The Adonis research data preservation project for digital humanities in France
What is Adonis?

The aim of the Adonis project is to provide computing, data storage, data access and data management facilities to the French research community on Digital Humanities, based on a shared infrastructure.

It should also promote and coordinate high level application software development dedicated on digital humanities content treatment.

Adonis was created 2 years ago by the CNRS. The reflection really started 1 year ago.
The CNRS is the National Center for Scientific Research. It is a national institute and it covers all scientific research activities from digital humanities to HEP.

~25000 collaborators working on as scientists, engineers, techniciens and administratives.

It works in coordination with the french universities and also with 5 other specific national institutes (CEA, IFREMER, INRA, INRIA, INSERM).
Research on digital humanities:
- Ministry of education and research
- Ministry of culture

Digital humanities data are spread over:
- CNRS
- Universities
- Archive of France (national, regional, …)
- Museums (national, regional, …)
- Libraries (BNF, …)

Hopefully, Adonis is only concerned by research data!
European contacts

Similar activities are starting in other European countries and some projects are done at European Community level like DARIAH, CLARIN, …

Adonis is involved in these European projects.  
  hope to have national infrastructures that can exchange informations
The Adonis project is based on the network of the ~12 Numerical Resources Centers (CRN) and 2 computing centers (CC-IN2P3 & CINES). The CRN are dedicated to specific data subjects and studies.

These CRN’s have limited resources and can not manage long term preservation:

- externalization of the data to computing centers (200-300 TB)
- use a shared infrastructure.
- concentration on high level application for data content treatment
CC-IN2P3 & CINES in Adonis

• The CC-IN2P3:
  • Data management, data access, data storage
  • Computing
  • Web servers, databases, …
• The CINES:
  • long term data preservation (format, physical support)
  • data format check & metadata check
Infrastructure architecture

- CMS
- Data validation

- Workflow ?
- Authentication ?
- Access rights ?
- Scalability ?

Distributed working space (fedora-commons)

Distributed working space (iRods)

Long term preservation system
Technologies used

• Numerical object level:
  • fedora-commons (web services, api’s)
• File level:
  • iRods & i-commands
  • iRods api’s (java, php)
  • java & windows iRods browsers
• Authentication:
  • LDAP ? Shibboleth ? GSI ? …
  • Authentication transmission ?
• Access rights management:
  • LDAP ? RDB ? …
• Video streaming

We need performances at each level.
Thank you