LongRec – Records Management over Decades

Part 1: The research project

1 Objectives

Primary objective:

Persistent, Reliable and Trustworthy Long-Term Archival of Digital Documents, with Emphasis on Availability and Use of Documents.

The particular problems addressed by LongRec typically emerge when document lifetime exceeds 20 years, and LongRec imposes no upper limit on the lifetime. The main objectives for the LongRec partners are:

Objective 1: To enable transition to digital original documents and digital work processes for information that must be available and in use over decades.

Objective 2: To explore the potential for commercial products/services in this area.

Case studies addressed by LongRec include documentation for physical objects that are in use for decades (ships, oilrigs, power plants, and others) and documents in public registries. Results may be generalised to other cases, e.g. health information.

LongRec goes beyond the "digital preservation" area addressed by libraries and (public) archives in that documents also need to be used (retrieved, updated, verified) subject to constraints related to ownership and authorisations. All parts of a document's environment (technology, processes, organisations, roles/people, and ownership) must be expected to undergo several changes during the lifetime of the document.

LongRec goes beyond state of the art in "records management" by addressing long-term aspects and preservation not only of availability and readability, but also of semantic value (meaning, context) and evidential value (trustworthiness).

Secondary objectives

1. Identify obstacles for, and innovation potential from, digital work-processes:
   - Identify obstacles to use of long-term, digital original documents,
   - Identify requirements/means to solve these obstacles,
   - Explore the potential for innovative products and services to support these requirements.
2. Develop a set of publicly available guidelines/methods supporting "best-practice" in long-term document management, focussing on the complete document life-cycle (creation, maintenance, use, discarding), and providing:
   - Availability and readability of documents,
   - Semantic value of documents (context, metadata, etc.),
   - Trustworthiness of documents (authenticity, reliability, security, history).
3. Design and develop reusable (across many scenarios) mechanisms for implementation of the guidelines.
4. Pilot digital work-processes based on long-term digital documents:
   - Verified implementation strategies based on both case studies and piloting,

- Evaluate and recommend long term strategy for the project participants' use of digital storages and transformation to digital work processes

5. Strengthen and promote the understanding of management of long-term digital documents in the educational system, to the academic research community, and to the society at large:
   - Four Ph.D. degrees,
   - Publication in international journals and conferences, as well as to national conferences, workshops, newspapers, and journals.
   - Submission of guidelines/methods to international standards bodies where appropriate, with necessary follow-up work to pursue adoption.

2 Frontiers of knowledge and technology

Our modern knowledge based industry and society heavily depend on digitally available information. Due to low storage cost it appears as if data storage is out of control with annual growth rates of 30 % [1, 2]. Focus has therefore turned more and more towards records management to handle creation, management and preservation of information [3]. A variety of standards and procedures for records management, software applications and metadata are available like ISO 15489 (Information and Records Management) [38], MoReq (Model Requirements for the Management of Electronic Records) [4], DoD 5015.2-STD (guidelines for record management) [5], DIRKS (methodology for the design and implementation of recordkeeping systems) [6], ISO14721 (minimum requirements for an Open Archival Information System (OAIS) [39]; it provides however no design or implementation specification). The National Archive Services of Norway (Riksarkivaren) maintains the NOARK functional specifications, which are mandatory for the public sector in Norway [7]. Work on NOARK-5 is ongoing. The NOARK work has achieved international acclaim.

Although recordkeeping systems and IT standards tightly depend on one another there seems to be a lack of participation from archivists and towards development of IT standards [8]. DNV has performed a series of internal studies clearly pointing towards the need for a clarification of the requirements connected with the (future) usage of document and records [9, 10]

Data storage is cheap – information management however is expensive and policies must be in place to select what to preserve. Here, legal aspects will be determining but visions about potential future use (and thereof business) may have even more influence [11]. It is hard to imagine what metadata some future user might want know about a century-old digital object especially as the concept of quality rests totally on *"…satisfying implied needs"*(ISO 8402). If too little background information (metadata [12]) is preserved, data will become useless or non-interpretable even if the syntax is readable. A self-describing information repository (data and metadata) could be generated using XML. XML is both a file format and a text-based human-readable markup language, which is hardware and operating system independent [13] but it seems there is still a generic data architecture lacking for archiving and analysis [14]. Large amounts of data and metadata may have to be stored in order to maintain trust in the authenticity of a digital object and to be able to interpret or execute it correctly.

One aspect of interpretation is to present the object to a human in the way intended by the producer of the object. Another aspect is describing the meaning of information in a standardised way such that computers can understand and handle it. This is the goal of the semantic web initiative [15]. This initiative tries to achieve semantics by using markup languages, a resource description framework (RDF) as a way to exchange machine understandable descriptions [16, 17], and ontologies which define relations between digital objects (Web Ontology Language (OWL) [43] and ISO 15926 [41]). The main benefit of ontologies is considered to be the enhanced (automated) search accuracy [18] but it

still remains an open issue how much knowledge an ontology can effectively represent [19]. In addition, the meaning of words may change over time and vary geographically, representing a further hurdle in storage and for search algorithms [20].

