

Preserving Websites Of Research & Development Projects

Daniel Bicho
daniel.bicho@fccn.pt

Daniel Gomes
daniel.gomes@fccn.pt

Research Project: ARchive COmmunities MEMories



ABOUT ARCOMEM ▼

TECH DEMOS

SYSTEM DEMOS

TRAINING

OPEN SOURCE

VIDEOS

PUBLICATIONS

ABOUT ARCOMEM

ARCOMEM is about memory institutions like archives, museums, and libraries in the age of the Social Web. Memory institutions are more important now than ever: as we face greater economic and environmental challenges we need our understanding of the past to help us navigate to a sustainable future. This is a core function of democracies, but this

<http://www.arcomem.eu/>

Information about scientific events



[ABOUT ARCOMEM ▾](#) [TECH DEMOS](#) [SYSTEM DEMOS](#) [TRAINING](#) [OPEN SOURCE](#) [VIDEOS](#) [PUBLICATIONS](#)

IPRES 2013




10th International Conference on
Preservation of Digital Object

1st International Workshop on Archiving Community Memories
6 September 2013, Lisbon, Portugal

www.arcomem.eu | twitter.com/arcomem | [#arcomem](https://twitter.com/arcomem)

Dissemination and training materials




[ABOUT ARCOMEM ▾](#) [TECH DEMOS](#) [SYSTEM DEMOS](#) [TRAINING](#) [OPEN SOURCE](#) [VIDEOS](#) [PUBLICATIONS](#)

TRAINING


To explain the ARCOMEM project to detail we have created online training material so that the end-users of the ARCOMEM system can train themselves in understanding and using the AROMEM system.

ARCOMEM MADE EASY

As a beginner you can start with having a look at the "ARCOMEM made Easy" presentation. This presentation shows you a general overview on how the ARCOMEM system has been set up. If you want to dig deeper you can follow the links to the Powerpoint presentations on the different modules of the ARCOMEM system.



Demonstrations about the project



ABOUT ARCOMEM ▼ TECH DEMOS **SYSTEM DEMOS** TRAINING OPEN SOURCE VIDEOS PUBLICATIONS

SYSTEM DEMOS

Memory institutions like archives, museums, and libraries are more important now than ever. But our understanding of the past faces stiff new challenges in face of the Social Web. History is not only written in books anymore. The radical changes in information creation, communication and citizen involvement that currently characterise our information society (e.g., there are now more social network hits than Google searches) need a [new approach on preservation](#). Social media are becoming more and more pervasive in all areas of life. But how do we separate the wheat from the chaff in Social media, how is it being preserved and will it still be available twenty years from now?

ARCOMEM will enable political archivists and journalists to preserve Social media around certain entities, topics or events. To give you a quick insight [here is a short movie](#) on how journalism can benefit from Social media archiving.

We would like to share our work with you so please come and have a look at our system demos! We have installed

Content is not available anymore

404
PAGE NOT FOUND

Waste of significant investments

European Union FP7 Work Programme invested **59 million euros in R&D projects.**

Part of this funding was **spent** developing R&D projects websites.



Loss of Knowledge


R&D project sites publish important scientific outputs (e.g. data sets, tools).

Work as aggregators of project outputs (ex. news and events).

ARCOMEM NEWS

ARCOMEM @ SAMOS SUMMIT 2013
Workshop E: The ARCOMEM project platform The ARCOMEM...
[READ MORE](#)

ARCOMEM WORKSHOP ON SOCIAL MEDIA SEMANTICS (SMS2013) AT OTM 2013
The Social Web has become the first and main medium...
[READ MORE](#)

FEATURE STORY

HELLENIC PARLIAMENT ORGANISES ARCOMEM WORKSHOP

**Web archives can preserve this
valuable information.**

Web Archives **need to identify project URLs** to preserve their content.

Where to get project URLs?

[Sitemap](#) | [Legal notice](#) | [Contact](#) English (en) ▼

 **European Union Open Data Portal**

EUROPA > Open Data Portal > Data > Publisher > Publications Office > CORDIS - EU research projects un...

