

«Pérennisation de l'information numérique»

**Les changements spectaculaires du paysage
national et du contexte européen.**

Jeudi 10 avril 2008

Organisation : Aristote groupe PIN

Coordination scientifique :
– *Claude Huc (Aristote PIN).*

Amphithéâtre Becquerel, École Polytechnique, Palaiseau

<http://www.aristote.asso.fr>

Contact : info@aristote.asso.fr

Edition du 22 pluviôse an CCXVII (*vulg.* 10 février 2009) ©2009 Aristote

Table des matières

1	Programme de la journée	5
1.1	Introduction	5
1.2	Programme	6
2	Présentations	7
2.1	Towards a European/Global infrastructure to support digital preservation	7
2.2	PILAE, plate-forme pilote d'archivage électronique	17
2.3	SPAR (système de préservation et d'archivage réparti)	26
2.4	PAC, plate-forme d'archivage du CINES	32
2.5	Etat d'avancement du standard NAS – EN 9300 LOTAR	36
2.6	Un contexte normatif très enrichi	40
2.7	L'évolution de l'offre des tiers archiveurs en réponse à l'évolution des besoins	47
2.8	PILAE : vers une offre complète en composants réutilisables pour l'archivage	52
2.9	SPAR : l'analyse et les choix majeurs du système logiciel	56

Chapitre 1

Programme de la journée

1.1 Introduction

Pendant une période d'une quinzaine d'années, la question de la conservation à long terme de l'information sous forme numérique n'a été un sujet de préoccupation et de réflexion que pour de rares institutions scientifiques ou patrimoniales qui ont joué un rôle moteur dans la prise de conscience des risques encourus et dans l'émergence de normes de référence et de méthodologies dans ce domaine. La progression constante et inexorable du numérique dans tous les domaines avec pour conséquence l'élimination progressive des technologies qui lui précédaient ont accéléré cette prise de conscience jusqu'au plus haut niveau de la hiérarchie des entreprises et des administrations. De ce fait, d'importantes évolutions du Droit ont permis la reconnaissance, du point de vue juridique, du support numérique.

Les années 2005 à 2007 ont été le théâtre de l'émergence d'une série de projets, tant dans la sphère publique que dans la sphère privée, tant au niveau national qu'au plan européen, visant à développer des infrastructures matérielles et logicielles dédiées à la conservation de l'information et à la mise en place de véritables services d'archivage long terme des informations sous forme numérique. Un enrichissement de l'offre commerciale mais aussi l'émergence des logiciels libres accompagne ce mouvement dans plusieurs domaines : archivage et records management, mais également stockage et plusieurs secteurs d'activité (entreprise, documentaire, scientifique, patrimoine). Un état des lieux pourra être dressé. Il pourra être complété, au cours de la table ronde, par un examen des nouvelles questions qui s'ouvrent en matière de certification des Archives.

1.2 Programme

9h15-9h30	Claude Huc Groupe PIN, Aristote	Introduction
9h30-10h25	Seamus Ross W Director HATII University of Glasgow	Towards a European/Global infrastructure to support digital preservation
10h25-10h55	Françoise Banat-Berger Direction des Archives de France	PILAE, plate-forme pilote d'archivage électronique : le contexte du réseau public des archives.
10h55-11h10	<i>Pause café</i>	
11h10-11h40	Thomas Ledoux Bibliothèque nationale de France	SPAR (système de préservation et d'archivage réparti) : genèse et démarche au sein de BNF
11h40-12h20	Olivier Rouchon CINES	PAC, plate-forme d'archivage du CINES : enjeux et perspectives pour l'information scientifique et technique
12h20-12h40	Pierre Duchier Airbus	Etat d'avancement du standard NAS-EN 9300 LOTAR Long-term archival and retrieval of aerospace product information
12h40-14h00	<i>Déjeuner</i>	
14h00-14h40	Marie-Anne Chabin Société Archive17	Un contexte normatif très enrichi : Moreq2 à la rencontre du Modèle OAIS
14h40-15h10	Vincent Jamin Société ORSID	L'évolution de l'offre des tiers archiveurs en réponse à l'évolution des besoins
15h10-15h40	Alain Borghesi et Bruno Ricci Société Cecurity.com	PILAE : vers une offre complète en composants réutilisables pour l'archivage
15h40-15h55	<i>Pause</i>	
15h55-16h25	Charlotte Fabre, Laurent Sollier Société ATOS Origin	SPAR : l'analyse et les choix majeurs du système logiciel
16h25-17h15		Table ronde et conclusions

Chapitre 2

Présentations

2.1 Towards a European/Global infrastructure to support digital preservation

Prof. Seamus Ross (Director HATII, University of Glasgow)

Towards a European/Global Infrastructure to Support Digital Preservation

Pérennisation de l'information numérique les changements spectaculaires du paysage national et du contexte européen

10 Avril 2008

Seminaire Aristote
Ecole Polytechnique (Palaiseau)

Prof Seamus Ross
Director HATII, University of Glasgow

1

!!!WOW!!!!

"A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing"
DGiaretta

• In 2006, the amount of digital information created, captured, and replicated worldwide was 161 exabytes (161 billion gigabytes) - roughly 3 million times the information in all the books ever written!

Between 2006 and 2010, the information added annually to the digital universe will increase more than six fold from 161 exabytes to 988 exabytes

IDC (2007) The Expanding Digital Universe

Objectives of digital longevity



- Digital preservation aims to ensure that future users will be able discover, retrieve, render, manipulate, interpret and use digital information in the face of constantly changing technology
- It involves conservation, renewal, restoration, selection, destruction, enhancing, updating, and annotating
- It is a risk management activity at all stages of the longevity pathway -- translating uncertainties into manageable risks
- Digital Preservation is an ongoing activity to ensure recurring value of digital objects.



Charles Dollar visits HATII, 2004

3

Preservation Fracture Points

- Loss/inaccessibility of "the bits"
 - The mechanisms to identify the location of data, for example a DNS entry pointing to a host machine, may no longer be resolvable.
- Failure of any chain of preservation
 - Hardware (including media), software or support environment.
 - Custodian of the data, an organisation or project may cease to exist.
 - changing standards – loss of "standardisation" and commonality
- Changes in the Knowledge Base of the Designated Community
 - Loss of understandability/usability/intelligibility/performance...
- Loss in the chain of evidence
- Lack of certainty of provenance or authenticity.

4

An Agitating Buzz about Things Digital



- We have successfully socially amplified the perception of risks associated with digital materials
- While there are 'risk amplifiers' and 'risk attenuators', perceptions of risk arise complex social and cultural processes
- Have we socially constructed our perceptions of preservation risk within a vacuum

5

We need to make a radical change



- We need to stop 'building that agitating buzz'
- We need to deliver actual solutions that work.
- We need solutions that scale.

6

Preservation Risk is Actual



- It is technological.
- It is social.
- It is organisational.
- And it is cultural.
- Actual risks can be assessed and measured—actual risks can be managed.

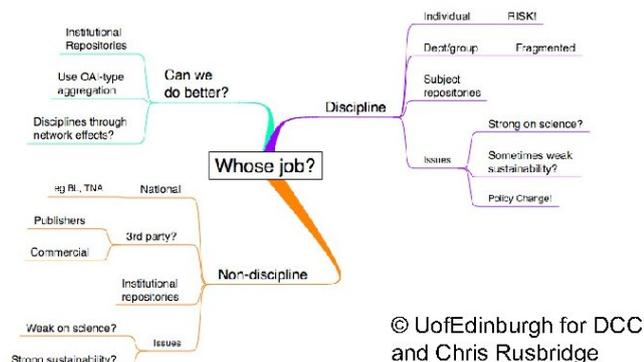
7

A Sad Tale, the AHDS

- Arts and Humanities Data Service
 - 12 year old UK Agency promoting preservation practices and providing preservation services for the Arts and Humanities.
 - Internationally recognised model for activity in this area.
- Arts and Humanities Research Council of the UK, the primary funder of the AHDS voted in March 2007 to withdraw funding from end of March 2008.
 - “Council believes that long term storage of digital materials and sustainability is best dealt with by an active engagement with HEIs rather than through a centralised service

8

Who are the curation players?



9

DPE Research Roadmap



- Analysing the state of the art in Digital Preservation research and existing research agendas.
- Redefinition and refocusing of the Preservation research agenda
- Basis for development of research problem basis
- Provides a foundation for communication about research needs.

10

Reviewed Published Research Agendas from 1991 to 2006



- **UEI** – *Understanding Electronic Incunabula: A Framework for Research on Electronic Records* [9] by Margret Hedstrom, 1991.
- **PDI** – *Preserving Digital Information* [17], edited by John Garrett and Donald Waters, 1996.
- **DPNU** – *An Investigation into the Digital Preservation Needs of Universities and Research Funders* [15] by Denise Lievesley and Simon Jones, 1998.
- **SoDP** – *The State of Digital Preservation - An International Perspective* [5] contains articles by various authors, 2002.
- **IAT** – *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation* [3] was published by the NSF in 2003.
- **I2S** – *Invest to Save* [4] was prepared for the NSF-DELOS working group on digital archiving and preservation in 2003.
- **eScience** – *e-Science Curation Report* [16] by Philip Lord and Alison McDonald was published in 2003.
- **Cyber** – *Revolutionizing Science and Engineering Through Cyberinfrastructure* [18] was created by the Blue-Ribbon Advisory Panel on Cyberinfrastructure of the NSF in 2003.
- **DigiCULT** – *The Future Digital Heritage Space: An Expedition Report* [6] was published as a DigiCULT thematic issue in 2004.
- **Epanet** – *Electronic Resource Preservation and Access Network* [11] was a European Commission funded project which ran from 2001 until 2004.
- **Warwick** – *Digital Curation and Preservation: Defining the research agenda for the next decade* [1] reports on the Warwick workshop held in 2005.
- **DRR** – *Digital Repositories Roadmap - Looking Forward* [29] by Rachel Heery and Andy Powell, 2006.

11

Review Conducted Segmented by



- Digital Object Level
- Collection Level
- Repository Level
- Process Level
- Organisational Environment

12

The Landscape



- lack of common understanding
- loss of focus
- lack of practical experience
- fragmentation
- frictional losses
- lack of training

13

Refocused Landscape



- Restoration
- Conservation
- Management
- Risk
- Significant Properties of Digital Objects
- Interoperability
- Automation
- Context
- Storage
- Experimentation

14

Out of Scope, but not out of mind



- Research also needed in areas of
 - policy and procedures,
 - organisational structure and communication,
 - education,
 - business case development, or
 - legal arena.



© HARTILL Foundation, 2003

15

www.alliancepermanentaccess.eu

16

- Taskforce on Permanent Access to the Records of Science recommended its foundation.
- Alliance construing a sustainable business model to ensure long term viability.
- An Alliance Working Group did initial planning and brought together a group of catalysing members.
- Brussels Conference November 07 formal public statement of intent.
- Successful bid for PARSE.Insight.

17

Whose in the Sandbox:

- **Funding agencies**
 - European Science Foundation
- **Key European Research Organisations**
 - CERN, European Space Agency,
 - Max Planck Gesellschaft/Max Planck Digital Library,
 - Science and Technology Facilities Council (STFC)
- **National Archives and Libraries**
 - Swedish National Archives
 - British Library,
 - Koninklijke Bibliotheek,
 - Deutsche Nationalbibliothek
- **Associated National 'coalitions'**
 - Digital Preservation Coalition
 - Nestor
 - Perennisation des Informations Numeriques (PIN)
- **Publishers**
 - International Association of Scientific, Technical and Medical Publishers

18

What is the Alliance doing?

- Foster consensus towards a European Digital Information Infrastructure (EDII), starting with the records of science;
- Push forward the creation of components for the EDII;
- Collaborate with commercial enterprises, centres of competence, coalitions, and international partners.
- Act as a catalysing force for national governments and the EU;
- strengthen role of European participation international preservation and access initiatives.
- build, articulate and maintain continuing R&D programme.
- Provide a launch-pad and incubator for new projects

[- Alliance for Permanent Access -] ¹⁹

Alliance Work Programme

- Aims to consolidate the Alliance Programme of Activity
- Working within Communities
- Coordination and stimulation of activity
- Community building in Europe and beyond
- Stimulation and co-ordination of R&D
- Engage the community to make the Alliance Sustainable.

20

PARSE.Insight

- Bid to EU unit which funds GEANT and EGEE
 - 1MEuro, 9 partners, 2 years
- Designed to help them direct their e-Sci funding in 2009/2010
- Will produce
 - Roadmap for Infrastructure to support preservation of data
 - Survey of what is already there and what is planned
 - Gap Analysis
 - Impact Analysis tool
 - Sustainability study.

21

WHAT will the EDII look like?

According to the Report of the Task Force on Permanent Access (2005) it will:

- ✓ Identify **core physical digital archives/repositories**
- ✓ **OAIS-compliant** to ensure proper archiving, interoperability and long-term preservation
- ✓ Framework for **metadata**, Framework for **persistent identifiers**, and number of **registries**, possibly **other standards**
- ✓ Cost-effective **preservation methods and services** must be available
- ✓ Common framework of **principles and guidelines for management of access and rights** (underlying the technical tools to implement this framework)
- ✓ Create **Financial mechanism** for developing and testing implementation tools, techniques and services, and for strengthening collaboration and training
 - ✓ a. **Certification service providers**, accredited according to
 - ✓ b. Common **European accreditation mechanism**.

[- Alliance for Permanent Access -] ²²

Competence Centres.



- DPE survey of Competence Centres in Europe
- Assessment of the character of Competence Centres.
- Definition of guidance on how to benchmark Competence Centres (Capacity, Context, Credibility, Commitment, Certification, Competition, Communication).

23

Why Establish Competence Centres



- The main objectives of the federation of competence centres should be to:
 - Establish a vibrant research programme that is informed by the requirements and experiences of a range of stakeholder communities
 - Nurture strong community relationships from a range of disparate stakeholders
 - Work with user communities and technology providers to develop tools and resources as well as relevant and valuable services
 - Achieve the 'virtuous circle' whereby expertise, experience and user requirements inform the federation's research and development activity

24

Our Approach to Competence Centres



- a federated approach to the provision of support and guidance
- a life-cycle approach to the provision of support and guidance
- better coordination of disparate competence centres' activities
- improved integration of competence centres with industry
- a more collaborative approach to the provision of training and outreach activities
- increased research capacity within competence centres to help push forward the international research agenda
- the introduction of competition between competence centres to drive performance
- the investigation of business models that will help to provide sustainable funding for competence centres' activity.

25

What is DPE?



- *Digital Preservation Europe (DPE)* fosters collaboration & synergies between existing national initiatives across the European Research Area.
- *DPE* addresses the need to improve coordination, cooperation, and consistency in current activities to secure the longevity digital materials.