Exchange of data in electronic collaboration has raised the need for common ontologies describing the semantics of data exchanged. Initiatives such as OASIS UBL [40] and the Ontology Forum [42] address this issue.

Simply printing the digital file on paper may work for text documents but not for spreadsheets with embedded formulas or databases (like DNA sequences with terabytes of data). How can digital information be accessed when new versions of databases, spreadsheets, or word processors are released every two to three years [21]? The fragility of digital media and the rate at which computer hardware and software become obsolete have so far prevented the development of preservation methods guaranteeing that information will be readable and understandable centuries from now [22, 23].

To avoid loss of data, persistence in storage must be guaranteed, possibly by use of redundancy, and data must routinely be migrated from old to new media (within the same technology and on changes to new storage technology) [24, 25, 26, 27]. Actors like national libraries seem to be confident in their approach to this logistics challenge. As an example BBS [28] has guaranteed storage for decades to customers. This confidence however rests on a fragile assumption: that migration does not alter the characteristics of the migrated and the original digital object.

In addition, different digital objects may require different preservation strategies [29] as the authenticity of a document may be altered. An example is preservation of digitally signed documents. A digital signature binds not only to content but also to the format of the signed object, and by necessity the signature cannot survive conversion. One may convert a digitally signed object by adding signature verification "evidence" to the migrated object, stating that this object was originally signed. In this case, the conversion process may need to be controlled by a neutral party (removing the signatures of your counterparts e.g. contracts and replacing them with your own statements about signature validity is not recommended). The lifetime of a digital signature is limited also by lifetime of cryptographic algorithms and keys (that can be broken given advances in computing power or discovery of weaknesses like the recent attack on the MD5 hash algorithm [30]), lifetime of signature and certificate formats (which may also need to be migrated), and possibly also lifetime of trusted actors (e.g. for certificate status information) [31]. Once encryption is broken, fake digital signatures – indistinguishable from genuine ones – can be created [32], although it is possible to protect weak signatures by adding further signatures enveloping the signed object. ETSI has developed standards for archive formats for digitally signed objects but these are not in widespread use [33, 34].

LongRec addresses live records, i.e. records that are in active use and may therefore be subject to changes. Hence digital logs may be kept as they essentially represent forensic fingerprints of IT processes [35]. They shed light on what, who, when, where and how an IT process was involved in alteration of e.g. documents and so minimize legal risk as well as increasing trust in the recorded data or document [36, 37].

3 R&D challenges

Topic and overall challenges

The overall research challenge for LongRec is to establish theory, mechanisms, and technology that enable companies to trust long-term (several decades) storage of digital original documents, and to be able to use and update the documents throughout their lifetime. LongRec addresses *records management*. A (archive) record is created when one or more documents and supporting information (context information, semantic information, and presentation information) are assembled and stored.

LongRec addresses the case where such records must have an active lifetime (in the sense of being updated and used) of several decades.

An archive record is created and used in work processes related to *document management*. LongRec addresses such processes through the case studies, ensuring that the records management solutions developed are capable of supporting document management.

To achieve persistent, reliable, and trustworthy storage and use of digital original documents over decades, the problem area has to be covered in full breadth and sufficient depth. Any single remaining issue may be enough to halt the transition to digital documents! Thus, a broad scope is necessary, with research focussing on the areas where state of the art is not sufficiently developed at present. In other areas, state of the art will be applied and improved, and best practice recommendations will be developed. Based on the analysis done during the writing of the LongRec proposal, the following main research challenges are identified:

1. Records transitions survival,
2. Long-term usage,
3. Preservation of semantic value,
4. Preservation of trust and security,
5. Legal, social, and cultural framework.

These challenges are detailed in the following.

Research challenge 1: Records transitions survival

A record will experience many major changes over a lifetime of decades (say, 50 years or more). The record (or at least the information conveyed by the record) must be expected to live longer (meaning also being useful) than any other component in the system; be that technology, actors, or organizations, and must be able to survive transitions related to any of these other components. This imposes the following challenges:

- Limited lifetime of technology:
  - Storage media and related technology. This is briefly discussed in chapter 2.
  - Formats and related technology – document formats, formats of metadata, presentation information, signature formats, etc. Support for format conversion is necessary since one cannot rely on continued support for old formats. Reliable format conversion is a challenge, and is only possible if formats are open and specified in a proper way.
  - Lifetime of technology platform – hardware, base software (OS etc.), archive and document management systems, databases and so on. Export and import functionality between platforms is needed. Again, open formats are required.
  - Lifetime of algorithms and functions (e.g. cryptographic algorithms and their application, such as digital signatures).
- Limited lifetime of organizations and organizational units. This includes sell-outs and mergers and even termination, as well as changes in ownership and authorisations.
- Limited lifetime of ownership. When an object (say, a ship) is sold, the accompanying documentation must follow the object, change ownership, and possibly be migrated to new systems and new technology.
- Limited lifetime of actors, roles and processes. Work procedures, responsibilities and persons

change over time. This has particular relevance to security (ownership, authorisations and more), but also to semantic interpretation of documents. Outsourcing of (parts of) processes and operations may pose particular challenges.