Data Applications Linked Data Developers' corner About 🔒 Data provider's area Share

CORDIS - EU research projects under FP7 (2007-2013)

Publisher
[Publications Office »](#)

Description

This dataset contains projects funded by the European Union under the seventh framework programme for research and technological development (FP7) from 2007 to 2013. Grant information is provided for each project, including reference, acronym, dates, funding, programmes, participant countries, subjects and objectives. You can also find a separate file with organisation information (project participants, coordinators) and another with results in brief written by CORDIS for selected FP7 projects.

Reference data (countries, funding schemes/types of action, subjects (SIC codes)) can be found in this dataset: <https://data.europa.eu/euodp/en/data/dataset/cordisref-data>

eurovoc domains
[International relations](#), [Production](#), [technology and research](#)

Resources

[DOWNLOAD](#) FP7 Projects [CSV](#)

[DOWNLOAD](#) FP7 Projects [EXCEL](#)

Licence:
[Legal Notice](#)

EuroVoc concepts

- [technical cooperation](#)
- [EU research policy](#)
- [research and development](#)
- [scientific exchange](#)
- [cooperation policy](#)
- [research](#)
- [Framework Programme for Research and Development](#)
- [scientific cooperation](#)

Catalogue record

Added to data.europa.eu/euodp
2015-07-29

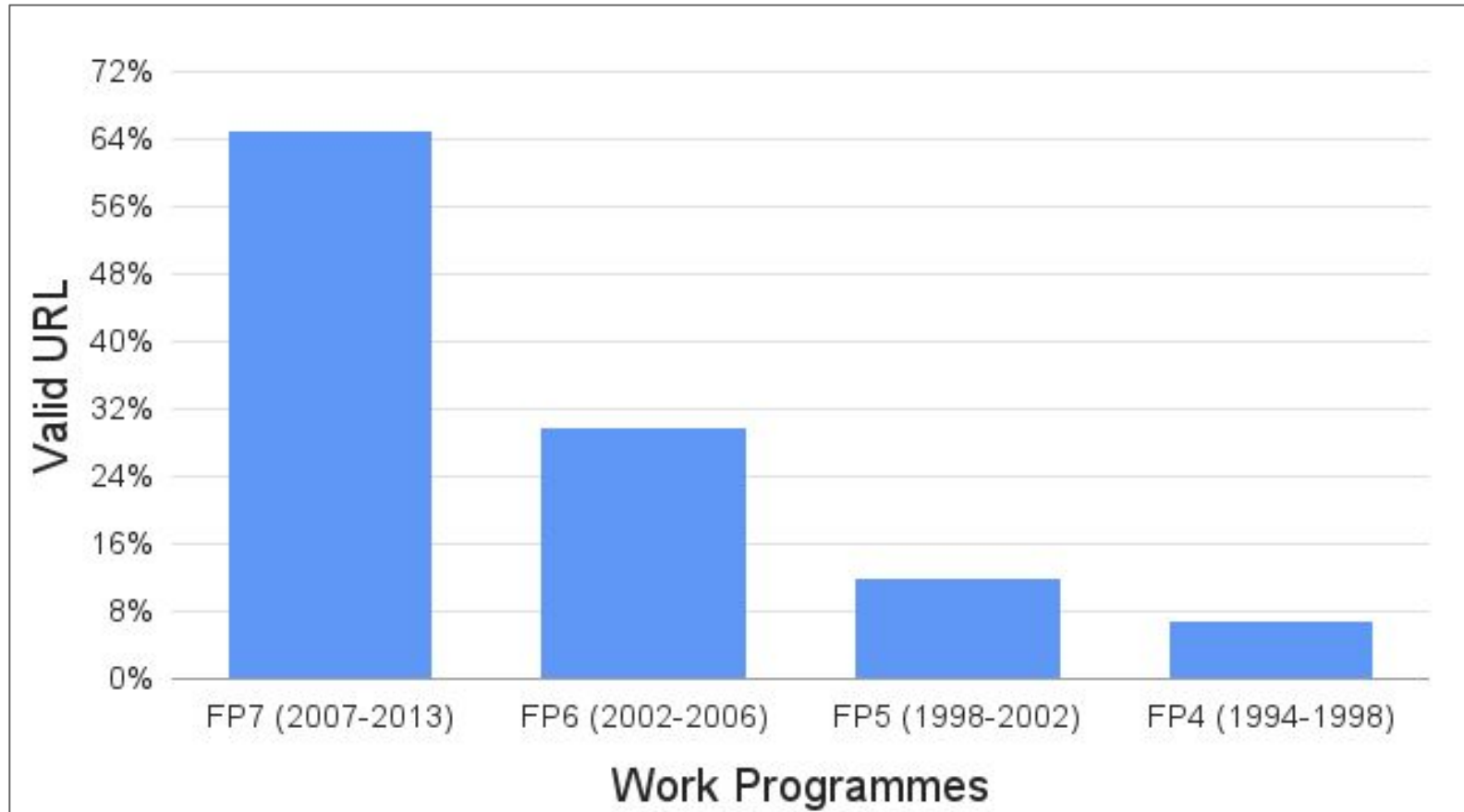
<https://open-data.europa.eu/>

European Open Data Portal Datasets

CORDIS EU research projects funded under work programmes (FP4, FP5, FP6, FP7).

Include information like **project URL**, acronym, title, dates, funding, objectives etc.

Valid project URLs in 2016



How much of these projects are really being preserved?



BRISK

Biofuels Research Infrastructure for Sharing Knowledge

[Home](#)
[Partners](#)
[Facilities](#)
[Funding](#)
[Activities](#)
[BRISK Events](#)
[Conferences etc.](#)
[Publications](#)
[News](#)
[Contact](#)
[Partner area](#)

BRISK is a research infrastructure. Its main activity is to fund researchers from any EU country to carry out research at any of the 26 partners' facilities.

The partners are listed here
 The facilities available are listed here
 Details on how to apply for funding are provided here

Other activities in BRISK

Dissemination
 Protocols, Databases and Benchmarking
 New methodologies for feedstock characterisation
 Advanced measurement methods and procedures in thermo-chemical biomass conversion
 Improving methods for testing and optimisation in catalytic conversion processes
 Beyond BRISK