26

Collaborative Arrangements



- Established collaborative working arrangements with international activities and with European Projects.
- Organised workshop in Nice (FR) to facilitate training collaboration across Europe.
- Promoting collaborative action with FP6 Call 5 Projects
 - <http://www.wepreserve.eu>

27

Training and Continuing Professional Development



- Creation of a Trainer and Training Materials Repository
- Coordination Framework for Training and Education
- Co-ordinate and Deliver DPE Training Programme
- Foster Collaboration on Training between DPE, CASPAR, and Planets

28

DCC

- DCC partners
 - STFC,
 - U Glasgow,
 - U Bath,
 - U Edinburgh

29

OAIS (ISO14721)

- Open Archival Information System Reference Model
 - referenced in just about any serious work on digital preservation
 - Development hosted by CCSDS Panel 2
- 5 year ISO review underway
 - minor corrections and updates
 - No major changes
- Revised version due early 2008

30

Authenticity

- Traditionally archivists have been extremely concerned about authenticity
 - Central to any “archivist” discussion about preservation
- InterPARES project:
 - “When we refer to an electronic record, we consider it essentially intact and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered.”
 - Document, business process oriented
 - Does this apply to data? [What is “the” message intent of data?](#)
- Maintaining authenticity
 - Technical aspects [hash codes etc](#)
 - Social aspects [do I trust him/her?](#)
- Perhaps not so important for science data in the short term but will become increasingly important over time

31

Audit & Certification

- Need for a way to judge digital repositories has been demanded for > 10 years
- RLG/NARA produced TRAC (Trusted Repository Audit Checklist) (DG member of working group)
- Follow on work in CCSDS aims to lead to ISO standard on which an accreditation and certification process can be based
- Such a certification process could shape digital “market” in future

32

Links

- Digital Curation Centre
 - <http://www.dcc.ac.uk>
- RAC group Wiki:
 - <http://wiki.digitalrepositoryauditandcertification.org>
- TRAC document
 - <http://www.crl.edu/PDF/trac.pdf>
- CASPAR project
 - EU project on digital preservation – Science, Culture and Arts data
 - Infrastructure, tools and detailed case studies – what does one need to actually “understand” the data?
 - <http://www.casparpreserves.eu>

33

ISO process status

- New group set up with the primary aim of producing an ISO standard
 - Repository Audit and Certification (RAC)
- OPEN process
 - Wiki open to all
 - Mailing list open to all
 - Virtual meetings normally every week
 - See <http://wiki.digitalrepositoryauditandcertification.org>
- Into ISO via CCSDS – same route as OAI
- Some organisational/procedural changes in CCSDS
- Currently a Birds of a Feather (BoF) group
 - To demonstrate adequate support for the work
- Subsequently should become a Working Group
- Documents agreed by the WG will then be reviewed by CCSDS and more broadly via international ISO review process

34

Current status

- Reviewing and comparing
 - TRAC
 - NESTOR
 - DCC documents
- Do we need another ISO standard?
 - Could we simply add to existing standards e.g. ISO 27001
 - The view is that ISO 27001 CANNOT be modified adequately
 - It’s view of Information is too limited
- Started drafting a straw man document
 - Taking TRAC and add concepts from other docs

35

Trusted Digital Repositories

- Invited group, hosted by Research Library Group (RLG)
- Concerned with organisational and financial issues
- Trusted Digital Repositories: Attributes and Responsibilities (TDR)
 - <http://www.rlg.org/legacy/longterm/repositories.pdf>

36

Audit and Certification Services



- Collaboration Agreements with DCC and
- Define audit and certification practices
- Create Self-Audit Toolkit
- Piloting of the DPE Audit and Certification Scheme
- Certification Result Repository
- Repository Validation Scheme

37

Participation in Audit and Certification International Efforts



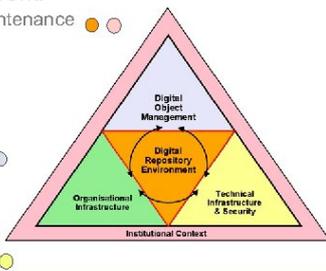
- DPE contributed to the final definition of the Trustworthy Repositories Audit and Certification Checklist (released by RLG/OCLC/CRL)
- Developed collaborations with the Center for Research Libraries (Chicago), and Nestor
- Developed ten core principles that can be used to characterise a digital repository.
- Delivered in collaboration with DCC *The DRAMBORA Toolkit*

38

Repository Environments



- Ten principles conceived for Digital Repositories
- An intellectual context for the work:
 - Commitment to digital object maintenance
 - Organisational fitness
 - Legal & regulatory legitimacy
 - Effective & efficient policies
 - Acquisition & ingest criteria
 - Integrity, authenticity & usability
 - Provenance
 - Dissemination
 - Preservation planning & action
 - Adequate technical infrastructure



Objectives



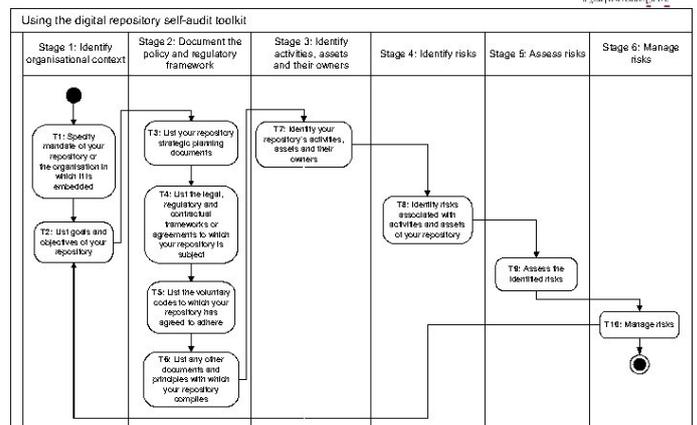
- The purpose of the DRAMBORA toolkit is to facilitate the auditor in:
 - defining the mandate and scope of functions of the repository
 - identifying the activities and assets of the repository
 - identifying the risks and vulnerabilities associated with the mandate, activities and assets
 - assessing and calculating the risks
 - defining risk management measures
 - reporting on the self-audit

Benefits of DRAMBORA



- Following the successful completion of the self-audit, organisations can expect to have:
 - Established a comprehensive and documented self-awareness of their mission, aims and objectives, and of intrinsic activities and assets
 - Constructed a detailed catalogue of pertinent risks, categorised according to type and inter-risk relationships
 - Created an internal understanding of the successes and shortcomings of the organisation
 - Prepared the organisation for subsequent external audit

DRAMBORA Workflow



Audit and Certification Services



- Collaboration Agreements with DCC and
- Define audit and certification practices
- Create Self-Audit Toolkit
- Piloting of the DPE Audit and Certification Scheme
- Certification Result Repository
- Repository Validation Scheme

37

Participation in Audit and Certification International Efforts



- DPE contributed to the final definition of the Trustworthy Repositories Audit and Certification Checklist (released by RLG/OCLC/CRL)
- Developed collaborations with the Center for Research Libraries (Chicago), and Nestor
- Developed ten core principles that can be used to characterise a digital repository.
- Delivered in collaboration with DCC *The DRAMBORA Toolkit*

38

DRAMBORA Interactive



43

Research Roadmap



- Analysing the state of the art in Digital Preservation research and existing research agendas.
- Redefinition and refocusing of the Preservation research agenda
- Basis for development of research problem basis
- Provides a foundation for communication about research needs.

44

Refocused Landscape



- Restoration
- Conservation
- Management
- Risk
- Significant Properties of Digital Objects
- Interoperability
- Automation
- Context
- Storage
- Experimentation

45

Experimental Testbeds

- integrate, automate, and evaluate a framework for digital entity preservation by integrating and combining the testbed framework and evaluation metrics
- tools to automate selected steps of the preservation process, such as ingest validation, preservation experiment set-up and control, preservation criteria definition, and verification of formal transformation, to support semi-automatic alternative evaluation.
- to investigate the potential metrics for measuring the effectiveness of different preservation strategies in the context of complex digital objects
- integration of software tools to support the digital preservation testbed framework.

46

CASPAR Consortium

EU FP6 Integrated Project

Total spend approx. 16MEuro (8.8 MEuro from EU)

Started April 2006, for 42 months

David Giarretta is Co-ordinator



<http://www.casparpreserves.eu>

High Level Objectives

The guiding principle of CASPAR is the application of the OAIS Reference Model

- to research, develop and integrate advanced components to be used in a wide range of preservation activities.
- to create the CASPAR framework:
 - the software platform that enables the building of services and applications that can be adapted to multiple areas

The CASPAR consortium will seek to guarantee the future evolution of CASPAR. This ambitious goal will be pursued through:

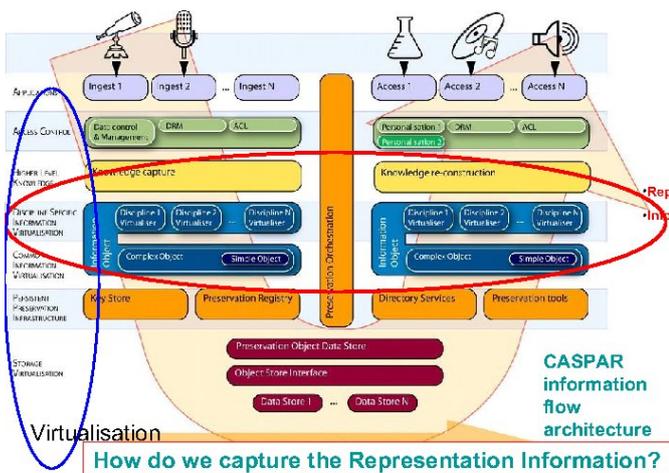
- the building of the CASPAR preservation user community
 - creating consensus around the initiative and gathering a critical mass of potential users
- embedding the CASPAR framework and components within key memory organisations, both national and international.

CASPAR Aims

- Produce tools and techniques to support digital preservation and make it easier to share the cost
 - must be relatively easy to use
 - must have a low “buy-in” in terms of effort required for adoption
 - must avoid requiring wholesale change of everyone else’s systems
 - must be decentralised and reproducible so that it can live on after the formal end of the CASPAR project
 - must be “preservable”
 - must be open: open source, open standards
- Cannot do everything
 - Working closely with other projects

CASPAR proposed validation metrics

- Demonstrate theoretical basis
- “Accelerated lifetime” tests
 - Changes in hardware
 - Changes in environment
 - Changes in Designated Community
- Demonstrate increased trustworthiness
 - Measured using Certification process (as/when available)



DPEX Exchange Programme.



- The DPE Research and Industrial Exchange Programme has been launched
- It promotes:
 - communication
 - collaboration between digital preservation research groups.

2.2 PILAE, plate-forme pilote d'archivage électronique

Françoise Banat-Berger (Direction des Archives de France)

Le contexte du réseau public des archives.

Péremniation de l'information numérique

PILAE, plate-forme pilote d'archivage électronique : le contexte du réseau public des archives

Françoise Banat-Berger
Direction des Archives de France



Le pilote PIL@E

Le contexte : l'e-administration



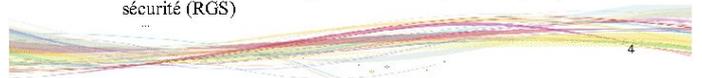
L'administration électronique

- Portée actuellement par la direction générale de la modernisation de l'Etat (DGME) et plus précisément par le service pour le développement de l'administration électronique, direction du MINEFE à vocation interministérielle
- Un schéma directeur encadrant les différentes actions entreprises dans les différents secteurs en vue de dématérialiser des processus métier



L'administration électronique

- Dans le cadre d'une évolution législative et réglementaire importante depuis la loi du 13 mars 2000 qui renouvelle le droit de la preuve
 - Une même valeur de preuve est reconnu à l'écrit sur support papier et à l'écrit sur support électronique à condition qu'on puisse précisément identifier l'auteur du document et que ce dernier soit conservé dans des conditions susceptibles de garantir son intégrité
- Jusqu'à l'ordonnance du 8 décembre 2005
 - Concerne les autorités administratives aux niveaux central et local
 - Donne un fondement juridique aux téléservices et téléprocédures
 - Introduit les référentiels généraux d'interopérabilité (RGI) et de sécurité (RGS)



- « Services Intégrés » : Particuliers, Entreprises, Collectivités, Associations
- « Services Sectoriels » : Appui aux schémas directeurs ministériels
- « Fonctions Soutien » : Finances, Aide à la décision, Ressources Humaines
- « Fonctionnel Transverse » : Identité numérique, Archivage Numérique, Accueil, Production et gestion de contenu, Information géographique
- « Construction des SI » : Développement informatique, Systèmes d'échanges, Echanges de données dématérialisés, Pilotage et données
- « Infrastructures » : Postes terminaux, Certificats et cartes, Télécom & réseaux, Production



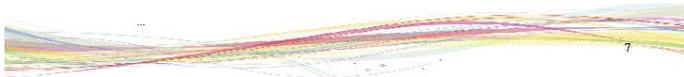
L'initiative IF-08

- Ses objectifs
 - sensibiliser les acteurs du processus d'archivage : producteurs des données, services informatiques, records manager, archivistes
 - élaborer des référentiels
 - appuyer la création de plates-formes d'archivage



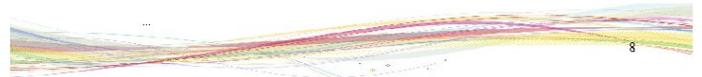
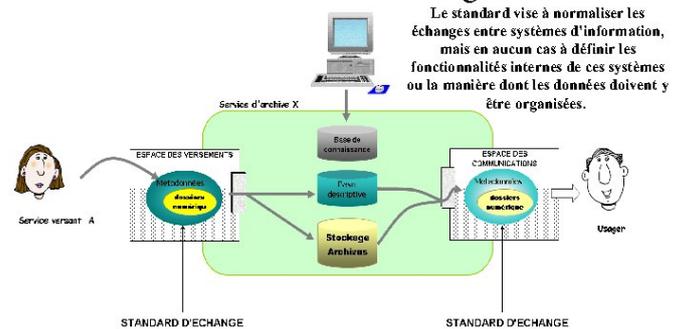
L 'initiative IF-08 et ses référentiels

- Étude en partenariat avec la direction centrale de la sécurité des systèmes d 'information (DCSSI) (cabinet Caprioli et Associés)
- Publication du référentiel : juillet 2006 (<http://www.ssi.govv.fr/fr/confiance/archivage.html>)
- Les livrables :
 - une politique d 'archivage type pour le secteur public (particularités du secteur public avec les missions et attributions réglementaires des services d 'archives prises en compte) - à décliner ensuite sur place pour une déclaration des politiques d 'archivage (le comment)
 - une prise en compte non seulement de la sécurité (réducteur) mais également du métier (fonctionnalités)



7

Le standard d 'échange



8

L 'initiative IF-08 et ses référentiels : le standard d 'échange

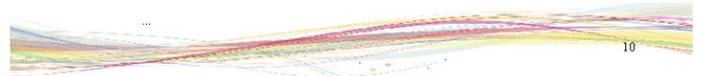
- Il s 'agit par conséquent avant tout d 'un format visant à faciliter les transferts entre un service et un autre à des fins de prise en charge pour l 'archivage
- Un format de métadonnées orienté transferts
- Soit une modélisation du bordereau de versement traditionnel et de la façon dont les métadonnées (le bordereau) et les données sont encapsulées dans une enveloppe XML au moment du transfert
- A leur arrivée dans le service d 'archives, les métadonnées sont intégrées dans la base de données descriptive des archives (qui permet de retrouver les archives de multiples producteurs de données) et les données sont écrites sur les espaces de stockage sécurisés



9

L 'initiative IF-08 et ses référentiels : le standard d 'échange

- Le standard d'échange de données pour l'archivage est destiné à être mis en œuvre par :
 - les producteurs d'archives ;
 - les éditeurs de logiciels sectoriels (gestion du personnel, finances, gestion sociale, gestion d'équipements, messagerie, etc.), dont les outils doivent pouvoir exporter automatiquement des données à archiver conformes au standard ;
 - les services d'archives, publics ou privés ;
 - les tiers-archivistes ;
 - les éditeurs de logiciels de gestion d'archives, dont les outils doivent notamment pouvoir accueillir automatiquement des données versées conformes au standard.