- Limited lifetime of external reference data and services from external actors. If a piece of (contextual) information is maintained by an external actor, continuation of this actor's support may be needed.

Moreover, not only individual records need to survive transitions, but also contextual information, history information, and relationships between records.

The transition survival challenge can only be solved by proper maintenance procedures for records and their storage. Since any technology (and organizations, processes etc.) in use at the time of record creation will be obsolete long before the expiry of the lifetime of the record, and one cannot predict the nature of the technology (organizations, processes etc.) that will constitute the replacement (say, 15 years from now), specifications must be sufficiently independent of technology. At the same time, specifications must be supported by existing technology. Note that in particular conversion and versioning of context information (metadata etc.) is a fairly new research area where the partners of LongRec are in a good position to contribute.

LongRec will address all aspects listed above, with particular focus on formats, organizational issues, and processes. Research in records transition survival will partly be carried out by one Ph.D. student localised at NTNU.

Research challenge 2: Long-term usage

LongRec focuses on information that is accessed, used and changed during its entire lifetime. Thus, availability and accessibility is very important. All aspects, such as search, retrieval, presentation, verification (see research challenge 4), and change, need to be addressed.

An important aspect is the ability to find all relevant information. Partly, this can be achieved by structuring (and maintaining over time) relationships between documents and their context. However, a general trend today is that search functionality becomes increasingly important. Thus, structuring the documents and their context information in a way that facilitates reliable searching will become an important measure.

Searches in old documents brings on challenges related to terms and relevance; which terms are used in a 30 years old document, and which of today's terms cannot be used or have to be translated into their "old" more or less equivalent terms?

LongRec will address these topics with particular focus on search. Research in long-term usage will partly be carried out by one Ph.D. student localised at NTNU.

Research challenge 3: Preservation of semantic value

When committing to digital collaboration it is a prerequisite that semantics of documents used as input and output to business processes are agreed. This particularly applies for processes where documents are exchanged between systems. However, semantics are important also for processes where documents are exchanged digitally between human users.

Semantic information is partly stored in the records themselves and partly as separate objects in an archive (typical for information common to many records). Additionally, referral may be made to information maintained by external parties. Examples of distributed information sources in Norway are www.lovdata.no (laws and regulations), Enhetsregisteret (public company registry), and sources of relevant metadata (e.g. Oppgaveregisteret for reporting to the public sector). Thus, one may need to

rely on the continued support and trustworthiness of external parties.

Not all contexts are storable (e.g. a person's explanation of why the answer to a particular question is as indicated), and a document and its context may over time make less sense due to changes in terminology and advances in knowledge in the field of the document. The inevitable reduction in semantic value will imply decreased trustworthiness and usefulness of the document. Updates to the document may mitigate this.

Methodologies for knowledge engineering have been around for some years (UML was defined in the mid nineties; OWL is a few years old). In a long term view, the methodologies used for describing semantics are immature and the best of breed are rapidly changing. On this background the following research challenges are important for the semantic aspects of LongRec:

- Transition of ontologies between generations of ontology methodologies,
- Maintaining distributed ontologies,
- Requirements to the initial ontology used,
- Identify areas of ontology engineering that possibly will be exposed to significant changes.

The research area must be regarded as relatively immature, but the LongRec consortium is in a good position to significantly advance state of the art in the field.

Research in preservation of semantic value will partly be carried out by one Ph.D. student inaugurated at NTNU but hosted by DNV IQM.

Research challenge 4: Preservation of trust and security

In the context of LongRec, digital original documents must be trusted over time. Additional properties with respect to the items in research challenges 1-3 are confidence in preservation of: Availability (to authorised actors), integrity (correctness), confidentiality (to unauthorised actors), IPR protection (of ownership), and accountability (traceability of actions and events related to the document).

These properties are threatened by errors, mistakes and failures, but also by security events (i.e. intentional attacks). LongRec proposes to use the *evidential value* of a record as an index for the degree of trust, and to derive guidelines for preservation of evidential value over time. With respect to a record, the following questions must be answered:

- What are the requirements for evidential value of the record? E.g. which compliance requirements (legal, privacy, etc.) must be addressed?
- How are these requirements met at the time of creation, and how do they change over time? E.g. requirements for confidentiality and availability can typically be relaxed over time.
- How can evidential value best be preserved over time according to the (partly changing) requirements of the record?
- How can evidential value be verified when required? E.g. requirements for trust in the verification process and related software, and requirements for use of independent actors.