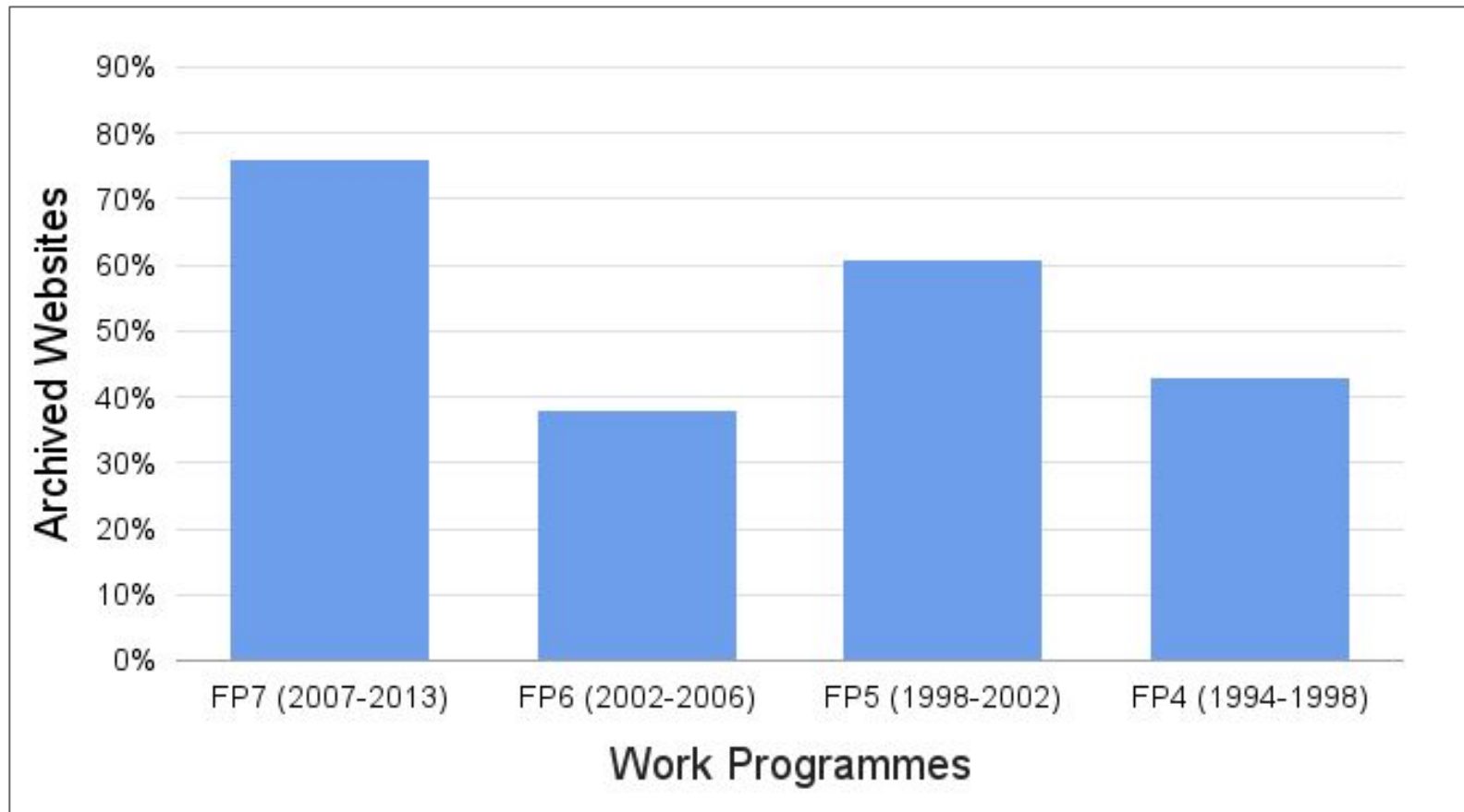




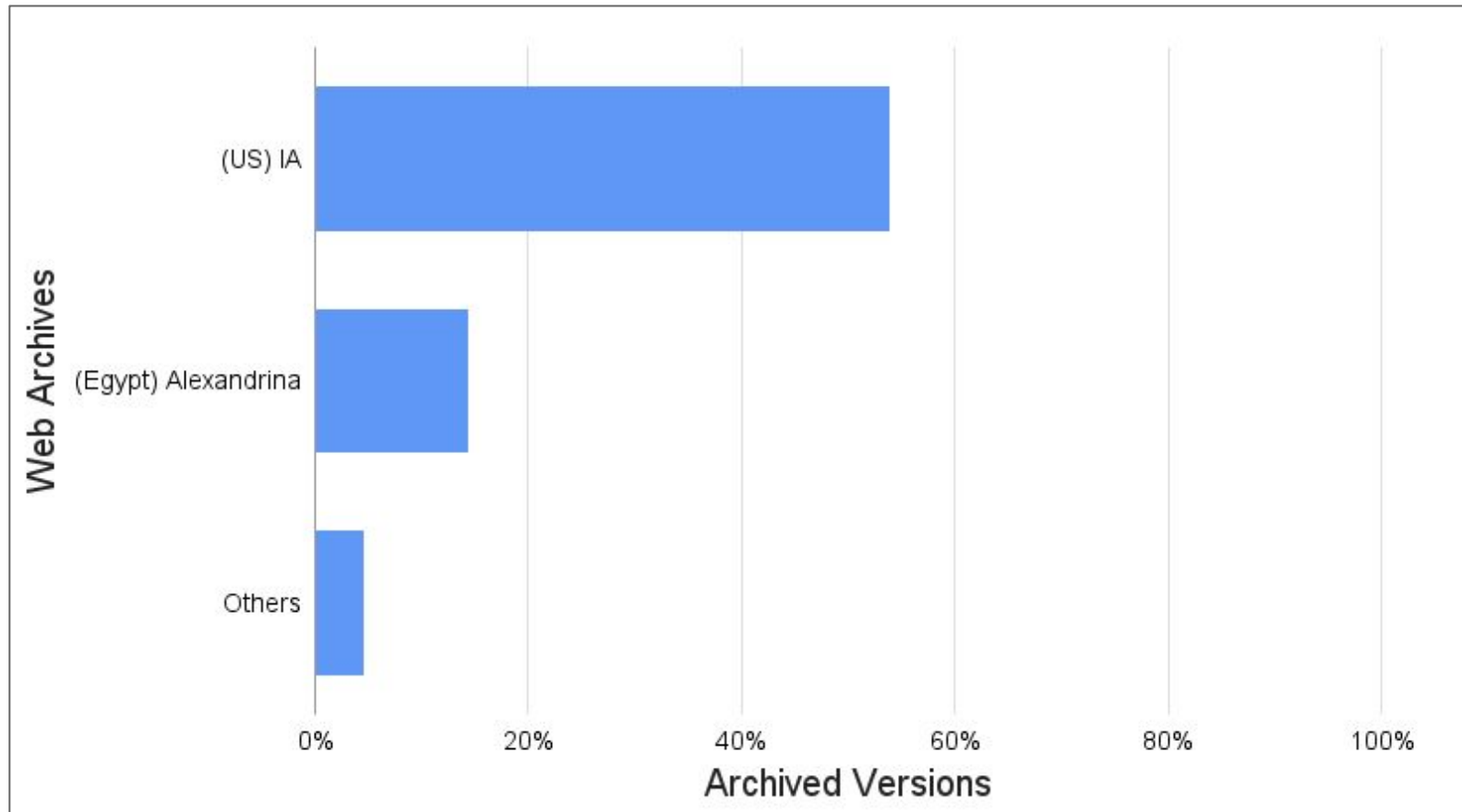


<http://timetravel.mementoweb.org/>

Web-archived project URLs since FP4



Project URLs of EU-funded research are mostly being preserved outside Europe



CORDIS datasets provided **incomplete information**

25 000 projects were funded by the FP7 work programme.

Only **8%** have an associated **project URL**.

Missing information regarding **project URL**

acronym	title	projectUrl
ALFRED	ALFRED - Personal	
TIBETMETH	Microbial Biomark	
SMALL_MAM_RECOL	Post-glacial reco	
MOMEFAST	Molecular Mecha	
RNF4 IN THE DDR	Identifying the targ	
ARIEL	Archaeological Inve	
NANODYNATCELLVATIO	Nano -structural a	
TBKO	Synthesis and Bio	
MICROBIOELECTROSYN	cathode in bioelect	
THINFACE	Thin-film Hybrid In	
IMPACTS	The impact of the	
FLEXISTAT	Production Flexibil	
3FLEX	Depth enabled wo	
GUIDENANO	Assessment and m	
AZNETAC	A zebrafish model	
PLASMANANOSMART	Plasma- and electr	
RASMIM	Reactivity of Alum	
FAMILIESANDSOCIETIES	Changing families	http://www.familiesandsocieties.eu/
ACTICOSPACK	Reducing the level	
MULTIMERMAID	Development of a	
GAGAUTOSYN	Automated Glycos	
COMPOSITES 4 IU	Grant to develop se	
MCEC	Ministerial Confere	
BACTINSECT	Molecular mecha	
SOLAR FUEL BY III-VS	Direct photoelectro	
FIBRILLATION	The structure-base	

Problem

R&D websites provide **important** information but quickly **disappear**.

Funded projects datasets are **incomplete**.

Thousands of projects are too many to be humanly identified.

Our study:

How to **automatically** identify
project URLs to be preserved
(with **limited resources**)?

Proposed approach to automatically identify project URLs

Search Engines already index the Web.

Open Data Portal provides meta-data about R&D projects.

