

Nikolaos Kasioumis

CERN







BlogForever



European project (FP7)

March '11 – August '13

12 Partners







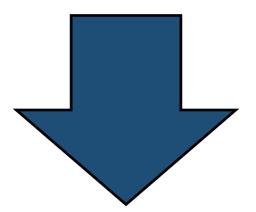
- Blogs: well established dynamic media type
- No existing solution for blog archiving







- Blogs: well established dynamic media type
- No existing solution for blog archiving

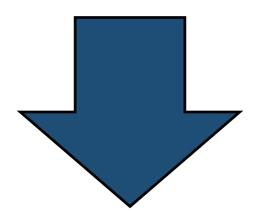








- Blogs: well established dynamic media type
- No existing solution for blog archiving



- Robust digital preservation, management and dissemination facilities for blogs
- Study the structure and semantics, define a digital preservation policy, implement a blog digital repository and specific case studies









Bloggers

Unstructured information

Web services, Blog APIs



- Real-time weblog monitor
- HTML data extraction engine
- Spam and noise filtering
- Web services extraction engine



Weblog posts,

Comments

Metadata (Content, DRM, Technical)

Semantic information,

Microformats

Associations between different blogs, posts, users, etc.



- Digital preservation
- Data migration
- Quality assurance checking
- Collections curation
- · Public access APIs

- Personalized services
- Data mining, memetracking
- Information retrieval
- Public Web Interface, Browse, search, export









Need for a tailored digital repository software







- Need for a tailored digital repository software
- Extend and adapt the widely used Invenio







- Need for a tailored digital repository software
- Extend and adapt the widely used Invenio
 - Fully featured web portal
 - OAIS compatibility
 - Long term digital preservation
 - Digital Rights Management (DRM)
 - Spam filtering
 - Billing system







Currently

- Ingestion, METS and OAIS compatibility
- User interface and functionality improvements and additions







Currently

- Ingestion, METS and OAIS compatibility
- User interface and functionality improvements and additions

Future

- Scalability
- Digital Rights Management (DRM)
- Extended indexing