As a rule, it is inevitable that the evidential value decreases over time. Any operation on a document, such as format conversion, implies a risk and involves actors and components that must be trusted.

LongRec will develop theory and methods to as far as possible preserve evidential value over time, and to express, measure, and verify evidential value. As all security related work, analysis of threats, risks and vulnerabilities must be applied. A general risk analysis may be augmented by specific analysis related to case studies. Two particular topics that will be addressed are signatures and IPR (Intellectual

Property Rights).

Digital signature as a mechanism for integrity and accountability receives particular attention due to legal and security requirements. As a digital signature binds also to the formats used, a digital signature by necessity cannot survive format conversion. Long-term aspects of digital signatures will be studied by LongRec, including methods to cope with removal of signatures without too much loss of evidential value (use of neutral actors and their services may be necessary).

IPR issues, in a broad sense, can if not properly addressed impede long-term storage and access activities. Ownership and IPR necessarily change over time, must comply with relevant laws and regulations, and dictate management of authorizations given to other actors. IPR can be considered as parts of the context information of a record. Both customary access management solutions and a DRM (Digital Rights Management) approach will be studied by LongRec.

Research in preservation of evidential value will partly be carried out by one Ph.D. student inaugurated at NTNU but hosted by NR.

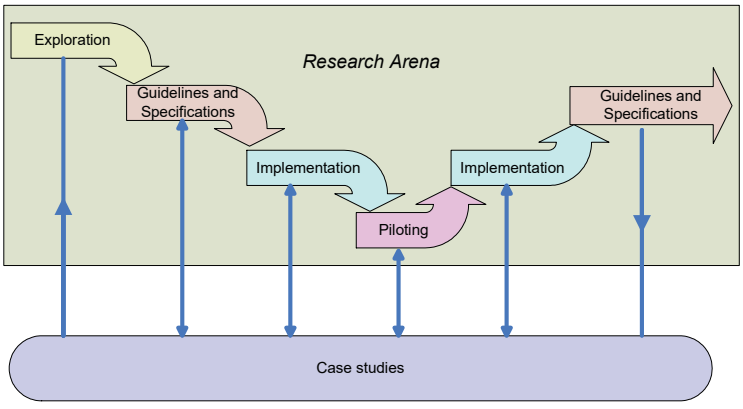
Research challenge 5: Legal, social and cultural framework

LongRec will primarily address issues raised by a changing and evolving environment (with respect to the actors) during the exploratory phase of the project. Deep research in this area will require other competence and possibly other partners, such as experts in law, economics, society, culture, and politics. However, existing issues in areas such as compliance requirements will be explored by LongRec.

Since legal, social and cultural framework must be viewed in an international perspective, the partners of LongRec will consider an application for a project under the 7th Research Framework Programme of the EU particularly addressing research challenge 5.

4 Research approach/methods

A phased approach is selected for the research in LongRec: Exploration, guidelines and specifications, implementation, and verification and validation trough piloting. A primary arena for this work will be case studies run in the partners' organizations. The primary research processes are illustrated in the following figure and describe further in the text below:



Exploratory phase:

The project starts out with an exploratory phase focusing on both the obstacles inhibiting the transition to digital work processes and the possibilities and innovation potential that might be exploited after such a transition has taken place. The exploratory phase includes case studies for the partners (primarily requirements capture) and establishment of state of the art in related research and in product/service support. Requirements from the partners' environment are also collected (legal, compliance etc.).

Guidelines and specifications phase:

In this phase, a first version of guidelines (processes and organization) and specifications (for technical

solutions) are developed as a set of publicly available documents. Additionally, case studies will develop strategies for long-term digital documents (processes, organization, and technology) for the partners.

The specification activities continue until the end of the project, producing revised versions according to the progress and results of the project. Guidelines and specifications may be submitted to international standards bodies during the project's lifetime or by the end of the project, provided that resources can be allocated for the necessary participation and follow-up.

Implementation phase:

Following development of guidelines and specification, this phase aims at implementation of mechanisms to support products and services. The implementations must be aligned with the needs of the pilots to be run in the following phase. For the case studies, specifications for work processes will be developed.

Pilot phase:

Finally, selected aspects of long-term digital documents are piloted. The main focus for pilots will be real work processes for partner case studies, supported by technology. However, some piloting may be in the form of prototypes and demonstrators.

Results dissemination, patenting, commercialization, reference group:

Results dissemination is organised as a separate work package in LongRec. Four Ph.D. theses shall be produced. See section 15 for further details.

A series of partner workshops throughout the project will be an important arena for discussions and feedback of results, as well as supporting internal mobilization of resources among the project partners.

Search for existing, relevant patents will be done in the exploratory phase of LongRec. No search has been carried out during the preparation of the project proposal. Patenting of results will be considered by the commercialization partners during the course of the project.