Combine open data sets with search engines to identify R&D project URLs.

Automatic workflow to identify R&D project URLs

ACRONYM	TITLE	PROJECT URL	ACRONYM	TITLE	PROJECT URL
DIP3	The 3Ps of Distributed Information delivery ...	missing	DIP3	The 3Ps of Distributed Information delivery ...	www.dip-3.eu



Identify project URL
through Bing Web Search
API

Web search by project Acronym +Title



FAMILIESANDSOCIETIES Changing families and sust 

Web | Imagens | Vídeos | Os meus itens guardados

1 630 RESULTADOS | Idioma ▼ | Região ▼

Families And Societies Traduzir esta página
www.familiesandsocieties.eu ▼
... Changing families and sustainable societies: ... ©FamiliesAndSocieties.
Multidisciplinary approach: We combine a wide range of expertise in social sciences, ...

FamiliesAndSocieties - Changing families and sustainable ...
www.age-platform.eu/images/stories/EN/FamiliesAndSocieties... · Ficheiro PDF
FamiliesAndSocieties - Changing families and sustainable societies: Policy contexts
and diversity over the life course and across generations

Experiment

Evaluate heuristics to **maximize** the performance of the automatic identification of R&D project URLs.

+Acronym +Title

+Acronym +Title -Cordis

+Acronym +Title -Cordis -EC

+Acronym +Title -Cordis -EC +CommonTerms

Evaluating the heuristics to automatically identify project URLs

A test collection was built based on the FP7 dataset by **manually** validating each R&D project URL.

The project URLs returned by each heuristic were compared to the test collection to measure if they **matched**.

ACRONYM	TITLE	PROJECT URL
FAMILIEANDSOCIETIES	Changing families and ...	www.familiesandsocieties.eu

Heuristics Performance (F-measure)

Heuristics	Top 1 Results
+Acronym +Title	44%
+Acronym +Title -Cordis	45%
+Acronym +Title -Cordis -EC	47%
+Acronym +Title -Cordis -EC +project	48%

*F-measure is a combination of **recall** and **precision**.*

Preserving R&D projects websites

Applied heuristic with the best performance to the CORDIS Dataset.

23 588 missing project URLs.

20 429 new project URLs were automatically identified.

Before: 23 588 missing URLs

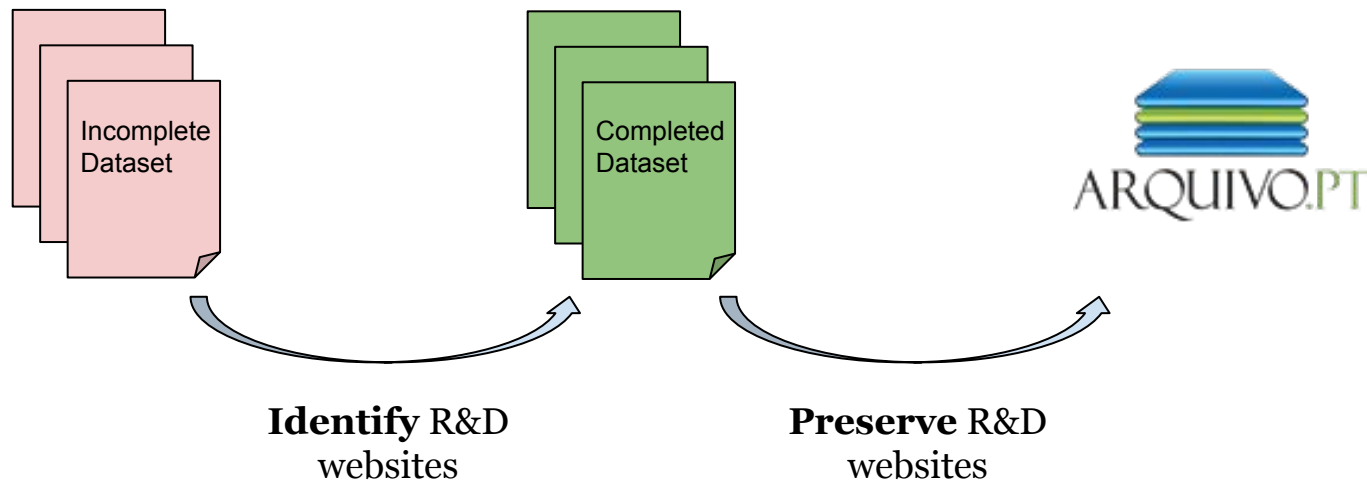
acronym	title	projectUrl
ALFRED	ALFRED - Person	
TIBETMETH	Microbial Biomar	
SMALL_MAM_RECOL	Post-glacial reco	
MOMEFAST	Molecular Mecha	
RNF4 IN THE DDR	Identifying the ta	
ARIEL	Archaeological ir	
NANODYNATCELLVATION	Nano -structural	
TBKO	Synthesis and Bi	
MICROBIOELECTROSYN	Microbially cataly cathode in bioele	
THINFACE	Thin-film Hybrid I	
IMPACTS	The impact of the	
FLEXISTAT	Production Flexit	
3FLEX	Depth enabled w	
GUIDENANO	Assessment and	
AZNETAC	A zebrafish mode	
PLASMANANOSMART	Plasma- and elec	
RASMIM	Reactivity of Alu	
FAMILIESANDSOCIETIES	Changing familie	http://www.familiesandsocieties.eu/

