and is not intended to compete with other initiatives in that area. Moreover, it is important for repositories to be aware that at the present time there does not exist any recognised international authority accredited for the auditing or certifying of repositories. There do exist initiatives which are aimed at standardising the audit process, for example the Birds of a Feather Group on Digital Repository Audit and Certification[14]. However, until such groups establish accepted standards in this area it will be up to individual repositories to define their own standards and procedures for establishing trust in consultation with their stakeholders. We can identify this is a primary goal of the repository:

Goal 0: In consultation with stakeholders, establish criteria for trust

The criteria to be used should draw on the existing initiatives in this area - for example the TRAC[15] and nestor[16] checklists, the DRAMBORA toolkit[17], OAIS[18] etc. For example, a repository might set as its criteria for trustedness that it complete a self-assessment exercise based on the DRAMBORA toolkit with the results of that self-assessment to be audited by two recognised experts in the field. Another repository might prefer to base its criteria for trust on a checklist approach. Some of the issues to consider in choosing an audit strategy are

- Which existing audit tool (if any) is most suitable for the particular repository?
- Is an external audit necessary and if so what criteria should be used in selecting external auditors?
- When and how often should an audit be carried out?
- Is failure a possible outcome of the audit process and if so what plans need to be made to deal with that eventuality?

PLATTER has been designed to support both checklist and risk-analysis based approaches to audit.

---

14 http://wiki.digitalrepositoryauditandcertification.org/bin/view

15 "Trustworthy Repositories Audit & Certification (TRAC) : Criteria and Checklist"

http://www.crl.edu/PDF/trac.pdf
http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

16 The nestor catalog of criteria for trusted digital repositories

http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

17 A McHugh, S Ross, R Ruusaleep & H Hofman, The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)', http://www.repositoryaudit.eu. 2007. ISBN: 978-1-906242-00-8

18 The Open Archives Initiative Protocol for Metadata Harvesting

http://www.openarchives.org/OAI/openarchivesprotocol.html

6.1. PLATTER AND CHECKLISTS

The discussion of the SOPs in Section 5 lists a large number of points to be addressed in the development of detailed objectives and targets. These are overwhelmingly drawn from the TRAC and nestor checklists. We have not sought to identify which concepts have been drawn from TRAC and which from nestor or other sources, so a repository which chooses its own priorities amongst them may find that it does not satisfy all the checkpoints on its chosen checklist. Therefore any repository seeking to satisfy such a checklist comprehensively should use it in combination with PLATTER during the planning stage. Nevertheless PLATTER does try to be comprehensive, in the sense of covering all the major points identified on other checklists. Thus any repository which adopts the PLATTER philosophy, that all points covered in the discussion of the SOPs must be addressed, will be guaranteed to have covered the majority of the points made in both the TRAC and nestor checklists.

6.2. PLATTER AND DRAMBORA

PLATTER is designed to complement DRAMBORA and a repository planned using PLATTER will be strongly placed to use DRAMBORA as a self-assessment tool. In DRAMBORA, the initial stages of the risk analysis require the repository to identify and document its goals. The repository then proceeds to describe the activities it undertakes in pursuit of those goals and the assets which it deploys. The DRAMBORA risk analysis then consists of identifying threats to the achievement of those goals. When PLATTER has been used for planning of objectives, a repository will be in a very strong position to carry out an effective DRAMBORA analysis because all its current objectives will be thoroughly documented. The combination of PLATTER and DRAMBORA therefore represents a powerful tool in the development of Trust.