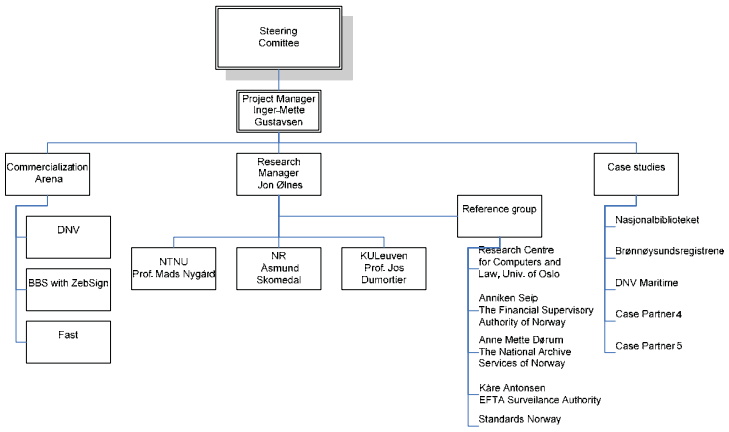
Liaisons to important, project external Norwegian actors are kept by establishing a reference group, which will be used for review of project results and will meet 3-4 times each year for workshop-type meetings providing feedback on the project's direction and progress.

5 Project organisation and management

The partners have been selected to be as complementary as possible and with a minimum of competition between them. The case partners also provide a breadth of business and problem domains to ensure validity and relevance of the results. The following organizations participate in the project:

Name	Partner role	Key competencies
Det Norske Veritas AS	Project owner and commercialisation	Information Quality Management (IQM) department, within both semantics and security. Primary research focus on research challenge 3, but will also do work across all the research challenges.
Nasjonalbib. (the National Library of Norway)	Partner/user	Preservation of digital information for the future.
Brønnøysundsregistrene (Brønnøysund Register	Partner/user	Maintenance of extensive, digital public records with high demands on availability and long-term

Centre		storage.
DNV Maritime Classification	Partner/user	Long-term records management of ship information.
Case partner 4	Partner/user	
Case partner 5	Partner/user	
BBS (with ZebSign)	Commercialisation	Technology solutions and services for storage and use of documents; security and PKI/signature solutions.
Fast Search & Transfer	Commercialisation	Integrated technology solutions based on best practice for information search and retrieval, analysis and use.
NTNU, Dep. Of Computer and Information Science	Research	Strong competence in storage, and in search, retrieval and use of documents. Primarily research challenges 1 and 2, but will also support as needed in other areas. Several research groups within the department will contribute, co-ordinated by Prof. Mads Nygård.
NR, Dep. Of Applied Research in Information Technology	Research	Security and trust, formal methods, privacy protection and DRM. Primarily research challenge 4, but will also participate in the work on the other challenges
Katholieke Universiteit Leuven, Interdisciplinary Centre for Law and Information Technology	Research	Prof. Jos Dumortier has worked within security and long-term management of digital documents, as well as with the legal challenges in this area. They also provide a link to the InterPARES 2 project.



The project has ongoing dialogue with several candidate partners to fill the slots for Case Partner 4 and 5.

The figure to the left shows the organization of the project.

6 International co-operation

Co-operation is established with the Interdisciplinary Centre for Law and Information Technology at the Katholieke Universiteit Leuven through Prof. Jos Dumortier. This Centre has also been a participant in the Canadian/American project InterPARES 2 [45] that concludes in 2006. The main

contributions will be related to research challenges 1, 2 and 4 and will be in the form of both short residences in Norway and workshop participation. Access will be provided to InterPARES 2 results

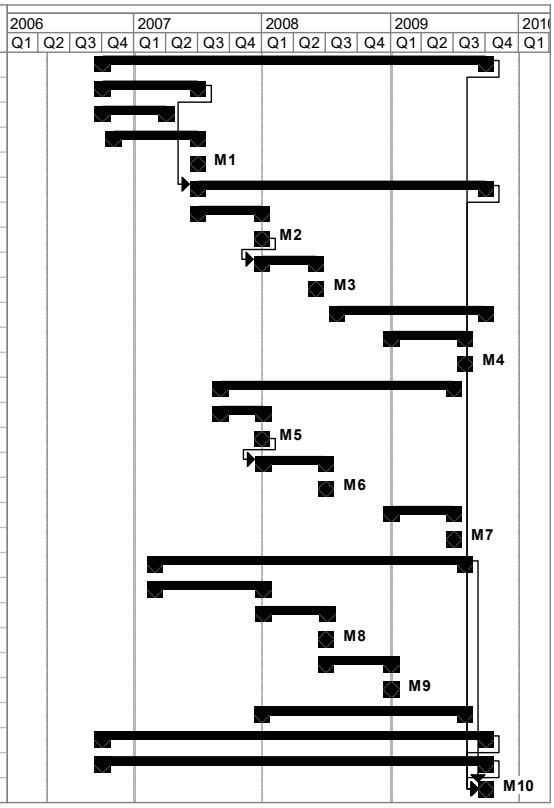
DNV has for several years worked with the European Space Agency (ESA) and is currently involved in project activities with ESA on concurrent engineering, and in this context long-term storage of information. The idea is to facilitate effective information sharing and collaborative, geographically distributed engineering through all phases of a project from conceptual design, though implementation, operation and maintenance. Safe and effective long term storage and interchange of information with partners and suppliers is vital in this connection. The relationship is with Massimo Bandecchi, Manager of ESA Concurrent Design Facility and it will be maintained by Narve Mjøs from DNV.