10

L 'initiative IF-08 et ses référentiels

- Ces référentiels sont d 'ores et déjà intégrés dans le référentiel général d 'interopérabilité (RGI) qui doit être publié par arrêté au journal officiel en 2008
- S 'imposera au niveau central et local
 - Dans les volets organisationnel, sémantique et technique
 - il concernera également les acteurs du secteur privé à travers les échanges avec l 'administration (appels d 'offres publics par exemple)
- Au niveau du cycle de vie des documents : institution d 'une règle obligatoire d 'intégrer le cycle de vie des données d 'une application dès la mise en œuvre de celle-ci
- Au niveau de la politique d 'archivage : organisation et processus
- Au niveau du format d 'échange : le standard d 'échange de données pour l 'archivage doit être implémenté pour transmission vers un service d 'archives
- Au niveau des formats de conservation (images, structurés, non-structurés, sons et vidéo, graphiques..., composites)
- Au niveau des supports de conservation



11

Le projet PIL@E

Le contexte



12

Le projet PILAE : contexte

- Etude sur les coûts d'une plate-forme d'archivage électronique (Parker-Williborg) à la demande de la DAF, en 2005.
- Les référentiels élaborés dans le cadre du partenariat avec la direction générale de la modernisation de l'Etat



13

Le projet PILAE : enjeux

- Ouverture du nouveau centre des Archives nationales de Pierrefitte sur Seine et de sa plate-forme d'archivage électronique
 - Utilisation du pilote au sein du site de Fontainebleau (service Constance) durant une période transitoire 2008-2011
- Développement de l'administration électronique



14

Le projet PILAE

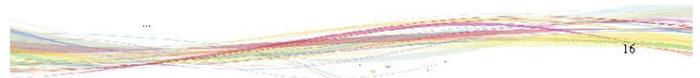
- Repose sur la norme OAIS
 - Description fonctionnelle du pilote
 - Structuration des données (distinction entre les paquets d'archives à soumettre / à conserver / à diffuser)
- Repose sur les normes de description des archives (norme ISAD-G)
- Repose sur les référentiels (politique d'archivage, standard d'échange de données pour l'archivage)



15

PIL@E

Présentation de ses fonctionnalités



16

Vue d'ensemble

- Objectif :
 - recevoir, valider, conserver, communiquer les données et documents nativement numériques produits par les services centraux de l'Etat
 - pour l'instant les utilisateurs de l'application sont :
 - les archivistes des services d'archives dans les ministères
 - les Archives nationales qui reçoivent, contrôlent, valident et conservent
- Non ouvert :
 - aux producteurs
 - au grand public



17

Type de données concernées

- Pouvoir tester le traitement de plusieurs natures d'archives numériques comprenant toutes les productions électroniques existantes depuis plusieurs dizaines d'années
 - données extraites de bases de données
 - documents bureautiques issus des outils de traitement de texte/tableurs/présentations
 - documents issus de GED et décrits par une base de donnée
 - messageries électroniques
 - flux de données sécurisées...



18

Le projet PILAE : contraintes d'ordre organisationnel

- Un point déterminant : l'adhésion sans faille de la directrice des Archives de France et par la suite l'adhésion de la nouvelle directrice du service à compétence nationale des Archives nationales mais...
- Une équipe projet restreinte (DAF/DGME) et un DSI avec des ressources réduites
- Une équipe Constance très qualifiée avec une expérience bien rodée pour l'archivage des bases de données mais en phase d'expérimentations pour les autres typologies documentaires
- Des archivistes dans les ministères en phase d'appropriation de l'archivage électronique
- Des inconnues :
 - visibilité moyenne sur l'appropriation de l'outil
 - son utilisation
 - les volumes de données archivées



19

Le projet PILAE : sa réalisation

- Impossibilité de réaliser une plate-forme en interne aux équipes de la maîtrise d'ouvrage et de la maîtrise d'oeuvre
- Choix de la réalisation d'un pilote (étant donné le degré d'innovation mise en oeuvre) avec une volumétrie modeste mais pouvant évoluer sur les prochaines années
- Réalisation d'un pilote national de plate-forme d'archivage électronique sur la base d'outils existants
 - un coffre-fort électronique (disposant en interne d'outils logiciels assurant les mécanismes de scellement, signature, horodatage, journalisation ad-hoc propres à assurer le maintien de l'intégrité des archives prises en charge) : édité par la société Cecurity.com
 - une infrastructure de stockage IBM assurant les fonctions ad-hoc de sécurisation des supports, de réplication, de sauvegarde



20

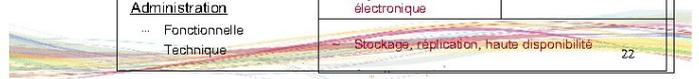
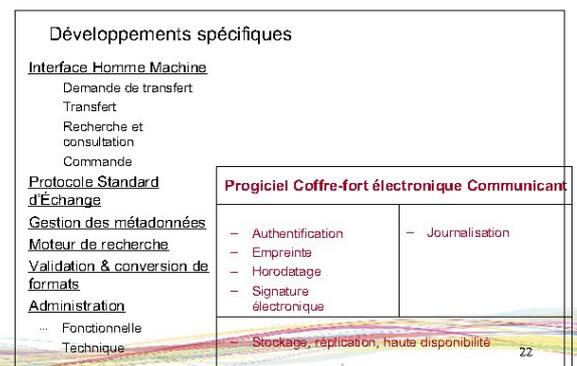
Vue d'ensemble

- Développement :
 - de la partie standard d'échange de données (implémentation)
 - des fonctionnalités archivistiques (pour la prise en charge, la recherche, la consultation, les éliminations, la production de statistiques métier)
 - les fonctions documentaires de recherche et consultation sont développées a minima en attendant l'accostage futur avec des systèmes de gestion documentaire d'archives plus perfectionnés qui intègrent également les archives papier
 - d'un workflow entre services versants et services d'archives
 - intégration d'outils d'identification et de conversion des formats



21

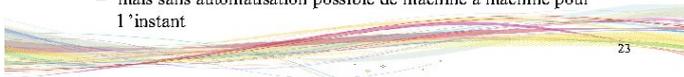
Vue d'ensemble



22

Avec un certain nombre de contraintes

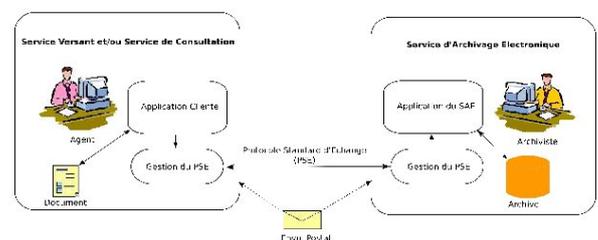
- Pourvoir accepter des versements non encore formatés suivant le standard d'échange de données pour l'archivage (encore la très grande majorité des sources ne possèdent pas de format d'export de ce genre)
 - un fonctionnement dans ce cas « dégradé » avec la prise en charge manuelle de ces versements (en passant par des interfaces de gestion documentaire d'archives à remplir manuellement) et un formatage qui se fait ensuite, au moment du transfert au sein de l'application PILAE
 - l'idéal est bien sûr de recevoir des versements formatés car évite la saisie manuelle versement par versement
 - mais sans automatisation possible de machine à machine pour l'instant



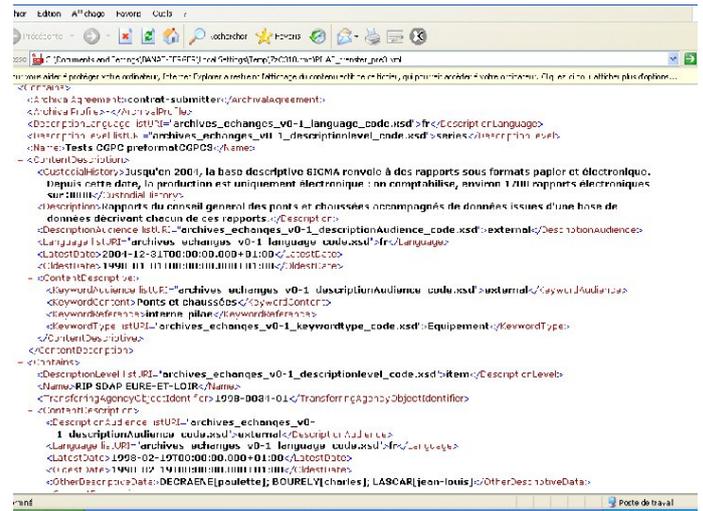
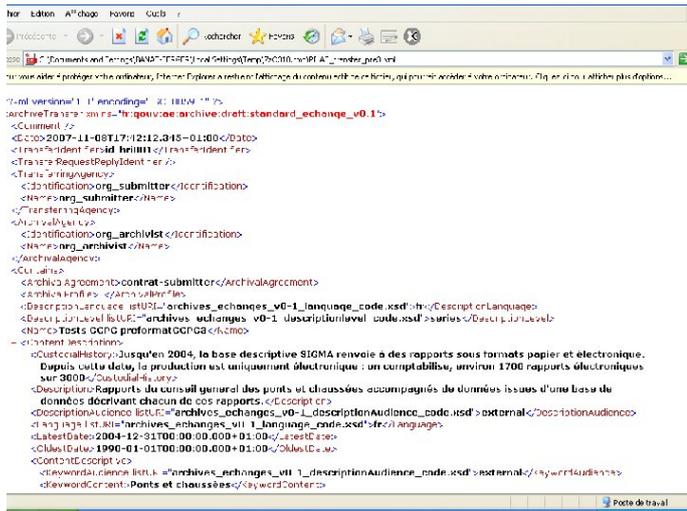
23

Architecture applicative : le principe du standard d'échange

Systeme cible

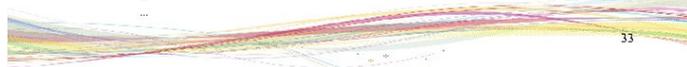


24



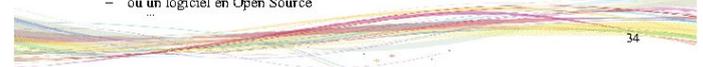
Les formats de données : objectifs

- Tester des conversions de formats par lots en entrée du système
 - Définition des formats de fichiers destinés à l'archivage
 - Définition des formats admissibles en entrée
 - Sélection des outils de conversion
 - Elaboration des stratégies de conversions et des règles de gestion de celles-ci
- Quelques arbitrages
 - Refus des formats MS Office comme formats d'archivage
 - Conservation des versions N, N-1 et la version initiale en cas de migration des formats



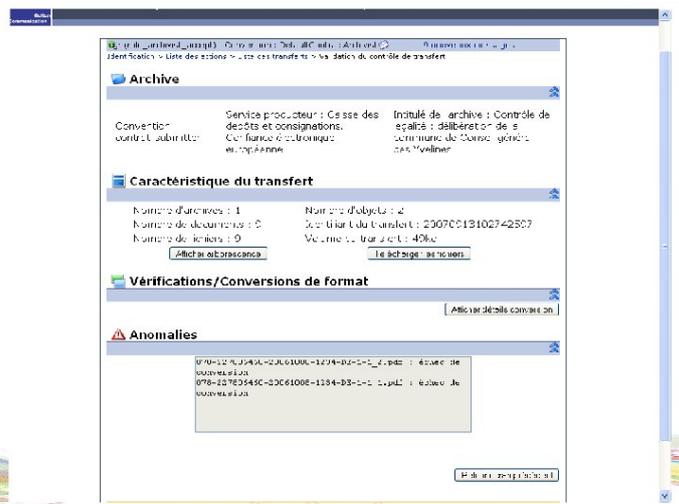
Critères de choix des formats cibles d'archivage

- Respect du Référentiel Général d'Interopérabilité (RGI)
- Format très largement répandu
 - et/ou disposant d'une norme européenne ou internationale
- Les spécifications doivent être publiques et facilement accessibles
- La stabilité des formats doit être « raisonnable »
 - Une version nouvelle au maximum tous les 3 ans
- Conditions imposées :
 - deux logiciels d'éditeurs différents disponibles sur le marché français ou européen
 - ou un logiciel en Open Source



Formats en entrée/cibles

Type de données	Format accepté en entrée (Format(s) supporté(s))	Format cible d'archivage
Images fixes	JPG	JPG
	PDF	PDF
	PDF 2007 (2)	PDF 2000
	PDF 1.4 (3)	PDF
	PDF	PDF
	PDF	PDF
Bases de données	Microsoft Layer II (MPS)	Microsoft Layer II (MPS)
	XML	XML
Documents structurés et non structurés	XML	XML
	OpenDocument	OpenDocument
	Microsoft Office	OpenDocument
Formats composés	ZIP	Format Original
	Microsoft Office	Format Original
Documents structurés	XML	XML
	XML	XML
	XML valide par XSD	XML
	XML	XML
	XML	XML
	XML	XML
Documents graphiques	JPG	JPG
	JPG	JPG
	JPG	JPG
	JPG	JPG



Conclusion

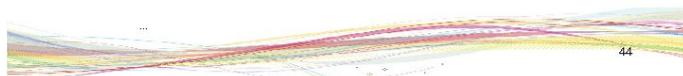
- Mais trois inconnues
 - le degré d'adhésion à l'utilisation de l'outil par les équipes des archives nationales
 - le degré d'adhésion à l'utilisation de l'outil par les archivistes dans les ministères
 - la mise en œuvre dans les applications de production d'exports au format du standard d'échange de données (partenariats à construire ou à approfondir entre archivistes des ministères et informaticiens), condition pour que l'utilisation de PILAE soit efficace en permettant une grande simplification du travail d'ingestion et de contrôle/validation



43

Conclusion

- Au-delà des archives nationales, le modèle de PIL@E pourra être utilisé et repris dans d'autres environnements
 - chez les producteurs
 - chez les archivistes
 - dans les ministères, dans les collectivités territoriales



44

Conclusion

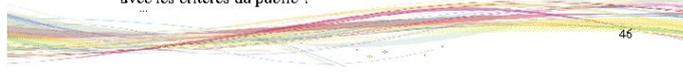
- D'ores et déjà, d'autres réalisations
 - Archives départementales des Yvelines, réalisation d'une plate-forme d'archivage électronique (réalisation en interne au conseil général) qui permet la réception de versements déjà formatés (type contrôle de légalité)
 - Archives municipales de Limoges : utilisation d'une autre solution de coffre-fort électronique
 - Minefe, plate-forme Arcade : solution de coffre-fort interfacé avec un outil de gestion d'archives (pour l'instant pour la partie prise en charge et non versement)
 - Projet au conseil général/archives départementales de l'Aube avec un rapprochement prometteur entre une solution de coffre fort et un éditeur d'outil de gestion et de description d'archives (jusqu'à la papier)



45

Conclusion

- L'objectif principal est d'impulser la mise en place de plates-formes d'archivage
 - Avec des modalités et périmètres variés suivant les contextes
 - Des organisations variées (solutions de mutualisation, d'externalisation)
 - Des mises en œuvre différenciées suivant les acteurs
- Une préoccupation fondamentale des mois à venir
 - Participer à l'organisation de la labellisation/certification des services d'archives électronique (publics, tiers-archivistes), ainsi que des outils logiciels
 - projet de norme ISO en cours de rédaction
 - pour les outils de records management, spécifications MOREQ 2
 - initiative dans le secteur privé prometteuse (le Coref) : rapprochement avec les critères du public ?