Appendix: Dependencies Between Taxonomic Axes and SOP Objectives

In this section we attempt to identify in tabular form some of the key relationships between the goals outlined in the discussion of the SOPs in Section 5 and the questions posed in Section 3 on Repository Classification. Generally speaking, the individual taxonomic axes defined by the questions posed in Section 3 are of relevance to at least one, and typically several of the goals which individual repositories will have to formulate. The only question we have not explicitly linked to any specific goals is question 4.2 on interoperability standards. This is not because interoperability is unimportant to a repository's objectives. On the contrary, we firmly believe that interoperability in all areas of repository operation is and will be of increasing importance throughout the foreseeable future. We have, rather, chosen not to list the dependencies of operational goals on interoperability precisely because they are potentially so wide reaching and so variable from repository to repository. For example, if a repository chooses to interoperate with other repositories in the use of services associated with preservation (e.g. format characterisation) then this will couple question 4.2 to the Preservation Plan SOP. Other repositories may perhaps use interoperation to create coordinated search interfaces, or perhaps automated metadata extraction. Each of these will introduce quite new dependencies within the PLATTER framework. We would therefore foresee that in the near future, as the interoperability landscape becomes clearer, Question 4.2 will need to be refined to address the separate areas addressed by these new interoperability frameworks which are currently being developed.

We have also tried to identify dependencies between different goals, especially those in distinct SOPs. Several of the goals clearly have such strong dependencies on so many of the other goals that it would be confusing to list them all. One such goal is goal 3.1 on Staff Roles. Since all our goals are, by assumption, SMART it follows that they are all assignable, and therefore that every goal is associated with at least one Staff Role. Similarly all, or almost all, goals are linked to Goal 1.1, the maintenance of a business plan, since every goal will have some cost associated with achieving it. Finally, we have not listed all the goal linkages of the Data Plan because, again, there are so many. Essentially all the goals in the Data Plan are strongly linked, at least, to both the Acquisition Plan and the Access Plan, and also the Preservation Plan.

Goal	Coupled Goals	Coupled Questions
1.1 Business Plan	3.2 Staffing 4.1 Mission Statement 5.1 Technical Scale	1.2 Commercial Status
1.2 Financial Support	1.4 Outreach	1.1 Mandate
1.3 Financial Contingency Plans	7.1 Succession 8.1 Disaster	
1.4 Outreach	2.1 Acquisition 4.2 Designated Communities	1.1 Mandate 1.2 Commercial Status
2.1 Acquisition	4.2 Designated Communities	1.1 Mandate
2.2 Deposit Agreements	2.5 Maintain Agreements 1.1 Business Plan 6.1 SIP Objects 6.1.2 SIP Bibliographic Metadata 6.1.3 SIP Technical Metadata	1.3 Legal Acquisition 3.1 Acquisition Method 4.1 Source of Metadata
2.3 Physical Control	5.3 Technical Services	4.3 Storage Strategy
2.4 Monitor Ingest	6.1 SIP Objects 6.2 AIP Objects 6.4 SIP -> AIP	3.2 Data Complexity 3.3 Data Specialisation 4.4 Software Strategy
2.5 Maintain Agreements		

Goal	Coupled Goals	Coupled Questions
3.1 Staff Roles		2.3 Staff Size 1.2 Commercial Status 3.3 Data Specialisation 4.1 Source of Metadata 4.3 Storage Strategy 4.4 Software Strategy
3.2 Staffing		2.3 Staff Size 4.3 Storage Strategy 4.4 Software Strategy
3.3 Staff Skills		3.3 Data Specialisation 4.4 Software Strategy
4.1 Mission Statement	1.1 Business Plan 9.3 Community Understanding 1.4 Outreach	1.1 Mandate 1.2 Commercial Status 3.5 Access Rights
4.2 Designated Communities	5.1 Technical Scale 5.3 Technical Services	1.1 Mandate 3.5 Access Rights
4.3 Access Policy	5.2 Technical Security 6.2.1 AIP Metadata	1.2 Commercial Status 3.5 Access Rights