Fast has formalized research partnerships with: Professor Jamie Callan at Carnegie Mellon, Professor Lee Giles at Pennsylvania State University, Professor Mitsuru Ishizuka at Tokyo University, Professor Takenobu Tokunaga at Tokyo Institute of Technology, Professor Ruy Milifiu at PUC University Brazil, Professor Norbert Lossau at Bielefeld University and Professor Franz Guenthner at the University of Munich. Fast is also part of two EU FP6 projects and one application is under evaluation.

7 Progress plan – milestones

The project timeline is three years, with a planned start 1/10-2006 and finishing 30/9-2009. The following figure shows the overall project plan with the milestones for the main deliveries indicated.

ID	Task Name
1	**Project Management**
6	**Exploratory phase**
7	**State of the art in research and technology/service support**
9	**Partner exploration cases**
11	Case innovation potential described
12	**Specification phase**
13	**Development of best practices**
16	Best practice first draft completed
17	**Partner Long-term strategy development for each partner**
23	Suggested long-term strategies developed
24	**Partner long-term strategy implementation**
28	**Development of standards input**
30	Standards input completed
31	**Implementation phase**
32	**Mechanisms gap analysis**
36	Mechanisms gap analysis completed
37	**Development of mechanisms**
42	First version mechanisms ready for piloting
43	**Revision of Mechanisms**
46	Revised Mechanisms completed
47	**Pilot phase**
48	**Pilot identification and planning**
51	**Best practice piloting**
55	Best practice pilot completed
56	**Mechanism piloting**
61	Mechanism piloting complete
62	**Demontrators**
65	**Commercialization activities**
70	**Knowledge transfer and dissemination**
74	Project completed



The PhD students will have separate activity and delivery plans, which will be worked out during the project start-up phase and integrated into the overall project plan. The following table lists the primary deliverables for each milestone:

Milestone	Primary deliverables
M1	Case reports from partner cases identifying inhibitors, innovation potential and partner environment requirements
	State-of-the art in research and product/service support
M2	Best practice first draft
M3	Suggested long-term strategies for partners
M4	Guidelines and specifications (input to standardisation)
M5	Mechanisms gap analysis
M6	Mechanisms for piloting in partner cases
M7	Revised mechanisms
M8	Best-practice pilot evaluation report from partner cases
M9	Mechanisms piloting report from partner cases
M10	Final project report

8 Costs incurred by each project partner

The project costs are distributed among the partners as shown in the table below. (Note: Removed from this document as it is under revision.) The primary project cost aside from the partners' contributed time is the research done by research partners.

All necessary equipment needed for the project will be provided by the project partners. Other costs are mainly related to travelling in relation to partner workshops, meetings, conferences etc.

In the event that the slots for Case Partner 4 and 5 are not filled, one PhD will be removed and the amounts allotted to the research partners will be reduced.

9 Financial contribution by partner

Partners	Hour	Cash	TOTAL
Det Norske Veritas AS	1 500	1 500	3 000
Nasjonalbiblioteket	1 000	1 000	2 000
Brønnøysundsregistrene	1 000	1 000	2 000
DNV Maritime Classification	1 000	0	1 000
Case Partner 4	1 000	1 000	2 000
Case Partner 5	1 000	1 000	2 000
BBS/Zebsign	1 500	1 500	3 000
Fast	1 500	1 500	3 000
NTNU	400		400
Forskningsrådet (33%)		9 200	9 200
TOTAL	**9 900**	**17 700**	**27 600**

Each of the commercialisation partners contributes 3 mill NOK and most case partners 2 mill NOK to the project.

Funding from Forskningsrådet is 33% of the total project budget. (Note: Funding from Forskningsrådet has been granted as indicated in the table, provided that Case Partners 4 and 5 are identified and signed up.)

PART 2: Exploitation of results

10 Underlying idea

Work processes, both within a company and across company borders, increasingly are digital processes relying on digital documents. Solutions exist today that support such work processes, but these solutions do not solve the long-term (decades) perspective. As more and more work processes become digital the need for long-term storage becomes critical.

By establishing a scientific knowledge base for handling the long-term perspective, it becomes viable to build products and services for the long-term perspective. As illustrated by the research challenges,

there are many unresolved issues in the area, even if closely related areas such as digital preservation have many results that can be built upon.