acronym	title	projectUrl
ALFRED	ALFRED - Person	http://alfred.eu/
TIBETMETH	Microbial Biomar	http://www.thefreedictionary.com/Tib
SMALL_MAM_RECOL	Post-glacial reco	http://www.thefreedictionary.com/Ins
MOMEFAST	Molecular Mecha	
RNF4 IN THE DDR	Identifying the ta	http://www.dundee.ac.uk/research/m
ARIEL	Archaeological ir	http://ucy.ac.cy/ariel/
NANODYNATCELLVATION	Nano -structural	
TBKO	Synthesis and Bi	
MICROBIOELECTROSYN	al systems	
THINFACE	Thin-film Hybrid	http://www.nanogune.eu/projects/th
IMPACTS	The impact of th	http://www.sciencedirect.com/science
FLEXISTAT	Production Flexi	http://www.flexistat.eu/
3FLEX	Depth enabled w	http://www.hhi.fraunhofer.de/departm
GUIDENANO	Assessment and	http://www.guidenano.eu/News
AZNETAC	A zebrafish mod	
PLASMANANOSMART	Plasma- and elec	http://portal.tpu.ru/departments/cent
RASMIM	Reactivity of Alu	https://www.highbeam.com/doc/1G1
FAMILIESANDSOCIETIES	Changing familie	http://www.familiesandsocieties.eu/


After: 20 429
new URLs

Crawled the identified project URLs to be preserved.

Nr. Project URL Seeds	20 429
Stored Content (compressed)	1.4 TB




Project Websites Preserved

ALFRED

HOME NEWS PROJECT PARTNERS DISSEMINATION DOWNLOADS CONTACT


Interactive Assistant for Independent Living and Active Ageing

ALFRED – Personal Interactive Assistant for Independent Living and Active Ageing – is a project funded by the Seventh Framework Programme of the European Commission under Grant Agreement No. 611218. It will allow older people to live longer at their own homes with the possibility to act independently and to actively participate in society by providing the technological foundation for an ecosystem consisting out of four pillars:




User-Driven Interaction Assistant

Empowering people with age related dependencies to live independently for longer by delivering a virtual butler with seamless support for tasks in and outside the home. The virtual butler ALFRED will have a very high end-user acceptance by using a fully voice controlled and non-technical environment.




Personalized Social Inclusion

Fostering active participation in society for the ageing population by suggesting and managing events and social contacts.



Effective & Personalized Care

Improved care process through direct access to vital signs for carers and other medical staff as well as alerting in case of emergencies. The data is collected by unobtrusive wearable sensors monitoring the vital signs of older people.