Goal	Coupled Goals	Coupled Questions
4.4 Dissemination and Access	6.3.1 DIP Metadata 6.3 DIP Objects 6.5 AIP -> DIP 4.2 Designated Communities 5.3 Technical Services	3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights 4.4 Software Strategy
5.1 Technical Scale	1.1 Business Plan 3.3 Staff Skills	2.1 Data Quantity 2.2 Object Quantity 2.4 Enduser Quantity 4.3 Storage Strategy
5.2 Technical Integrity and Security	4.3 Access Policy 3.3 Staff Skills 9.5 Preservation Strategies	3.1 Acquisition Method 3.4 Data Sensitivity 3.5 Access Rights 4.3 Storage Strategy 4.4 Software Strategy
5.3 Technical Services	9.1 Technical Understanding 3.3 Staff Skills	3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights 4.3 Storage Strategy 4.4 Software Strategy

Goal	Coupled Goals	Coupled Questions
6.1 SIP Objects	2.2 Deposit Agreements	3.2 Data Complexity 3.3 Data Specialisation
6.1.2 SIP Bibliographic Metadata	2.2 Deposit Agreements	3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights 4.1 Source of Metadata
6.1.3 SIP Technical Metadata	2.2 Deposit Agreements	3.2 Data Complexity 3.3 Data Specialisation
6.2 AIP Objects	9.1 Technical Understanding 9.2 Format Understanding 9.5 Preservation Strategies	3.2 Data Complexity 3.3 Data Specialisation
6.2.1 AIP metadata	9.5 Preservation Strategies	3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights
6.3 DIP Objects	4.2 Designated Communities	3.2 Data Complexity 3.3 Data Specialisation
6.3.1 DIP Metadata	4.2 Designated Communities	3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights

Goal	Coupled Goals	Coupled Questions
6.4 SIP -> AIP		3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights 4.4 Software Strategy
6.5 AIP -> DIP		3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights 4.4 Software Strategy
7.1 Succession	1.3 Financial Contingency	1.2 Commercial Status 3.4 Data Sensitivity 3.5 Access Rights
8.1 Disaster	5.2 Technical Integrity and Security 1.3 Financial Contingency Plans 9.5 Preservation Strategies	4.3 Storage Strategy
9.1 Technical Understanding	3.3 Staff Skills	4.3 Storage Strategy 4.4 Software Strategy
9.2 Format Understanding	6.2 AIP Objects	3.2 Data Complexity 3.3 Data Specialisation

Goal	Coupled Goals	Coupled Questions
9.3 Community Understanding	4.2 Designated Communities	3.2 Data Complexity 3.3 Data Specialisation 3.4 Data Sensitivity 3.5 Access Rights
9.4 Preservation Understanding	3.3 Staff Skills	3.2 Data Complexity 3.3 Data Specialisation
9.5 Preservation Strategies	6.2 AIP Objects 6.2.1 AIP Metadata 8.1 Disaster 5.2 Technical Integrity and Security	3.2 Data Complexity 3.3 Data Specialisation 3.5 Access Rights 4.3 Storage Strategy 4.4 Software Strategy
9.6 Appraisal		1.1 Mandate

Project information

Project acronym:	**DPE**
Project full title:	**DigitalPreservationEurope**
Proposal/Contract no.:	**IST-2006-034762**

Project Officer: Manuela Speiser

Address:	INFSO-E3 Information Society and Media Directorate General Content - Learning and Cultural Heritage Postal mail: Bâtiment Jean Monnet (EUFO 1167) Rue Alcide De Gasperi / L-2920 Luxembourg Office address: EUROFORUM Building - EUFO 1167 10, rue Robert Stumper / L-2557 Gasperich / Luxembourg
Phone:	+352 4301 33632
Fax:	+352 4301 33190
Mobile:	
E-mail:	manuela.speiser@cec.eu.int

Project Co-ordinator: Prof Seamus Ross

Address:	HATII, University of Glasgow
Phone:	+44 141 330 3635
Fax:	+ 44 141 330 2793
Mobile:	+44 797 435 7006
E-mail:	s.ross@hatii.arts.gla.ac.uk