By offering trusted, long-term storage services the service provider takes on a great responsibility to provide services for an unknown future. In order to make this a viable business it is absolutely essential to be able to build on sound, scientifically established standards and methods that are also supported by technology. Otherwise the business risk would be far too great and/or the service would not be trustworthy for the customers.

11 Innovation/degree of novelty

LongRec seeks to build a foundation for technology and services that have a much richer service portfolio than mere storage of a file for x years. The latter is possible today; BBS (LongRec project partner) delivers storage for decades for files in a defined format. This leaves all the challenges of actually understanding and evaluating the information to the customer. By addressing the challenges on a larger scale, much more advanced services can be offered to customers.

LongRec seeks to provide support for living records that are in active use over decades, with preservation of semantics (understanding) and evidential value (trustworthiness) of the contents. The aim is not mere preservation, but business support. A lifespan of decades, e.g. documentation for physical objects with such a lifetime, implies a need to be able to manage transitions in many areas, such as technology, organization, people and ownership.

The LongRec partners are not aware of any actor providing technology or services in this area today, and surprisingly few research initiatives address the issues. This implies a high degree of novelty combined with significant innovation potential, based on the competence of the partners.

12 Plan for exploitation of research results in the individual company

Note: Removed from this version of the document.

13 Environmental impact

No direct environmental consequences are foreseen from this project. It can be argued that more use of digital storage will reduce the need for paper based storage, but the environmental consequences of creating and operating the needed digital storages are not known to the participants in LongRec.

14 Other value

The long-term aspects of digital documents have not received significant attention from the research community in Norway apart from initiatives in the digital preservation area. LongRec seeks to rectify this and at the same time build on the strengths already present in the research partners.

The general problem of handling long-term storage of live records is to a large extent ignored by organizations today. Many believe that technology will solve this “in the future”, but not many suppliers of technology focus on how documents can survive their technology. There are no complete solutions today to the problems addressed by LongRec.

15 Information and dissemination of results

As a research project with high ambitions, LongRec aims at extensive publication in reviewed international journals and conferences. A Ph.D. thesis must be expected to generate 3-5 publications and researchers taking part in LongRec are expected to produce further papers. No publication is expected in 2006 as the project merely starts this year. Early papers (2007) will typically be of a survey

type reflecting the exploratory phase. Papers in 2008 to 2010 will present results advancing the state of the art in the research areas, and additionally papers presenting empirical results from the case studies (individual cases or generalised) should be written. The year 2010 is included in the table below because some publication, notably journal papers with a long delay in the referee process, necessarily will be done after the completion of the project itself.

The split on journal and conference publications indicated below is only indicative. Selection of venue must be determined for each publication topic/paper. There are quite a few topical journals and conferences suitable for dissemination from LongRec, e.g. in document/information management and information security. No list is provided here.

Type of publication / Year	2007	2008	2009	2010
Academic publications in international, per-reviewed journals	1	4	5	5
Contribution to high-end academic conferences	3	4	4	2

References:

[1] Lyman, P; Varian, H. R. How Much Information? (2003). *http://www.sims.berkeley.edu:8000/research/projects/how-much-info-2003/execsum.htm*

[2] *Butler Group, (2005) Managing the Storage of Information*

[3] Abie,H; Skomedal, A (2005) A Conceptual Formal Framework for Developing and Maintaining Security-Critical Systems. *IJCSNS International Journal of Computer Science and Network Security, Vol.5, No.12, December 30, 2005*

[4] *http://www.cornwell.co.uk/moreq*

[5] Department of Defence. Design Criteria for Electronic Records Management Software. *US Department of Defence, 2002.* http://www.dtic.mil/whs/directives/corres/pdf/50152std_061902/p50152s.pdf

[6] http://www.records.nsw.gov.au/publicsector/DIRKS/final/introduction/introduction.htm

[7] NOARK-4 (1999) http://www.riksarkivet.no/noark-4/Noark-eng.pdf

[8] Kimberly J. Barata. (1998) The impact of information technology standards on recordkeeping systems development: Implications for records professionals. *Computer Standards & Interfaces, Vol. 19, 1 ,p 51-64*

[9] Oftedal, G; Dalhaug, K; Torstensen, A; Haukøy, S. (2005), Document & Records Management – evaluation of needs and solutions. *DNV Technical Report.*

[10] Gayorfar, H. (2006) Digital Document Preservation - Feasibility Study. *DNV Technical Report 2006-0514*

[11] Business Record Retention Guidelines MMA research and by Clark Hill

[12] Dublin Core Metadata Initiative 2005. http://xml.coverpages.org/ni2005-03-21-a.html

[13] W3C Recommendation, Extensible Markup Language (XML) 1.0 *(Second Edition), World Wide Web Consortium (2000) http://www.w3c.org/TR/REC-xml*

[14] Harvey C. C; Huc, C. (2003) Future trends in data archiving and analysis tools. *Advances in Space Research, Vol. 32, 3, p. 347-353*