46

2.3 SPAR (système de préservation et d'archivage réparti)

Thomas Ledoux (Bibliothèque nationale de France)

Génèse et démarche au sein de Bibliothèque nationale de France.

SPAR

Le Système de Préservation et d'Archivage Réparti de la BnF

Genèse et démarche au sein de Bibliothèque nationale de France

10 avril 2008

Article 10/04/2008

BnF SPAR : genèse et démarche

3

Plan

- Genèse du projet
 - Problématique technique
 - Problématique fonctionnelle
 - Problématique métier
- Enjeux
- Démarche
- Le marché SPAR réalisation

Article 10/04/2008

BnF SPAR : genèse et démarche

2

Genèse 1/3 : Problématique technique

- croissance des volumes
- difficultés de maîtrise des estimations
- difficulté liée des achats successifs

Article 10/04/2008

BnF SPAR : genèse et démarche

3

Difficulté de maîtrise des estimations

- Insuffisance de la maîtrise des nouveaux programmes
- Exemple : passage du N&B vers le niveau de gris
- Choix de numérisation :
 - N&B : TIFF G4 => ~50Ko/page
 - niveaux de gris : TIFF RAW => ~8Mo/page

Article 10/04/2008

BnF SPAR : genèse et démarche

4

Volumétrie BnF (estimation 2005)

		2005	2006	2007	2008	2009	2010	2014
		Période 1	Période 2	Période 3	Période 4	Période 5	Période 6	Période 7
Entité/ projet	Formats							
SA/ conversion Video, doc sonores	Mpeg-2, wav	0	0	0	0	400	412	460
SA/ CD audio pressés à migrer	CD-audio	0	0	0	0	50	56	80
SA/ vidéo à migrer								200
DBN/ Archives Web	html, PDF, Word, JPEG, GIF, PNG, fichiers AVI, RealAudio, RealVideo, MPEG audio et video, Java applets, fichiers Flash ..	80	160	240	320	400	480	800
DSC/NUM Marchés en cours	JPEG, TIFF	1,6	1,7	0,9	1	1	1,05	1,25
DSC/NUM Dunhuang	TIFF	1	1	1	1	1	1	1
DSC/NUM interne	TIFF	6,4	12,8	19,2	25,6	32	38,4	64
DSC/NUM presse	TIFF	9,5	19	28,5	38	47,5	57	95
DSC/NUM externe	TIFF							
DSC/NUM préservation	TIFF	1	2	3	4	5	6	10
DSR/ DRE	TIFF	3,07	3,77	4,47	5,17	5,87	6,57	9,37
PQR/ électronique	PDF	1,82	3,62	5,42	7,22	9,02	10,82	18,02
TOTAL en To		104	204	302	402	951	1 069	1 739

Article 10/04/2008

BnF SPAR : genèse et démarche

5

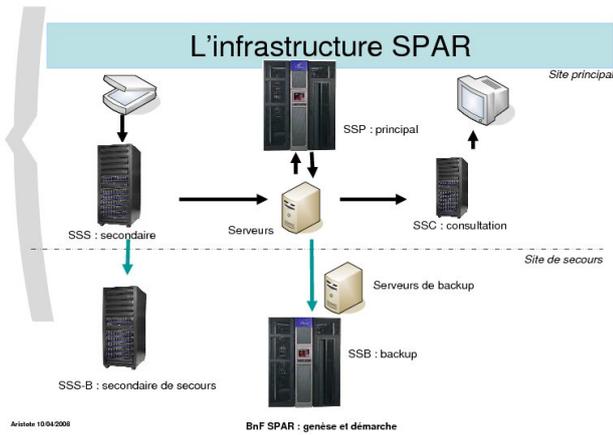
La saga des baies



Article 10/04/2008

BnF SPAR : genèse et démarche

6



Article 10/04/2008

7

Genèse 2/3 : problématique fonctionnelle

- au début de la numérisation,
 - exemplaire numérique=exemplaire de substitution
- Production documentaire uniquement sous forme numérique
 - Dépôt légal : Web, Presse, etc.
- Dématérialisation des supports analogiques
 - Opérations de numérisation de masse des collections papiers
- Numériques suite à l'obsolescence voire la disparition des équipements de restitution
 - Numérisation de sauvegarde des collections audiovisuelles
- Disparition progressive des moyens de productions de microforme
 - Remplacement des moyens de capture analogique

Article 10/04/2008

BnF SPAR : genèse et démarche

8

Genèse 3/3 : problématique métier

- extension du dépôt légal au Web (1^{er} août 2006)
- possibilité de dépôt sous forme électronique
- le domaine numérique n'est plus un moyen de répondre à la mission de diffusion
- la **mission de conservation** s'étend au domaine numérique
- Impulsion des instances dirigeantes de la BnF d'étendre à tout l'Établissement les enjeux du numérique

Article 10/04/2008

BnF SPAR : genèse et démarche

9

Disparition de la spécificité numérique

- en 1994, une mission pour la numérisation
- en 2000, un département de la Bibliothèque numérique
- en 2004, prémisse de l'utilisation du numérique comme support de conservation
- aujourd'hui :
 - un service du dépôt légal pour gérer le DL numérique
 - un service de la coopération pour gérer les partenariats
 - l'Agence Bibliographique Nationale dévient le Département de l'Information Bibliographique et Numérique
- ... le numérique : autre support à gérer dans les départements

Article 10/04/2008

BnF SPAR : genèse et démarche

10

Évolution des métiers

- Prolongation des métiers
 - gestionnaire de collection numérique
 - magasinier numérique
 - expert de préservation
- Relation renforcée avec la technique
 - données stockées sur l'infrastructure de stockage
 - opérations de migration sous le contrôle d'administrateur système
- Évolution du rôle du DSI : passage d'un fournisseur de services vers un partenaire de la gestion des collections

Article 10/04/2008

BnF SPAR : genèse et démarche

11

Plan

- Genèse du projet
 - Problématique technique
 - Problématique fonctionnelle
 - Problématique métier
- Enjeux
- Démarche
- Le marché SPAR réalisation

Article 10/04/2008

BnF SPAR : genèse et démarche

12

Le projet SPAR

- Imaginer et concevoir un système qui puisse permettre à la BnF de :
 - PRESERVER son patrimoine numérique
 - ARCHIVER l'ensemble de ses données
 - REPARTIR l'accès à ses données
- Nom de code : SPAR
- Mise en œuvre du système en 2 temps :
 - Marché infrastructure (fin 2005) : serveurs de pilotage, espaces disques, librairies de bandes
 - Marché réalisation (début 2008) : versement, gestion des données et accès, planification, migration, administration et exploitation

Article 10/04/2008

BnF SPAR : genèse et démarche

13

Enjeu économique

- Il s'agit de rationaliser à long terme les investissements :
 - Coût d'acquisition ou de remplacement de l'information
 - Coût des infrastructures
- Pour quelle durée :
 - Pour une période assez longue pour que l'on se soucie de l'impact :
 - Des changements technologiques (ceux nécessaires pour les nouveaux supports et les nouveaux formats de données)
 - De l'évolution de l'ensemble des utilisateurs sur les informations détenues par une archive.
- Période qui peut s'étendre à l'infini ...

Article 10/04/2008

BnF SPAR : genèse et démarche

14

Enjeu organisationnel (1/2)

- Politique de versement : caractéristiques de la négociation et du protocole de versement
 - Qui négocie ?
 - Que négocie-t-on ? : droit, format, volumétrie, flux
- Politique d'archivage : caractéristiques de la conservation
 - Que conserve-t-on ? : données, métadonnées, systèmes de représentation... original, master, produits dérivés
 - Comment conserve-t-on ? : à l'identique (train original des bits), émulation, migration
 - Combien de temps conserve-t-on ?

Article 10/04/2008

BnF SPAR : genèse et démarche

15

Enjeu organisationnel (2/2)

- Politique d'accès : caractéristiques de l'accès
 - Avec ou sans restriction
 - Avec ou sans services supplémentaires (ex. veille des formats)
 - Immédiat ou différé
 - Direct (applicatif DSI) ou indirect
 - Sur place, à distance
 - Volumétrie des transactions

Article 10/04/2008

BnF SPAR : genèse et démarche

16

Gestion du risque

- Technologique
 - Mesure de l'universalité des supports, des formats
 - Réduction des échelles de temps : compatibilité des *documents bureautiques dépassée en moins de 10 ans*
 - Accélération des évolutions de la technologie
- Environnemental et organisationnel
 - Information qui peut s'altérer sans que l'on s'en aperçoive et qui conduit à la perte de l'information
 - Condition de stockage, lieu, capacité, ...
 - Activité de pérennisation à mettre en place en repensant l'organisation du travail, en partageant les responsabilités, en utilisant les bonnes compétences
- Budgétaire
 - Coût solutions, moyen de la politique d'archivage... à long terme

Article 10/04/2008

BnF SPAR : genèse et démarche

17

Plan

- Genèse du projet
 - Problématique technique
 - Problématique fonctionnelle
 - Problématique métier
- Enjeux
- Démarche
- Le marché SPAR réalisation

Article 10/04/2008

BnF SPAR : genèse et démarche

18

Démarche

- Définition du besoin (Groupes de travail) 2006
- Étude technique des solutions possibles 2006
- Validation du scénario de réalisation par le comité de pilotage début 2007
- Marché de réalisation 2007
 - Rédaction du cahier des charges de réalisation
 - Réception des offres
- Réalisation par itération 2008 ...

Article 10/04/2008

BnF SPAR : genèse et démarche

19

Groupes de travail

- Les groupes de travail fonctionnels
 - G1 : groupe transverse de suivi
 - G2 : communauté d'utilisateurs
 - G3 : modèles d'information
 - G4 : gestion des risques
 - G2D : gestion des droits
- Groupe de travail technique

Article 10/04/2008

BnF SPAR : genèse et démarche

20

Groupe G2 : les filières

- Numérisation de conservation
 - imprimés, périodiques, images fixes
 - audiovisuel et multimédia
- Numérisation de consultation
- Dépôt légal automatique
- Dépôt légal négocié
- Production administrative/ technique
- Dons et acquisitions
- Tiers archivage

Article 10/04/2008

BnF SPAR : genèse et démarche

21

Groupe G3 : modèle d'information

- Empaquetage via METS :
 - Métadonnées descriptives : DC
 - Métadonnées techniques : dépendent de la filière (MIX, textMD, ...)
 - Métadonnées de préservation : PREMIS
 - Métadonnées de droits : ODRL
- 4 niveaux de granularité :
 - set: collection
 - group: objet numérique (monographie, lot d'images, ...)
 - object : élément numérique
 - file: objet-donnée (fichier numérique ou train de

Article 10/04/2008

BnF SPAR : genèse et démarche

22

Groupe G4 : gestion de risques

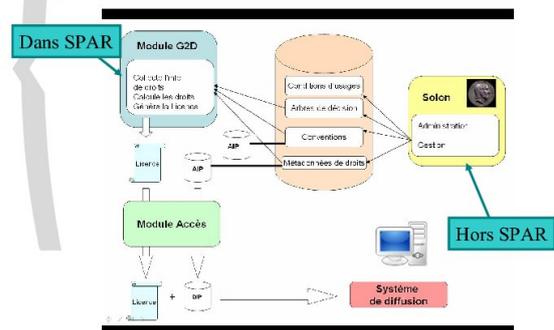
- On peut auditer le système en vue de la **certification**, ou en vue de prioriser les actions.
- La gestion des risques donne une vue d'ensemble des actions de préservation menées sur l'ensemble des objets numériques.
- Trois initiatives de certification/audit fortement convergentes mais qui restent distinctes pour des raisons géopolitiques :
 - DRAMBORA
 - RLG Checklist
 - NESTOR checklist
- Travaux du CCSDS : *ISO standard for Audit and Certification of repositories of digital information*
- Volonté de certification prise en compte dans le CCTP.

Article 10/04/2008

BnF SPAR : genèse et démarche

23

Groupe G2D : gestion des droits

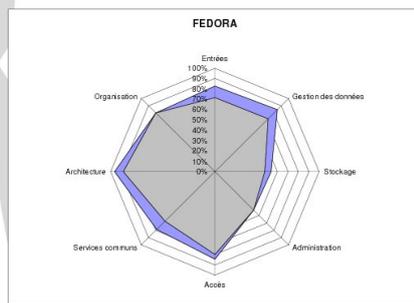


Article 10/04/2008

BnF SPAR : genèse et démarche

24

Groupe technique : choix de Fedora



Min : 173 (66%)
Max : 193 (73%)

Article 10/04/2008

BnF SPAR : genèse et démarche

25

Plan

- Genèse du projet
 - Problématique technique
 - Problématique fonctionnelle
 - Problématique métier
- Enjeux
- Démarche
 - Le marché SPAR réalisation

Article 10/04/2008

BnF SPAR : genèse et démarche

26

SPAR : rédaction du CCTP

Entrants :

Rapport d'analyse des solutions
Documents issus des groupes de travail
Marché d'infrastructure

Définition des filières :

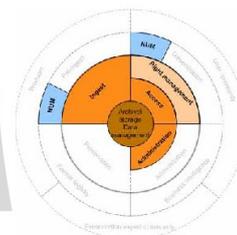
filière de numérisation de conservation,
filière de numérisation de reproduction,
dépôt légal des données audiovisuels,
dépôt légal des données électroniques,
archivage des documents administratifs,
tiers archivage

Article 10/04/2008

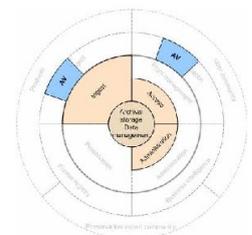
BnF SPAR : genèse et démarche

27

Réalisation : construction itérative



Tranche ferme



Tranche conditionnelle

Article 10/04/2008

BnF SPAR : genèse et démarche

28

SPAR : Réalisation du système

- 2008 Réalisation du noyau commun et de FIL_NUM_CONS_A
- 2009-2011, par étapes successives :
 - Ajout de filières
 - Ajout de fonctions avancées (préservation...)
- Cela permet de
 - prendre en compte l'aspect innovant du système,
 - mettre en oeuvre le système sans attendre d'avoir tout fini,
 - assurer au mieux l'acquisition progressive du système par les utilisateurs

Article 10/04/2008

BnF SPAR : genèse et démarche

29

Merci de votre attention

Questions ?