Physical & Cognitive Impairments Prevention

Prevailing age-related physical and cognitive impairments with the help of personalized, serious games.

Pages


Home
News
Partners

Links

Downloads
Dissemination
Subscribe Newsletter

Copyright

© ALFRED 2014
ALFRED – Personal Interactive Assistant
for Independent Living and Active Ageing





WEather haZARDS for aeronautics

☆ Home

☆ The Project

☆ The Consortium

☆ The Advisory Board

☆ Deliverables

☆ Documentation

☆ Publications

☆ Events

☆ Related projects

☆ Contact

Funded by:



European Commission



SEVENTH FRAMEWORK PROGRAMME

WEZARD Home

15 Feb 2013

Overview

The European WEZARD project (acronym standing for **Weather Hazards** for Aeronautics) aims at preparing the future research community in the area of air transport system robustness when it is faced with weather hazards. Its precise objectives are to provide: (i) an interdisciplinary and cross-sector network comprising relevant experts; (ii) a state-of-the-art review of the on-going research actions; (iii) an analysis which will identify the shortcomings, areas for improvements and the type of activity needed to limit the effects of disruptive events; (iv) a set of recommendations and a roadmap validated by the main stakeholders of the aeronautics community.




The WEZARD consortium consists of 3 airframers, 2 engine manufacturers, 1 system supplier, 1 network of meteorological offices, 4 research centers, 1 provider of test facilities and 1 civil aviation authority over 2 years. An Advisory Board gathering a panel of international experts in relevant domains has been set up to provide advice on the vision, priorities and directions proposed by the project.



The project runs for 24 months from July 2011 until June 2013.

Investigation Institutes



Quantum Optics and Spectroscopy

Institut für Experimentalphysik - Universität Innsbruck



[Home](#) [News](#) [People](#) [Research](#) [Contact](#) [Publications](#) [Open Positions](#) [Teaching](#) [Links](#)

[Home](#)
[About](#)
[News](#)
[People](#)
[Research](#)
 [2D Arrays](#)
 [Barium](#)
 [Cavity QED](#)
 [CryoTrap](#)
 [LinTrap](#)
 [Nanofiber](#)
 [PhiNet](#)
 [Precision](#)
 [QuIQS](#)
 [Rydberg](#)
 [SQIP](#)
 [Ultrafast](#)

Realization of Trapped Strontium Rydberg Ions (QuaSIRIO)


Note: The Rydberg Ions Experiment has moved to the Stockholm University.

In this project we focus on the realization of trapped Rydberg ions. This will bring together two quantum systems: trapped ions and Rydberg atoms. Joining them will form a novel quantum system with advantages from both sides. Trapped Rydberg ions are a promising system for applications in quantum information processing and quantum simulation.


What are trapped Rydberg ions?

Rydberg ions are orders of magnitude bigger than normal ions. They possess an outermost electron which resides far away from the atomic core so that it is almost not trapped anymore. The picture below demonstrates the dimension of a Rydberg ion with respect to other important dimensions in an ion trap. The size of a Rydberg ion can be bigger than its localization inside the trap, but will still be smaller than the typical ion-ion distance. The resulting physical properties make trapped Rydberg ions a promising system for quantum information processing. For instance, due to their size they should be able to sense other Rydberg ions across distances of several micrometers which would enable fast quantum operations.


Ion




Ion motion



Rydberg ion



Ion distance



On going work

Use the methodology to identify and preserve national funded projects.

Adapt the methodology to identify websites related to the Portuguese elections.

Conclusions

R&D websites are **important** but are quickly **disappearing**.

European Datasets about funded R&D projects are **incomplete**. Only **8%** have a project URL associated.

54% of the EU-funded R&D project URLs are being web-archived. Mostly **outside the European Union**.

Automatic heuristics to identify R&D project URLs.

Conclusions

All outputs of this study are available in open access

- Test collection
- Completed CORDIS data sets with new project URLs
- Extended technical report

arquivo.pt/publications

github.com/arquivo: Research-Websites-Preservation