[15] http://www.w3.org

[16] Abie, H; Aredo, D. B; Kristoffersen, T; Mazaher, S; Raguin, T. (2004) Integrating a Security Requirement Language with UML. I*n the Proc. of the Seventh International Conference on UML Modelling Languages and Applications – UML 2004 - October 11-15, 2004, Lisbon, Portugal, LNCS 3273, ISBN: 3-540-23307-5, Springer Berlin*

[17] Manola, F, Miller, E. (2004) RDF Primer, W3C Recommendation http://www.w3.org/TR/rdf-primer

[18] McGuinness, D. L. (2004) Question Answering on the Semantic Web. *IEEE Intelligent Systems, vol. 19, 1*

[19] Brewster, C; O'Hara, K; Fuller, S; Wilks, Y; Franconi, E; Musen, M.A; Ellman, J; Buckingham S. (2004) Knowledge Representation with Ontologies: *The Present and Future. IEEE Intelligent Systems, vol. 19, 1*

[20] Nørvåg, K; Overskeid Nybø, A. (2006). DyST: Dynamic and Scalable Temporal Text Indexing. T*o appear in Proceedings of the 13th International Symposium on Temporal Representation and Reasoning (TIME'2006), Budapest, Hungary, June 2006.*

[21] Hodge, G. (1999) Best Practices for Digital Archiving

[22] Boudrez, F., Dekeyser, H. and Dumortier, J., Digital Archiving: The new challenge?, Louvain-La-Neuve, *I.R.I.S., 2005, pp. 83-84*

[23] Lee, K.-H.; Slattery, O.; LU, R.; TANG, X.; McCrary, V. (2002). The state of the art and practice in digital preservation. *J. Res. NIST 107, 93–106*. http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf

[24] Rothenberg, J; Bikson, T. (1999) Digital Preservation: Carrying Authentic, Understandable and

Usable Records through Time.
[25] Where We Are Now: Obliquity and Squint? *The InterPARES report, Preservation Strategies for Electronic Records, Round 1 (2000-2001)*
[26] Granger, S. (2000) Emulation as a digital preservation strategy. *D-Lib Mag. 6, 10.* *http://www.dlib.org/dlib/october00/granger/10granger.html*
[27] Digital Preservation Testbed White Paper: Emulation: Context and Current Status (2003). *http://www.digitaleduurzaamheid.nl/bibliotheek/docs/White_paper_emulation_UK.pdf*
[28] http://www.bbs.no
[29] Paul Wheatley,– a CAMiLEON discussion paper(2001) *http://www.personal.leeds.ac.uk/~issprw/camileon/migration.htm*
[30] Wang, X; Feng, D; Lai, X; Yu, H. (2004). Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD" http://eprint.iacr.org/2004/199
[31] Ølnes, J. and Seip, A. (2002) On Long-Term Storage of Digitally Signed Documents. *Proceedings of the Second IFIP Conference on e-Commerce, e-Business, e-Government (I3E), Kluwer Academic Publishers*
[32] Dekeyser, H. (2005) Preservation of Signed Electronic Records, *DLM Conference.* http://europa.eu.int/comm/secretariat_general/edoc_management/dlm_forum/doc/11bis_Dekeyser.pdf
[33] Electronic Signatures and Infrastructures (ESI); Electronic Signature Formats, versjon 1.5.1, *ETSI TS 101 733* (2003)
[34] XML Advanced Electronic Signatures (XAdES), versjon 1.2.2, *ETSI TS 101 903* (2004)
[35] Danielson, J; Groven, A. K; Kristoffersen, T; Rivertz, H. J; Skomedal, A. (2005) Elektroniske spor, *NR Report 1008, ISBN: 978-82-53-90516-7*
[36] Kenneally, E. E. (2004) *Digital Investigation, Vol. 1, 2, p 94-101*
[37] Abie, H; Spilling, P; Foyn, B. (2004) A Distributed Digital Rights Management Model for Secure Information Distribution Systems, *International Journal of Information Security.*
[38] Information and Documentation – Records Management – Part 1: General and Part 2: Guidelines, *ISO 15489-1:2001 and ISO/TR 15489-2:2001* (2001)
[39] Space Data and Information Transfer Systems – Open Archival Information System – Reference Model, *ISO 14721:2003(E)* (2003)
[40] OASIS UBL http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl
[41] ISO 15926 http://www.posccaesar.com/ISO15926/iso15926.htm
[42] Ontology forum http://ontolog.cim3.net/cgi-bin/wiki.pl/
[43] Web Ontology Working Group of W3C http://www.w3.org/2001/sw/WebOnt/
[44] Ølnes, J. (2006) PKI Interoperability by an Independent, Trusted Validation Authority, *5th Annual PKI R&D Workshop* http://middleware.internet2.edu/pki06/
[45] The InterPARES 2 project at http://www.interpares.org/ip2/ip2_index.cfm