Thomas Ledoux
thomas.ledoux AT bnf.fr

Article 10/04/2008

BnF SPAR : genèse et démarche

30

2.4 PAC, plate-forme d'archivage du CINES

Olivier Rouchon (CINES : Centre Informatique national de l'enseignement supérieur)

Enjeux et perspectives pour l'information scientifique et technique.



Sommaire

- Le Centre Informatique National de l'Enseignement Supérieur
- La mission d'archivage du CINES
- L'information scientifique et technique
- Les enjeux
- Le projet PAC, Plateforme d'Archivage du CINES
- Les dates clés
- L'état des lieux
- Les projets d'archives au CINES
- Les perspectives



Séminaire Aristote - Pérennisation de l'Information

2



La présentation du CINES

Centre Informatique National de l'Enseignement Supérieur

- Basé à Montpellier (Hérault, France)
- Créé en 1999, succédant au CNUSC (Centre National Universitaire Sud de Calcul) – créé en 1980
- Placé sous la tutelle de la DGRI (Direction Générale de la Recherche et de l'Innovation) et de la DGES (Direction Générale de l'Enseignement Supérieur) du Ministère de l'Enseignement Supérieur et de la Recherche
- Principales missions
 - Calcul numérique intensif
 - Archivage pérenne de documents électroniques
 - Hébergement et suivi de serveurs d'applications
- Plus d'information : <http://www.cines.fr>



Séminaire Aristote - Pérennisation de l'Information

3



La mission d'archivage du CINES

Depuis 2004, le CINES travaille sur la mise en place d'un service pour l'archivage pérenne du patrimoine scientifique.

Cette mission a été confirmée par plusieurs décisions des organismes de tutelle :

- Arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue d'un doctorat
- Convention du 2 mai 2007 (faisant suite à celle du 15 octobre 2003) relative à la mise en ligne et l'archivage pérenne de données numérisées dans le cadre du programme Persée
- Lettre de cadrage du 12 février 2008 recentrant les activités du CINES autour de deux missions stratégiques :
 - le calcul intensif
 - l'archivage pérenne

Séminaire Aristote - Pérennisation de l'Information

4



L'information scientifique et technique (1/2)

L'information scientifique et technique (IST) désigne l'ensemble des informations produites ou reçues par les secteurs de la recherche, de l'enseignement

- Elle est définie par :
 - Des contenus constitutifs de connaissances
 - Des supports documentaires
 - Des canaux de communication spécifiques
- Elle s'inscrit dans des formes documentaires diverses :
 - Revues scientifiques, thèses, rapports, actes, ouvrages spécialisés, manuels, bibliographies, résumés, prépublications...
 - Brevets, cartes, banques d'images et de vidéos, données statistiques...
 - De manière plus informelle, hors publication : échanges interpersonnels, courriers électroniques, groupes de discussion, notes de travail, données brutes, etc.

Séminaire Aristote - Pérennisation de l'Information

5



L'information scientifique et technique (2/2)

- Elle est en phase de mutation
 - Les supports documentaires et canaux de communication évoluent
 - Passage massif au numérique
 - Documents nativement au format électronique
 - Opérations de numérisation de masse
 - Les volumes d'information croissent de façon exponentielle
- Ses limites demeurent imprécises
 - Très liée à d'autres types d'informations produites hors du champ de la recherche
 - Phénomène accentué par les profondes transformations des canaux de communication

La préservation à long terme de cette information est impérative, et compliquée par la nouvelle nature de son support (électronique)

Séminaire Aristote - Pérennisation de l'Information

6



Les enjeux (1/3)

La réussite au niveau national d'une stratégie pour l'archivage pérenne de documents électroniques produits par la communauté IST passe par la réalisation d'un certain nombre de défis – sur lesquels le CINES est actif :

- Acquisition d'une nouvelle compétence métier
 - Mise en place d'un service dédié à l'archivage
 - Définition de nouveaux processus et méthodes à mettre en œuvre et à maîtriser pour piloter un système d'archives
 - Participation à différents groupes de travail ou initiatives sur le thème de la préservation de documents numériques
 - Nationaux : PIN, CCT CNES
 - Européens : DPE, CASPAR, Alliance, DP-Net (programmes européens)

Les enjeux (2/3)

- Sensibilisation de la communauté IST à la problématique de la préservation à long terme des documents numériques
 - Renforcement des collaborations entre informaticiens, archivistes et bibliothécaires
 - Organisation de journées de sensibilisation et d'information à l'archivage pérenne
 - Participation à divers séminaires (JRES, Aristote) ou journées d'études (Rencontres RNBM Marseille, URFIST Bordeaux)
- Sensibilisation des décideurs sur l'importance de l'enjeu
 - Progression inexorable et exponentielle du numérique
 - Sélection des informations à archiver face au volume d'information atteint
 - Caractère impératif de la préservation de l'information scientifique et technique
 - Difficulté de prise en compte du très long terme
 - Coûts

Les enjeux (3/3)

- Émergence et reconnaissance d'un acteur majeur dans le domaine de l'archivage pérenne pour l'IST
 - Étude du CNES (2007)
 - Étude du CERN (2008) pour le TGE-Adonis
- Intégration de la chaîne fonctionnelle d'archivage
 - Des producteurs de documents et services versants jusqu'au service d'archives
 - D'autres services d'archives interviennent dans cette chaîne selon l'origine et la nature des documents
 - Archives départementales (Universités)
 - Archives Nationales (EPA Education Nationale/Enseignement Supérieur)
 - BnF (dépôt légal)
 - Respect du contexte législatif

Le projet PAC

Pour remplir la mission d'archivage, le CINES a mis en place le projet PAC, qui vise à se doter d'une plate-forme et d'un service d'archivage numérique pérenne

- L'équipe actuelle
 - 1 chef de projet
 - 4 ingénieurs
 - 1 archiviste
- 2 projets pilotes en cours
 - Archivage des thèses électroniques
 - Documents nativement au format électronique
 - Archivage des revues SHS du portail Persée
 - Documents issus de la numérisation de revues au format papier

Les dates clés

- 2004** • Démarrage du projet, constitution de l'équipe
 - Etude des normes, participation au groupe de travail PIN
- 2005** • Etude du cabinet Ourouk pour l'aide au choix d'une solution logicielle pour l'archivage électronique
 - Décision d'un développement interne, aucune solution existante n'étant satisfaisante
- 2006** • Documentation des spécifications fonctionnelles pour la solution souhaitée
 - Développement en interne d'une première plateforme
- 2007** • Premiers tests d'intégration avec l'outil STAR de l'ABES pour le versement de thèses électroniques
 - Déploiement de PAC v1.0 en production et démarrage de l'exploitation
- 2008** • Appel d'offre pour une plateforme PAC v2.0 notifié à la société SUN
 - Déploiement prévu au deuxième trimestre

L'état des lieux (1/2)

- Phase 1 : Développement interne d'une première plateforme pour valider les services attendus sur le projet ABES/STAR, capacité de stockage réduite (300 Go) – PAC v1.0
 - Basée sur les standards du domaine
 - Modèle OAIS
 - Protocole standard d'échange de données pour l'archivage
 - Métadonnées Dublin Core
 - Liste des formats de fichier acceptés volontairement limitée
 - Formats publiés, largement utilisés, normalisés si possible
 - HTML, PDF, TXT, XML
 - GIF, JPEG, TIFF, PNG
 - Architecture basée sur les logiciels libres
 - Java, PostgreSQL, Jhove, ImageMagick
 - Premiers tests d'archivage des thèses en Mars 2007
 - Début de l'exploitation en production fin 2007

L'état des lieux (2/2)

- Phase 2 : Appel d'offres notifié fin 2007 pour l'acquisition d'une plateforme de stockage capable de gérer de larges volumes (20 To extensibles à 40To) – PAC v2.0
 - Basée sur les standards du domaine
 - Modèle OAIS
 - Protocole standard d'échange de données pour l'archivage
 - Métadonnées Dublin Core
 - Liste des formats de fichier acceptés identique à PAC v1.0
 - Architecture basée sur du matériel SUN, le logiciel Arcsys et des logiciels libres
 - Java, MySQL, Jhove, ImageMagick
 - Premiers tests de versement et de migration (documents archivés sur PAC v1.0) en Mars 2008
 - Début de l'exploitation fin du deuxième trimestre 2008

Le projet d'archives « Thèses électroniques » ABES

Initié suite à l'arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue d'un doctorat

- Les doctorants déposent leur thèses au format électronique dans la BU (Bibliothèque Universitaire) de leur lieu de soutenance
- Les BU versent les thèses électroniques à l'ABES via l'outil STAR
- Après 3 étapes de validation, les thèses éligibles à l'archivage sont transférées sur la plateforme PAC
- Une copie est éventuellement disponible en ligne pour la communauté des internautes sur un site de diffusion
- L'outil STAR centralise les demandes de communication d'archives

Ce projet est actuellement en phase de production, l'archivage des thèses a débuté fin 2007

<http://www.abes.fr/abes/page,555,manuel-de-lutilisateur-star.html>

Le projet d'archives « revues SHS » Persée

Initié en 2006 pour répondre à un projet de numérisation massive et de préservation de collections rétrospectives de revues en Sciences Humaines et Sociales par l'équipe Persée (Université Lumière – Lyon 2).

- La chaîne de numérisation assure
 - une digitalisation de masse
 - une centralisation et une robotisation des traitements
 - un archivage pérenne des données
- La chaîne de documentation comprend
 - un outil de description des collections
 - des outils de documentation et de suivi
 - des étapes de contrôle qualité et de validation
 - des outils pour la diffusion des données générées (<http://www.persee.fr>)

Ce projet est actuellement dans sa phase préliminaire (définition des objets à archiver, du protocole de versement). Les premiers tests d'intégration sont planifiés pour le deuxième trimestre 2008

Les projets en cours d'étude

Les projets suivants sont en cours d'étude

2. Archivage pérenne des documents versés dans les Archives Ouvertes (HAL – Hyper Article en Ligne, <http://www.archives-ouvertes.fr/>)
 - Étude préliminaire en cours
 - Calendrier prévisionnel fin 2008
3. Autres projets de numérisation et d'archivage proposés
 - IRHT
 - TGE-Adonis
 - CERIMES
 - EFEO

Tous les projets d'archives partageront la même plateforme

- Mutualisation de l'infrastructure matérielle d'archivage
- Protocole de versement générique
- Diminution des coûts de mise en place et d'exploitation

Les perspectives

La mise en place d'une stratégie pour la préservation à long terme de documents électroniques issus de l'IST en est à ses débuts.

Elle réussira si la chaîne fonctionnelle d'archivage est claire et les acteurs identifiés. Le CINES y jouera un rôle clé dans la mesure où :

- Il est maintenant un acteur reconnu du domaine de la préservation à long terme de documents numériques :
 - Expérience dans l'archivage pérenne (documents nativement produit au format électronique ou issus de la numérisation de publications, photos, etc.)
 - Une étude récente menée par le TGE-Adonis préconise une collaboration étroite IN2P3 – CINES pour l'accès et la préservation de l'information SHS
 - Rôle confirmé par le ministère de tutelle
- Il reçoit de nombreuses sollicitations de laboratoires et d'universités pour divers services :
 - Aide et conseil dans la construction de projets d'archivage à long terme, retours d'expérience, propositions de projets d'archives
 - Dans le respect du contexte législatif archivistique français

Questions

&

Réponses

2.5 Etat d'avancement du standard NAS – EN 9300 LOTAR

Pierre Duchier (AIRBUS)

«Long-term archival and retrieval of aerospace product information»



ASD-STAN
ASD-STAN

PROSTEP
VIP

AIRBUS

FAI SYSTEMS

DASSAULT
AVIATION

EADS

Snecma

Long Term Archiving and Retrieval of Digital Technical Product Data in the aerospace community

Aristote Conference
Etat d'avancement du standard NAS – EN 9300 LOTAR, « Long-term archival and retrieval of aerospace product Information »
10th april 2008

Presented by: Pierre Duchier (CIMPA on behalf of AIRBUS)
Prepared by: JY Delaunay - Airbus

Follow up of the presentation held on the 16th of March 2006



ASD-STAN
ASD-STAN

PROSTEP
VIP

AIRBUS

FAI SYSTEMS

MTU

Snecma

Long Term Archiving and Retrieval of Digital Technical Product Data in the aerospace community

Aristote Conference
"Longevity of digital Information" – Paris – Ecole Polytechnique
16th of March 2006

Presented by: JY Delaunay - Airbus

Table of content



- Follow up of the presentation held on the 16th of March 2006
- The LOTAR work group
- Main activity
- Overview of NAS 9300 - EN 9300 LOTAR standards
- Fundamental & concept
- LOTAR specific concept (not in the OAIS)
- Overview of the LOTAR project: version
- Overview of the LOTAR project: Planning
- Evolution of the design method
- LOTAR Part 100 - Common concepts of CAD 3D mechanical Information
- LOTAR Part 110 - CAD mechanical 3D Explicit geometry Information
- LOTAR Part 120 - 3D Explicit Geometry with Geometric Dimension & Tolerances associated with Form-Features
- Released and ballot documents
- Future work
- Summary – next actions

ASD-STAN

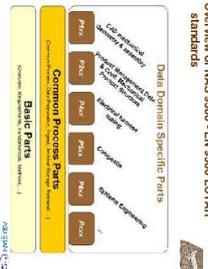
The LOTAR work group



Site web : <http://www.prostep.org/lotar/>

ASD-STAN

Overview of NAS 9300 - EN 9300 LOTAR standards



Basic Parts

Common Process Parts

Data Domain Specific Parts

Main activity

LOTAR is based on a page



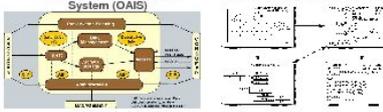
LONG TERM ARCHIVAL & RETRIEVAL OF DIGITAL PRODUCT DATA

Fundamental & concept

LOTAR is mainly based on 2 standards:

ISO 14721 – Reference Model for an Open Archival Information System (OAIS)

ISO 10303 - STEP (Standard for the Exchange of Product model data)

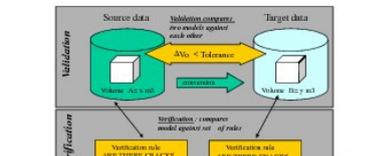


ASD-STAN

LOTAR specific concept (not in the OAIS):

- Data preparation phase (before ingest)
- Qualification methods principle :

 - Verification → data quality control
 - Validation → conversion control



ASD-STAN

LOTAR Part 110 Long Term Archiving and Retrieval of CAD mechanical 3D Explicit geometry information

➤ **Conversion in STEP AP214 or AP203 format of the exact boundary representation shape**

➤ **Verification:**

- 7 mandatory rules: Large Segment Gap / Non-Tangent Segments / Large Edge Gap / Free Edge / Over-Used Edge / Fragmented Surface / Empty Model
- 8 optional rules

➤ **Validation :**



Volume, centroid and area Clouds Of Points

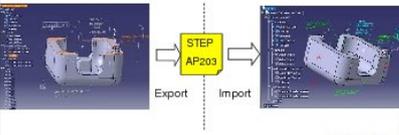
Released and ballot documents

EN 9300-001	Structure	released
EN 9300-002	Requirements (V2)	ballot Q2 2008
EN 9300-003	Fundamentals and concepts	released
EN 9300-004	Description Methods	released
EN 9300-005	Authentication and Verification	ballot Q2 2008
EN 9300-007	Terms and References	ballot Q2 2008
Common Process Parts		
EN 9300-010	Overview Data Flow	released
EN 9300-011	Data Preparation	released
EN 9300-012	Ingest	released
EN 9300-013	Archival Storage	released
EN 9300-014	Retrieval	released
EN 9300-015	Removal	released
Data Domain Specific Parts		
EN 9300-100	Fundamentals and concepts	ballot Q2 2008
EN 9300-110	Explicit Geometry	ballot Q2 2008
EN 9300-115	Explicit Assembly Structure	ballot Q2 2008

LOTAR Part 120 - 3D Explicit Geometry with Geometric Dimension & Tolerances associated with Form-Features

➤ **The Challenge** is to avoid 2D drawing with the use of 3D annotations

➤ **Pilots project** is in progress for the archiving of this kind of information



Export Import

Future work

Support Process Parts

- EN 9300-03X Preservation Planning

Data Domain Specific Parts – CAD

- EN 9300-120 3D Explicit Geometry with Geometric Dimension & Tolerances associated with Form-Features (parts)
 - Version 1: representation as polylines
 - Version 2: full representation
- EN 9300-125 3D Explicit Assembly Structure with Geometric Dimension & Tolerances associated with Form-Features (assembly)
- EN 9300-130 3D parametric Geometry
- EN 9300-135 3D parametric Assembly Structure

Data Domain Specific Parts – PDM / PLM

- EN 9300-200 Fundamentals and concepts
- EN 9300-215 PDM/TDM 3D Configuration Assembly Structure

Summary – next actions

➤ **New challenges:**

- Development of new parts, for LT Archiving of:
 - PDM information,
 - CAD 3D Composite, CAD 3D electrical harness
 - Audit & Certification (link with new CAIS parts)
- Promote to CAD-PDM vendors, reluctant to support functions required by Aerospace manufacturers

➤ **LOTAR standards are related with other standards for Interoperability of Aerospace digital information:**

- Setting up of an European Governance structure for e-Business open standards, including:
 - Exchange, visualization, quality control, and LT Archiving of PLM data

Web based infrastructure to support the dev. of LOTAR standards

➤ **LOTAR public web site:** <http://www.prostep.org/lotar/>



Merci

2.6 Un contexte normatif très enrichi

Marie-Anne Chabin (Société Archive17)

Moreq2 à la rencontre du Modèle OAIS.

Un contexte normatif très enrichi : Moreq2 à la rencontre du Modèle OAIS



Séminaire Aristote du 10 avril 08

Marie-Anne Chabin, Archive17
www.archive17.fr

Sommaire

1. De MoReq à MoReq2
2. Des concepts communs ou presque communs avec l'OAIS
3. Records management et conservation numérique
4. Articulation des deux modèles

MA Chabin_Archive17

Aristote - 10.04.08

2

Sommaire

1. De MoReq à MoReq2
2. Des concepts communs ou presque communs avec l'OAIS
3. Records management et conservation numérique
4. Articulation des deux modèles

MA Chabin_Archive17

Aristote - 10.04.08

3

Les étapes de MoReq

1. 1996: décision lors du DLM Forum



5. 2001: publication de MoReq

- Model Requirements for the Management of Electronic Records
- Modèle d'exigences pour l'organisation de l'archivage électronique

6. 2004: traduction française par Marie-Anne Chabin (10 sponsors)

8. 2008: publication de MoReq2

- décembre 2008: présentation de la traduction française par la Direction des Archives de France au DLM Forum de Toulouse



MA Chabin_Archive17

Aristote - 10.04.08

4

La notion de records

- Un record est un document:
 - validé
 - figé (non modifiable)
 - « set apart as a record » (sélection sur des critères de risques de non disponibilité dans le temps)
 - qualifié (documenté)
 - Approche globale papier et électronique
 - le support est secondaire par rapport à la valeur de l'information
3. Les archives historiques (patrimoniales) sont hors du champ du records management

MA Chabin_Archive17

Aristote - 10.04.08

5

La cible

MoReq s'adresse:

2. aux utilisateurs d'un système d'archivage électronique (SAE) pour rédiger leur cahier des charges ou auditer un système existant
3. aux organismes de formation sur l'archivage
4. éditeurs
5. aux prestataires de service en archivage
6. aux utilisateurs des prestations de service
7. aux auditeurs de solutions d'archivage

MA Chabin_Archive17

Aristote - 10.04.08

6

Un plan de classement hiérarchique

1. classement

- (dans l'archivage). Identification et ordonnancement systématiques des activités métiers et/ou des *documents archivables* en catégories selon une structure logique prédéfinie, des méthodes et des règles concrétisées dans un plan de classement (cf ISO 15489)

2. plan de classement

- (dans MoReq2). Organisation hiérarchique des séries, dossiers, sous-dossiers, volumes et documents
- un document est rattaché prioritairement à un point du plan de classement mais peut faire l'objet de liens (copies dans d'autres dossiers)

3. code de classement

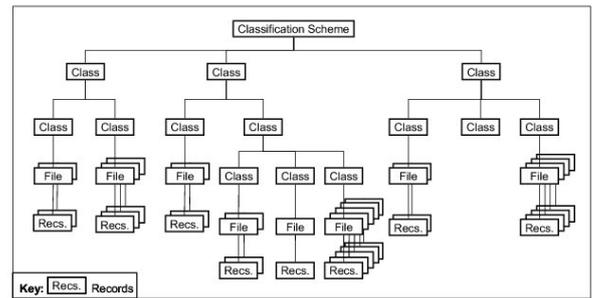
- Identifiant attribué à chaque série du plan de classement. Au sein de chaque série, les codes de classement des séries filles sont uniques.

MA Chabin_Archive17

Aristote - 10.04.08

7

Un plan de classement hiérarchique



source: MoReq2

class = série
file = dossier
records = documents archivés

MA Chabin_Archive17

Aristote - 10.04.08

8

La capture

- Capture comporte 4 opérations:
 - enregistrement (attribution d'un identifiant unique dans le système)
 - classement
 - ajout de métadonnées
 - stockage
- 2. Capture ou déclaration (sélection manuelle):
 - document à plusieurs versions: archiver toutes les versions du document comme une seule entité ; archiver une seule version du document ; archiver chaque version séparément
 - messages électroniques et leurs pièces jointes

MA Chabin_Archive17

Aristote - 10.04.08

9

Règles de conservation/destruction

1. Règles associées aux séries ou aux dossiers:
 - durée de conservation (en années) avec le point de départ
 - le devenir de ces documents à l'issue de ce laps de temps (sort final)
2. Toute entité archivée (capturée) doit être reliée à au moins une règle de conservation/destruction
4. Plusieurs niveaux d'application de la destruction
 - automatique, contrôle, conservation ou non des métadonnées

MA Chabin_Archive17

Aristote - 10.04.08

10

Exigences techniques

1. Traitement de tous types de documents/données:
 - non structurés mais aussi structurés,
 - tous formats et tous supports,
 - documents hétérogènes ou composites (page web..)
2. Automatisation maximale de la capture (des documents et des métadonnées)
3. Destruction et export
4. Contrôles et alertes
 - en cas de documents non conformes par rapport aux exigences de qualité définies
 - échéance de durées de conservation

MA Chabin_Archive17

Aristote - 10.04.08

11

La sécurité dans MoReq

1. Contrôles d'accès
 - un administrateur du SAE
 - des profils utilisateurs
 - des indices de sécurité pour les documents
2. Traçabilité
 - un historique de tous les événements qui modifient ou affectent les documents archivés (audit trail)
 - tous les changements de localisation sont tracés
3. Sauvegardes et restauration
4. Archives vitales
 - cf plan de continuité d'activité

MA Chabin_Archive17

Aristote - 10.04.08

12

Exigences non fonctionnelles

1. l'archivage des documents non électroniques (conservation et destruction des dossiers mixtes)
2. GED et workflow
3. signature électronique
4. interopérabilité
5. évolutivité

MA Chabin_Archive17

Aristote - 10.04.08

13

Les ajouts de MoReq2 (1)

- Prise en compte des évolutions juridiques et technologiques
 - la signature électronique et ses développements
 - formats et supports (mais MoReq2 n'en préconise pas)
- Les documents composites
 - composant: « train de bits » indépendant qui, seul ou avec d'autres, constitue un document
 - exemples :
 - un document HTML et les images JPEG qui constituent une page web
 - un document bureautique, composé d'un texte avec un hyperlien vers un tableur
 - un courriel avec ses pièces jointes

MA Chabin_Archive17

Aristote - 10.04.08

14

Les ajouts de MoReq2 (2)

1. Des modules optionnels pour l'intégration avec:
 - des applications métier, la messagerie
 - les outils de GED, workflow, système de gestion de contenu, système de gestion de dossier
 - les systèmes distribués, le travail à distance
3. Le modèle de métadonnées de MoReq2
 - jeux de métadonnées pour tous les niveaux du plan de classement
 - modèle relationnel des entités

MA Chabin_Archive17

Aristote - 10.04.08

15

Les tests de conformité



La grande nouveauté de MoReq2

3. batterie de tests logiciels
4. tester la conformité des logiciels du marché aux spécifications MoReq
 - ce n'est pas encore la certification
5. destiné
 - aux éditeurs
 - aux utilisateurs

MA Chabin_Archive17

Aristote - 10.04.08

16

Sommaire

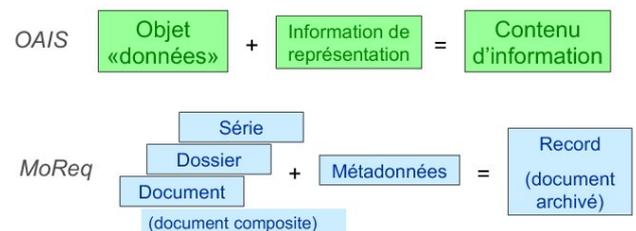
1. De MoReq à MoReq2
2. Des concepts communs ou presque communs avec l'OAIS
3. Records management et conservation numérique
4. Articulation des deux modèles

MA Chabin_Archive17

Aristote - 10.04.08

17

Le modèle d'information



Des terminologies différentes pour des objets documentés comparables

MA Chabin_Archive17

Aristote - 10.04.08

18

Le modèle fonctionnel

1. Deux bornes:
 - la mise en archive (l'entrée) réglementée avec le producteur
 - la sortie (pour l'utilisateur ou sortie définitive) contrôlée
2. et un intervalle:
 - La conservation
 - les supports: stockage, pérennisation
 - la recherche des informations
 - les contrôles d'accès

+

une **politique** (un budget pour un objectif)
 une **administration** du dispositif
 (compétences techniques et archivistiques)

Sommaire

1. De MoReq à MoReq2
2. Des concepts communs ou presque communs avec l'OAIS
3. **Records management et conservation numérique**
4. Articulation des deux modèles

Les deux temps de l'archivage

1. Sélectionner ce qu'il faut archiver

- Exigences de:
 - exhaustivité
 - authenticité
 - intégrité
 - exploitabilité

Records management

2. Conserver ce qui est archivé

- Exigences de:
 - pérennité
 - intégrité
 - Intelligibilité
 - disponibilité

Conservation numérique

Données structurées / Données non structurées

OAIS

- plutôt dédié aux données structurées
- données scientifiques (≠/≠ acte juridique)
- plutôt dédié aux données patrimoniales (la gestion des durées de conservation est possible mais n'est pas centrale)

MoReq

- plutôt dédié aux données non structurées
- hétérogénéité des documents archivables (tout type d'information de trace dans l'entreprise)
- traçabilité du cycle de vie (validation, diffusion, destruction)

Service orienté utilisateurs / Système orienté entreprise

OAIS

- Une Archive est dédiée à une communauté définie d'utilisateurs
- Préserver et mettre les contenus à disposition plutôt que gérer la valeur juridique des données

MoReq

- Le SAE doit répondre aux exigences réglementaires d'une entreprise
- Anticiper les risques d'audit et de contentieux (archiver et détruire à bon escient)

Poids de la pérennité / Poids de la sécurité

OAIS

- Pérenniser des To de données (plusieurs décennies)
- Programmer et effectuer les migrations nécessaires

MoReq

- Maîtriser l'exhaustivité de l'information archivable en termes de responsabilité
- Appliquer les règles de gestion (durée de conservation, droits d'accès) et tracer les événements

Sommaire

1. De MoReq à MoReq2
2. Des concepts communs ou presque communs avec l'OAIS
3. Records management et conservation numérique
4. **Articulation des deux modèles**

MA Chabin_Archive17

Aristote - 10.04.08

25

MoReq2 en amont de l'OAIS

- **Avant la mise en archive, définition des exigences fonctionnelles**
- 2. L'environnement réglementaire
- 3. L'évaluation des risques et le plan de classement
- 4. Les règles de conservation destruction (une grande partie des records sont éliminables à court ou moyen terme – 10 ans)
- 5. L'historique des événements: traçabilité de la capture et des accès dans une perspective d'auditabilité

MA Chabin_Archive17

Aristote - 10.04.08

26

L'OAIS en relais de MoReq2

1. L'obsolescence technologique et la **pérennisation** sont abordées dans les exigences non fonctionnelles de MoReq (chapitre 11.7)
 - une quinzaine d'exigences seulement (sur 800)
 - mention rapide et globale des solutions de migration et d'émulation
 - renvoi à ISO 18492:2005 – Conservation à long terme de documents sous forme numérique
 - allusion succincte à l'OAIS

MA Chabin_Archive17

Aristote - 10.04.08

27

Moreq2 à la rencontre de l'OAIS ?

- périmètres documentaires disjoints
- objectifs distincts
- public (utilisateurs) différents
- modèles proches
- technologies convergentes
- clients communs

une question de curseur ?

MA Chabin_Archive17

Aristote - 10.04.08

28

Consulter MoReq et MoReq2

http://ec.europa.eu/transparency/archival_policy/moreq
<http://www.moreq2.eu>

Guide « Apprivoiser MoReq pour archiver et conserver l'information »

<http://www.fntc.org>
<http://www.archive17.fr>

MA Chabin_Archive17

Aristote - 10.04.08

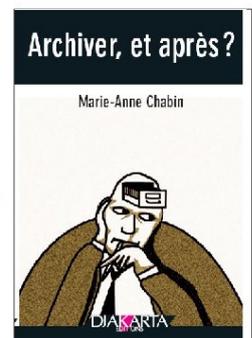
29

Une réflexion d'ensemble

Archiver, et après,

Marie-Anne Chabin
 éd. Djakarta, 150 p
 2007

3. Archiver pour quoi faire ?
4. Pour qui ?
5. A quel prix ?



MA Chabin_Archive17

Aristote - 10.04.08

30

Marie-Anne Chabin, *Archive 17*

marie-anne.chabin@archive17.fr



2.7 L'évolution de l'offre des tiers archiveurs en réponse à l'évolution des besoins

Vincent Jamin (Société ORSID)

Séminaire ARISTOTE
10 Avril 2008

L'évolution de l'offre des tiers archiveurs

Vincent JAMIN – Directeur Général Pôle Dématérialisation

ORSID

GRUPE LA POSTE

Sommaire

- ORSID
- L'évolution du tiers archivage
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- L'évolution du marché
 - L'évolution des offres
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID

Sommaire

- **ORSID**
- L'évolution du tiers archivage
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- L'évolution du marché
 - L'évolution des offres
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID

ORSID qui sommes nous ?

- Opérateur de services, créé en 1972
- Deux métiers autour du document :
 - Editique
 - Tiers archiveur
- Filiale de Doc@Post, groupe La Poste depuis 2006
- 40M €, 430 collaborateurs.

ORSID

Sommaire

- ORSID
- **L'évolution du tiers archivage**
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- L'évolution du marché
 - L'évolution des offres
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID

L'évolution du tiers-archivage (1/4)

- Environnement légal
- Environnement technologique
- Attentes du marché

ORSID

L'évolution du tiers-archivage (2/4)

L'environnement légal

- Evolutions permanentes :
 - Directives européennes
 - Lois, décrets,
 - Instructions DGI
 - ...
- Ecrit électronique : 13 Mars 2000
- Facture électronique : 20/12/2001 ; 11/01/2007
- Comptabilité : 24/01/2006
- Traçabilités : LSF; ...

ORSID¹

L'évolution du tiers-archivage (3/4)

L'environnement technologique

- Matériels :
 - Supports (ex : microfilm, CDs)
 - Puissances (avènement du numérique)
- Logiciels :
 - Compatibilités (standards, ...)
- Protocoles (langages, formats,...)
 - Standards, popularité
 - < 20 ans

ORSID¹

L'évolution du tiers-archivage (4/4)

Les attentes du marché

- Contraintes de conservation et traçabilités croissantes
 - Judiciarisation de la société
- Disposer de la bonne information au bon moment
 - Masse documentaire exponentielle
 - *dématérialisation*
 - Optimisation des process
 - *productivité*

ORSID¹

Sommaire

- ORSID
- L'évolution du tiers archivage
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- **L'évolution du marché**
 - L'évolution des offres
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID¹

Sommaire

- ORSID
- L'évolution du tiers archivage
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- **L'évolution du marché**
 - **L'évolution des offres**
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID¹

L'évolution des offres (1/4)

- Création des archives
- Conservation des archives
- Accès aux archives

ORSID¹

L'évolution des offres (2/4)

- Création des archives électroniques :
 - Dématérialisation
 - Conversions
 - Indexation, signature, horodatage
- Conservation :
 - Stockage sécurisé
 - Tiers-archivage à vocation probatoire :
 - Supports optiques (NF Z42-013; label FNTC)
 - Coffres-forts numériques

ORSID¹

L'évolution des offres (3/4)

- Accès aux archives :
 - Via Web
 - Intégration croissante avec les process métiers :
 - Distribution multi-canal
 - Gestion de dossiers clients; dossiers salariés
 - Facturation :
 - Échanges EDI
 - Communication papier (transpromo)
 - Call center, recouvrement

ORSID¹

L'évolution des offres (4/4)

Le contexte normatif

- Norme NF Z 42-013
 - Supports non réinscriptibles
 - Formats & normes d'enregistrement
 - Horodatage...
- Fédération Nationale des Tiers de Confiance
 - Créée le 16 janvier 2001 : Aproged, Conseil National des Greffiers de Tribunaux de Commerce, Conseil Supérieur de l'Ordre des Experts-Comptables, Chambre Nationale des Huissiers de Justice, Ialta, ORSID, ...
 - Travaux en cours : éthique, bonne pratique...

ORSID¹

Sommaire

- ORSID
- L'évolution du tiers archivage
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- L'évolution du marché
 - L'évolution des offres
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID¹

L'évolution des acteurs

- Tiers archiveur
- Archiveur papier
- Editeurs, constructeurs de solutions
- Prestataires d'encaissement
- Prestataires éditiques

ORSID¹

Sommaire

- ORSID
- L'évolution du tiers archivage
 - Environnement légal
 - Environnement technologique
 - Attentes du marché
- L'évolution du marché
 - L'évolution des offres
 - L'évolution des acteurs
- Exemple ORSID
 - La stratégie
 - L'évolution des offres

ORSID¹

Exemple ORSID : l'évolution des offres

- En 1999 :
 - Deux métiers séparés : éditique, archivage
 - Archivage :
 - Micrographie
 - GED
- Aujourd'hui :
 - Tiers archivage à vocation probatoire (chaîne de confiance)
 - Supports CD
 - Coffres-forts numériques
 - Hébergement internet
 - Hébergement tous flux (entrants, sortants, datas)
 - Coffre numérique
 - Gestion de dossiers clients/salariés
 - Création des archives in-situ
 - Workflow (facturation fournisseur,...)
 - Mixte éditique/dématérialisation :
 - Pilotage dynamique de la production via des bases hébergées
 - Mariage des flux entrants/sortants vers éditique

ORSID⁺

Exemple ORSID : Stratégie

- Etre en phase avec le marché :
 - Ecoute des enjeux clients
 - Présence active dans les organisations professionnelles
 - Ecoute des enjeux des métiers connexes
- Capacité technologique :
 - Veille technologique « just-in-time »
 - Maîtrises techniques internes
 - Evolution des compétences
- Disposer d'offres solides et novatrices :
 - S'appuyer sur des standards (interopérabilité, popularité)
 - Partenariat long terme
 - Réactivité face au marché

ORSID⁺

Synthèse

- Le **tiers-archivage** a subi, subi et continuera à subir des **transformations majeures** :
 - Evolutions de l'environnement légal
 - Evolutions des technologies
 - Evolutions des attentes du marché
- Les tendances actuelles continueront :
 - Chaîne de confiance
 - Importance croissante des technologies
 - Intégration croissante dans les process métiers

.....Métier passionnant !!.....

ORSID⁺

2.8 PILAE : vers une offre complète en composants réutilisables pour l'archivage

Alain Borghesi et Bruno Ricci (Société Cecurity.com)

Conservation
&
Sécurité

Security.com

Compliance
&
Security

Après PILAE : l'impact d'un grand projet sur
l'évolution de l'offre d'un éditeur de progiciel

Séminaire Aristote
Pérennisation de l'information numérique

10 Avril 2008

Security.com Sommaire de la présentation

2

- Security.com, acteur de l'archivage électronique
 - Pourquoi l'archivage électronique ?
 - Métiers & domaines de compétence
 - « Roadmap » du progiciel : vision de l'éditeur, intérêt pour l'utilisateur
- PIL@E et Coffre-fort électronique Communicant (CFEC)
 - Périmètre initial
 - Spécificités et enseignements du projet
 - Impact de PIL@E sur l'évolution du progiciel CFEC
- Les autres facteurs d'évolution du progiciel
 - L'environnement normatif
 - Perception des marchés connexes et politique de partenariat

Tous droits réservés © 2008 Security.com

Security.com Pourquoi l'archivage électronique ?

3

- Les évolutions technologiques, juridiques et économiques rendent possibles (et nécessaires) la dématérialisation
- La dématérialisation des flux et des documents crée la nécessité d'une conservation électronique sécurisée des stocks
- Les flux et les stocks dématérialisés peuvent être gérés dans un univers technologique commun
 - Signature électronique, Empreinte d'intégrité, Horodatage, Certificat électronique, Chiffrement
- Création de Security.com et apparition du concept de Coffre-fort électronique

Tous droits réservés © 2008 Security.com

Security.com Métier

4

Éditeur du logiciel Coffre-fort électronique Communicant solution opérationnelle depuis 7 ans

Avec la triple garantie :

- Conservation Sécurisée des documents immatériels
Certification et valeur probante de la Signature Électronique
- Serveur de preuves
Production, conservation et administration des preuves signées
- Protection des données personnelles (Privacy)
Compétence globale: techniques, juridique, marketing, organisationnelle



Réelle avance technologique

→ Soutenue par **oséo** **anvar**

Tous droits réservés © 2008 Security.com

Security.com Assurer la protection des données par la traçabilité et l'archivage légal

5



Tous droits réservés © 2008 Security.com

Security.com « Roadmap » du progiciel : vision de l'éditeur, intérêt pour l'utilisateur

6

- Les évolutions du progiciel qui aboutissent à une nouvelle version sont fonction
 - Attentes du marché
 - Demandes des clients
 - Environnement normatif
 - Évolutions technologiques
- Pour les nouvelles fonctionnalités
 - Arbitrer entre l'utilité, la différenciation, le délai de réalisation
 - Ne pas surestimer les capacités d'absorption de nouvelles fonctionnalités par le marché
 - Penser à la compatibilité ascendante et descendante
- L'utilisateur bénéficie du risque pris par l'éditeur, de sa vision du futur exprimée dans la « Roadmap »

Tous droits réservés © 2008 Security.com

Security.com Plateforme PIL@E et progiciel Coffre-fort électronique Communicant de Security.com 7

Plate-forme PIL@E

Développements spécifiques	
Interface Homme Machine Demande de transfert Transfert Recherche et consultation Commande	Progiciel Coffre-fort électronique Communicant - Authentification - Empreinte - Horodatage - Signature électronique - Stockage, réplication, haute disponibilité - Journalisation probatoire
Protocole Standard d'Échange (PSE) Gestion des métadonnées Moteur de recherche Validation & conversion de formats Administration Fonctionnelle Technique	

Tous droits réservés © 2008 Security.com

Security.com Spécificités du projet PIL@E 8

- Principales différences avec les autres références de Security.com (secteur privé, collectivités)
 - Besoins fonctionnels à la fois plus importants et moins précis
 - Importance du domaine archivistique par rapports aux habituelles préoccupations des DSI
 - Environnement de projet plus ouvert
 - Ex : le Comité de Pilotage du projet ne se limite pas au seul « client »
- Principales différences entre la réponse initiale et le déroulement du projet
 - Sous-estimation du temps nécessaire à l'IHM
 - Les cas d'utilisation du PSE ne représentent qu'une partie des besoins
 - Ajout du volet vérification et conversion des formats

Tous droits réservés © 2008 Security.com

Security.com Enseignements du projet PIL@E 9

- Principales difficultés rencontrées dans le cadre du projet
 - Concilier la dimension innovante et l'ampleur du projet
 - Évolution du périmètre initial en fonction des retours utilisateurs
 - Spécifications fonctionnelles détaillées itératives
 - Nouveaux domaines à maîtriser
 - volet documentaire, l'archivistique secteur public
- Quelle appropriation de l'OAIS ?
 - Facilitée par l'orientation du cahier des charges
 - Simplicité au niveau de la modélisation des données
 - Paquets d'information, composition des paquets
 - Mise en œuvre partielle de la séparation des serveurs
 - Il ne s'agit pas d'implémenter OAIS mais bien le PSE (qui repose sur OAIS)

Tous droits réservés © 2008 Security.com

Security.com L'intégration du Protocole Standard d'Échange 10

Offre initiale de Security	Utilisation API (<i>Applications Programming Interface</i>) HTTP/POST pour l'ensemble des échanges avec le CFEC	
Projet PIL@E	Éléments spécifiques au contexte	
	Éléments Rediffusables	Standard d'échange de données pour l'archivage électronique (PSE) de la Direction des Archives de France et de la direction générale de la modernisation de l'État (DGME)
Nouvelle Offre Security	API enrichie avec : <ul style="list-style-type: none"> Support transactions complexes Support modèle de données 	

Tous droits réservés © 2008 Security.com

Security.com Vérification et conversion des formats d'archivage 11

Offre initiale de Security	<ul style="list-style-type: none"> Contrôle des format des fichiers archivés embryonnaire (extension des fichiers) Pas de conversion de formats 	
Projet PIL@E	Éléments spécifiques au contexte	<ul style="list-style-type: none"> Contrôle, voire conversion, systématique pour tous les dépôts Volonté d'apporter une réponse pour les formats les plus complexes
	Éléments Rediffusables	<ul style="list-style-type: none"> L'apport du RGI et des documents DAF sur les formats cibles Modules unitaires de contrôle / conversion majoritairement open source
Nouvelle Offre Security	<ul style="list-style-type: none"> Ajout des modules contrôle et conversion activables en option pour certains formats (bureautiques, images) Ajout fonctionnalités de paramétrage de ces modules Traçabilité des opérations de conversion 	

Tous droits réservés © 2008 Security.com

Security.com Enrichissement des méta-données 12

Offre initiale de Security	Majoritairement des métadonnées techniques : <ul style="list-style-type: none"> Date/heure, Empreinte, Contexte dépôt (session/déposant), etc. 	
Projet PIL@E	Éléments spécifiques au contexte	<ul style="list-style-type: none"> Gestion communicabilité / confidentialité pilotée par des conventions Gestion de l'héritage et surcharge
	Éléments Rediffusables	<ul style="list-style-type: none"> Notion de contexte (dans une modélisation simplifiée)
Nouvelle Offre Security	<ul style="list-style-type: none"> Intégration d'un module paramétrable de prise en compte des métadonnées Plusieurs modélisation des métadonnées (PSE, Toolbox DublinCore) 	

Tous droits réservés © 2008 Security.com

Ceconomy.com Amélioration de la dimension recherche

13

Offre initiale de Economy.com	<ul style="list-style-type: none"> Recherche sur les métadonnées techniques & politique de nommage Complémentarité avec solutions GED et progiciels de gestion d'archives en amont 	
Projet PIL@E	Éléments spécifiques au contexte	<ul style="list-style-type: none"> Couplage de la recherche avec modèle de données arborescent Gestion de l'héritage et surcharge
	Éléments Rediffusables	<ul style="list-style-type: none"> Mise en œuvre du moteur de recherche Open Source Lucène
Nouvelle Offre Economy.com	<ul style="list-style-type: none"> Intégration d'une brique de recherche plus évoluée reposant sur Lucène Pertinence accrue par l'enrichissement de la dimension métadonnées 	

Tous droits réservés © 2008 Economy.com

Ceconomy.com PIL@E et évolution du CFEC : les zones de non recouvrement

14

- Les dimensions de PIL@E qui n'entreront pas à court terme dans l'évolution du progiciel de Economy.com
 - Workflow entre services versants et services d'archives
 - La prise en charge manuelle des versements au standard d'échange
- Les évolutions du progiciel de Economy.com qui ne présentent pas de liens directs avec le projet PIL@E
 - Amélioration de la dimension Chiffrement des données
 - Chiffrement du système ou chiffrement des données
 - Chiffrement transparent pour l'utilisateur
 - Vérification des documents signés électroniquement
 - Refonte de l'interface utilisateur de type « Web 2.0 »

Tous droits réservés © 2008 Economy.com

Ceconomy.com L'environnement normatif

15

- OAIS
 - Diffusion par intégration des apports de PIL@E au CFEC
- MOREQ
 - Participation aux actions de sensibilisation de la FNTC
- Z42-013
 - Participation à la refonte comme membre du Comité éditorial
 - Une innovation majeure :
 - Prise en compte « des systèmes disposant de supports réinscriptibles pour lesquels l'intégrité des archives et la sécurité des processus sont assurées par des moyens cryptographiques ».
 - Un aboutissement prévu en 2008
 - Formaliser l'apport des technologies issues de la signature électronique pour l'archivage électronique

Tous droits réservés © 2008 Economy.com

Ceconomy.com Marchés connexes et partenariats

16

- Gestion Électronique de Documents et Content Management**
 - Les grands acteurs du marchés ECM : de la vente de matériel à la maîtrise de l'information
 - Economy.com partenaire d'éditeurs de GED (Archimed, Digttech, ...) et de grands constructeurs pour le volet infrastructures (IBM pour PIL@E)
- Prestations de Tiers-Archiver**
 - Pas de dichotomie simpliste externalisation vs internalisation
 - Les archiveurs papier deviennent aussi tiers-archivers
 - Economy.com, fournisseur de technologie pour des tiers-archivers
 - Partenariat fort avec Orsid (Groupe La Poste)
- Progiciels de Gestion informatisée des services d'archives**
 - Intérêt du marché pour un système intégré destiné à la gestion des archives papier et électronique
 - Economy.com, partenaire de Sicem, éditeur du progiciel THOT ; première référence significative en 2008

Tous droits réservés © 2008 Economy.com

Ceconomy.com Pour conclure

17

- PIL@E : une étape important pour l'évolution stratégique de la société
 - De la sécurisation des données à l'archivage électronique sécurisé
 - Par le renforcement de la culture « archivage » au niveau de la culture « électronique sécurisé »
- Nouveau questionnement sur l'archivage électronique
 - Au-delà de la pérennité la question de la souveraineté

Tous droits réservés © 2008 Economy.com

Ceconomy.com Merci de votre attention

18

75 rue Saint Lazare
75 009 Paris
SAS au capital de 738 000 €
RCS Paris B 434 330 338

www.ceconomy.com

Tél : 01 56 43 37 37

contact@ceconomy.com



Tous droits réservés © 2008 Economy.com

2.9 SPAR : l'analyse et les choix majeurs du système logiciel

Charlotte Fabre, Laurent Sollier (Société ATOS Origin)




Le projet SPAR : Réponses applicatives aux choix fonctionnels
Système de Préservation et d'Archivage Réparti

Charlotte Fabre (AO) / Laurent Sollier (AO)
Séminaire Aristote, Groupe PIN, 10 avril 2008




Sommaire

- ▶ Présentation générale
- Versement
- Stockage
- Gestion des données
- Gestion des droits
- Accès
- Administration
- Préservation

2 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale



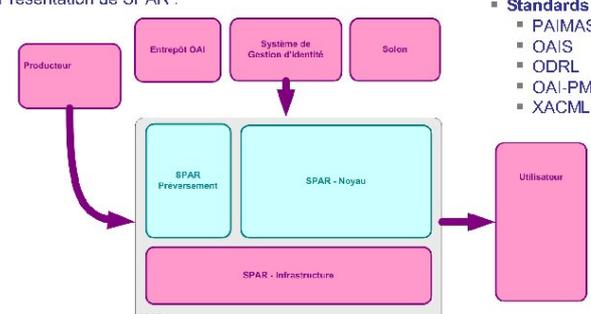

- **Contexte:**
 - La Bibliothèque nationale de France souhaite mettre en place un système assurant la pérennité de son information numérique. Volumétrie attendue : 1800 To d'ici 2014.
- **L'équipe Atos Origin** : 10 personnes
 - Chef de projet
 - Concepteurs techniques
 - Concepteurs fonctionnels
 - Développeurs...
- **Démarche modulaire**
 - **Tranche Ferme** : Création du noyau composé de modules et d'une chaîne
 - Numérisation de conservation chaîne A
 - **Tranche Conditionnelle** : Création des différentes filières
 - Dépôt légal du Web
 - Production administrative, Record management
 - Tiers archivage

3 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale




» Présentation de SPAR :



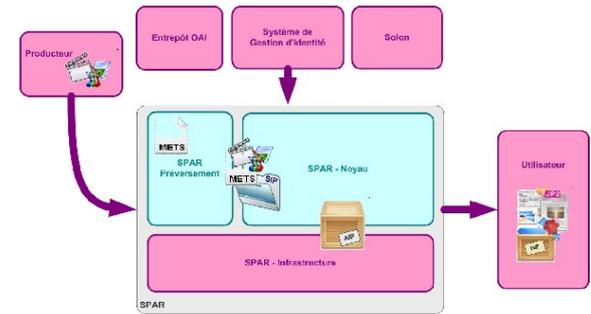
- **Standards:**
 - PAIMAS
 - OAIS
 - ODRL
 - OAI-PMH
 - XACML

4 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale



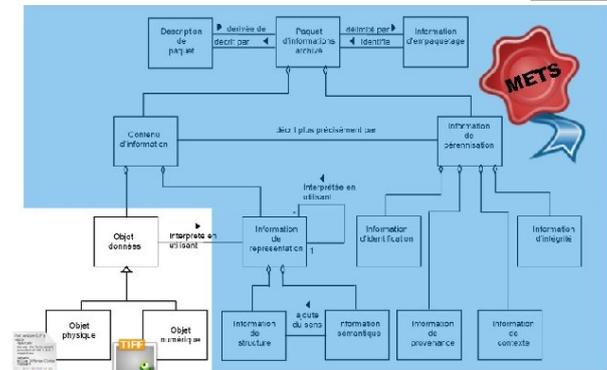

» Présentation de SPAR :



5 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale



6 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF
Atos Origin

Sommaire

- Présentation générale
- ▶ **Versement**
- Stockage
- Gestion des données
- Gestion des droits
- Accès
- Administration
- Préservation

13 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF
Atos Origin

Présentation générale

Versement

- » **Fonctionnalités clés**
 - Import de paquets SIP
 - Des différentes filières
 - Des descriptions de référence
 - Gestion des mises à jour
 - De nouveaux fichiers d'un paquet
 - De modification des fichiers d'un paquet
 - Contrôles
 - De l'intégrité
 - Du format du fichier
 - Du METS (.xml)
 - Génération d'un identifiant unique et construction d'un AIP
 - Envoi aux modules
 - Gestion de données
 - Stockage
- » **Mise en œuvre**
 - Les frameworks utilisés :
 - JHOVE
 - axiS
 - NOID
 - Struts²
 - schexation
 - Les développements spécifiques :
 - Enrichissement du manifeste METS
 - Appel des services de gestion de données et de stockage
 - Gestion de la transaction (ACID)

14 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF
Atos Origin

Sommaire

- Présentation générale
- Versement
- ▶ **Stockage**
- Gestion des données
- Gestion des droits
- Accès
- Administration
- Préservation

15 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF
Atos Origin

Présentation générale

Stockage

- » **Fonctionnalités clés**
 - Enregistrement d'un AIP
 - Recherche de la classe de service
 - Dialogue avec le service d'abstraction de stockage
 - Stockage de l'AIP et envoyer un AR
 - Audit d'un AIP
 - Vérification que le paquet existe
 - Contrôle de l'intégrité des fichiers
 - Récupération d'un AIP
 - Réception d'une demande du module Accès
 - Récupération asynchrone
 - Fourniture d'une copie de l'AIP
- » **Mise en œuvre : stockage / Service d'abstraction de stockage**
 - Les frameworks utilisés :
 - opensolaris
 - jargon
 - TRIPS
 - Les développements spécifiques :
 - Gestion de la transaction (ACID)

16 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF
Atos Origin

Sommaire

- Présentation générale
- Versement
- Stockage
- ▶ **Gestion des données**
- Gestion des droits
- Accès
- Administration
- Préservation

17 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF
Atos Origin

Présentation générale

Gestion des données

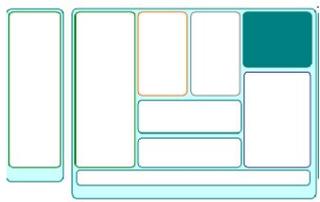
- » **Fonctionnalités clés**
 - Requête sur les métadonnées
 - Requête simple et synchrone
 - Requête profonde (toutes les MD)
 - Enregistrement des métadonnées
 - Mise à jour des index
 - Récupération des métadonnées
 - Demande du module Accès
 - Fourniture des MD
- » **Mise en œuvre**
 - Les frameworks utilisés :
 - mulgara (RDF)
 - Fedora Commons
 - Les développements spécifiques :
 - Gestion de la transaction (ACID)

18 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF Atos Origin

Sommaire

- Présentation générale
- Versement
- Stockage
- Gestion des données
- ▶ Gestion des droits
- Accès
- Administration
- Préservation



19 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale

Gestion des droits

» Fonctionnalités clés

- Collecte d'information de droits auprès de SOLON (syst. de gestion des droits de la BnF)
 - Métadonnées de droit
 - Arbres de décision et conventions
 - Mise à jour
- Génération d'une licence
 - Récupération des MD pour un DIP
 - Parcours de l'arbre ou de la convention
 - Génération de la licence



ODRL

» Mise en œuvre

- Les frameworks utilisés :

woodstox

axis

Proal

- Les développements spécifiques :
 - Gestion de la transaction (ACID)

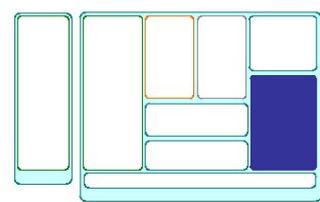


20 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF Atos Origin

Sommaire

- Présentation générale
- Versement
- Stockage
- Gestion des données
- Gestion des droits
- ▶ Accès
- Administration
- Préservation



21 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale

Accès

» Fonctionnalités clés

- Export d'un DIP
 - Synchrone
 - Gérer le cache de DIP
 - Asynchrone
 - Créer le DIP à l'aide des autres modules et le fournir
- Recherche dans les métadonnées
 - Liste de résultats d'ID de paquets
- Collecte des métadonnées
 - Récupération de MD pour un paquet



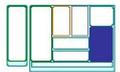
» Mise en œuvre

- Les frameworks utilisés :

woodstox

Struts²

- Les développements spécifiques :
 - Appel des services de gestion de données et de stockage
 - Gestion de la transaction (ACID)

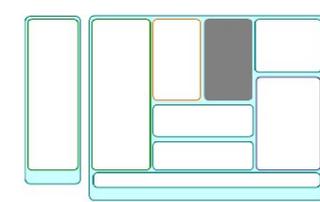


22 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

BnF Atos Origin

Sommaire

- Présentation générale
- Versement
- Stockage
- Gestion des données
- Gestion des droits
- Accès
- ▶ Administration
- Préservation



23 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale

Administration

» Fonctionnalités clés

- Gestion et fédération des habilitations
- Description des descriptions de référence
 - Description des AQS
 - Description des processus
- Planification
 - De l'Accès
 - De l'audit de Stockage
 - De la reconstruction des MD
 - De la planification de la migration
- Suivi d'un plan
- Surveillance du système
- Export des données d'administration
- Export de données financières

» Mise en œuvre

- Les frameworks utilisés :

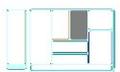
Struts²

LOG4J

OpenLDAP

Quartz Enterprise Job Scheduler

- Les développements spécifiques :
 - Journalisation des actions
 - ...



24 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Sommaire

- Présentation générale
- Versement
- Stockage
- Gestion des données
- Gestion des droits
- Accès
- Administration
- **Préservation**

25 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale

■ Préservation

- » Fonctionnalités clés
 - Description
 - Des informations de représentation
 - Des paquets de transformation
 - De la programmation de la migration
 - Collecte d'une information de représentation auprès d'un répertoire de format
 - Suivi d'une programmation de migration
- » Mise en oeuvre
 - Les frameworks utilisés :
 - Struts²**
 - Proai
 - Quartz Enterprise Job Scheduler
 - OAIHarvester2
 - Les développements spécifiques :
 - Journalisation des actions
 - ...

26 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Présentation générale

- » Répartition des frameworks :

27 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Questions / Réponses

28 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

WORLDWIDE IT PARTNER

Pour plus d'informations, contacter :

Charlotte Fabre Laurent Sollier
 charlotte.fabre@atosorigin.com Laurent.sollier@atosorigin.com

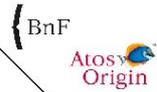
Atos Origin
www.atosorigin.fr

29 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Annexes

- » Liens standards
 - OAIS http://rds.cnes.fr/pin/documents/norme_oais_version_francaise.pdf
 - PAIMAS http://rds.cnes.fr/pin/documents/030115_CCSDS_651_R1.pdf
 - METS <http://www.loc.gov/standards/mets/>
 - OAI-PMH <http://www.openarchives.org/OAI/openarchivesprotocol.html>
 - ODRL <http://odrl.net/>
 - XACML <http://xml.coverpages.org/xacml.html>
 - Dublin Core <http://dublincore.org>

30 Le projet SPAR : Réponses applicatives aux choix fonctionnels – Présentation séminaire Aristote – Groupe PIN – 10 Avril 2008

Annexes

* Liens outils

- Fedora
<http://www.fedora-commons.org/>
- iROds
https://www.irods.org/index.php/irods:Data_Grids%2C_Digital_Libraries%2C_Persistent_Archives%2C_and_Real-time_Data_Systems
- Jhove
<http://hul.harvard.edu/jhove/>
- Schematron
<http://www.schematron.com/>
- OpenLDAP
<http://www.openldap.org/>
- NOID
<http://www.cdlib.org/inside/diglib/noid/>
- Proai
<http://sourceforge.net/projects/proai/>
- Mulgara
<http://www.mulgara.org/>
- SAM
<http://blogs.sun.com/samqfs/>

<http://www.aristote.asso.fr>

Contact : info@aristote.asso.fr

ARISTOTE Association Loi de 1901. Siège social : CEA-DSI CEN Saclay Bât. 474, 91191 Gif-sur-Yvette Cedex.
Secrétariat : Aristote, École Polytechnique, 91128 Palaiseau Cedex.
Tél. : +33(0)1 69 33 99 66 Fax : +33(0)1 69 33 99 67 Courriel : Marie.Tetard@polytechnique.edu
Site internet <http://www.aristote.asso.fr>