

# Preserving the .EU domain

Daniel Bicho  
daniel.bicho@fccn.pt

Arquivo.pt is a research infrastructure for the study of web archived material.

  [Advanced search](#)

## Search pages from the past

Millions of contents archived since 1996

[Meet the service](#)

# What is .EU domain?

- **.EU** is the top-level domain for the European Union (EU).
- Launched on 7 December 2005, the domain is available for organisations and residents of EU member states.
- There are more than 3.9 million registered .eu domain names. <https://www.eurid.eu/en/about-us>

# Why preserve .EU domain?

- An **European Identity** on the Internet for European citizens.

# European Identity

**Yanis Varoufakis**

thoughts for the post-2008 world

 Search



*"One very simple, but radical, idea: to democratise Europe."*

Home About Europe in Crisis Global Crisis Digital Economies/Money Greek Implosion Books Radio  
Videos&TV PRESS TALKS ESSAYS PPE Ελληνικά Contact Header Photo ΥΠΟΨΗΦΙΟΤΗΤΑ 2015

<https://yanisvaroufakis.eu/>

# Why preserve .EU domain?

- Used by the European Union **institutions** (europa.eu).

# European Commission



European Commission

English 

Search

HIGHLIGHT

Managing the refugee crisis

Jobs, Growth and Investment

Digital Single Market

Energy Union and Climate

Internal Market

EU-US Free Trade

Justice and Fundamental Rights

Migration

EU as a Global Actor

<http://ec.europa.eu/>

 ARQUIVO.PT

# European Parliament



The header of the European Parliament website features a dark blue background. At the top left, there is a language selector set to 'en'. To the right, a navigation menu includes links for News, MEPs, About Parliament, Plenary, Committees, Delegations, At your service, EPTV, and More. Below the navigation, the European Parliament logo is displayed on the left, and a banner on the right reads 'The European Parliament is in session' with a yellow arrow pointing to 'Strasbourg' and an image of the parliament building. A row of icons represents various sections: News, Video, Think tank, Committees, Plenary, MEPs, Organisation, and Parliament and you.



A row of social media icons for Facebook, Twitter, YouTube, LinkedIn, Instagram, Pinterest, and YouTube. To the right is a 'Sign up for email updates' button with an envelope icon.

Search on "European Parliament" 



The main content area is divided into three columns. The left column, titled 'News', shows 'Top stories' with a large image of a blue screen displaying white binary code. The middle column, titled 'The President of the European Parliament', features a portrait of David Sassoli and links to the Website, Facebook, and Twitter, along with an RSS icon. The right column, titled 'Live Broadcast', shows a '08:30 Plenary session: Strasbourg' with a red 'ON AIR' button.



# Why preserve .EU domain?

- European Research and Development projects

# Thorax project

The screenshot shows the THORAX project website homepage. At the top left is the THORAX logo, a blue circle with a stylized 'L' and 'O' inside, followed by the word 'THORAX' in blue. To the right of the logo is the tagline 'Thoracic injury assessment for improved vehicle safety' in a smaller blue font. Further right is a portion of the European Union flag. Below the header is a navigation bar with links: 'About THORAX', 'Consortium', 'News', 'Events', 'Downloads', 'Contact', and 'Links'. To the right of this bar are buttons for 'HOME', 'SEARCH', 'SITEMAP', and 'LOGIN'. The main content area is divided into several sections. On the left, there are three images: a car interior with a crash test dummy, a 3D anatomical diagram of the human thorax, and a 3D model of a crash test dummy. The central section features the title 'THORAX Final Workshop' and a paragraph: 'The THORAX project partners will present their results in the final workshop at the end of the project.' Below this is the date and location: 'Thursday 25<sup>th</sup> of April 2013, Bergisch-Gladbach, Germany at teh BAST.' A red link 'More information' is provided. Below the text is a small image of a car interior with a crash test dummy. To the right of this image is the text 'Co-funded under 7th FP (Seventh Framework Programme)' and the logos for the 'SEVENTH FRAMEWORK PROGRAMME' and the European Union. On the far right, there is an 'Events' section with the title 'THORAX Final workshop' and the date '25 April 2013' and location 'BAST, Bergisch-Gladbach, Germany'. A red link 'More info' is provided. Below this is a small image of a crash test dummy sitting on a chair. At the bottom right of the events section is a red link 'View all events »'.

**THORAX** Thoracic injury assessment for improved vehicle safety

HOME SEARCH SITEMAP LOGIN

About THORAX Consortium News Events Downloads Contact Links

**THORAX Final Workshop**

The THORAX project partners will present their results in the final workshop at the end of the project.

Thursday 25<sup>th</sup> of April 2013, Bergisch-Gladbach, Germany at teh BAST.

**More information**

Co-funded under 7th FP (Seventh Framework Programme)

SEVENTH FRAMEWORK PROGRAMME

**Events**

**THORAX Final workshop**  
25 April 2013  
BAST, Bergisch-Gladbach, Germany

**More info**

[View all events »](#)

<http://www.thorax-project.eu/>

# Neuroscience project

The screenshot shows the Neurobid website interface. At the top left is the Neurobid logo (a blue circle with yellow stars) and the text "Neuroscience on Barriers in Development". On the right, there are font size controls (A+, A, A-) and a search bar. The main content area is titled "General" and contains a "Summary" section. The summary text discusses the prevalence and costs of brain diseases in Europe, mentioning WHO data and the European Brain Council findings. It highlights the blood-brain barrier (BBB) as a major barrier to treatment. On the left side, there are two menu sections: "MAIN MENU" and "CONSORTIUM MENU". On the right side, there is a "LOGIN FORM" with fields for Username and Password, a "Remember Me" checkbox, and a "LOGIN" button. Below the login form is a "WHO'S ONLINE" section showing "We have 1 guest online". At the bottom right, there is a logo for "COOPERATION" featuring a stylized number 7.

Neuroscience on Barriers in Development

Home > About NEUROBID

## General

### Summary

Brain disease, encompassing both psychiatric and neurologic disease, represents a major public health problem in Europe and worldwide. It is one of the most prevalent and costly groups of diseases in Europe, with **estimated total annual costs (2004 data) amounting to 386 billion Euro or 829 Euro per individual living in Europe (1)**. Data collected by the World Health Organization (WHO) suggest that brain diseases are responsible for 35% of Europe's total disease burden (2). An analysis of all epidemiological and health economic studies in Europe published by the European Brain Council in 2005 found that across 28 European countries (the European Union plus Iceland, Norway, and Switzerland) with a total population of 466 Million, **127 million (or 27%) of Europeans**, children and adults at all stages, are affected by at least one form of brain disease (1). Improved prevention and treatment strategies are paramount in the attempt to reduce the individual and societal burdens associated with cerebral palsy, multiple sclerosis, stroke, schizophrenia and the many other devastating disorders that involve the central nervous system (CNS). In this scenario, prevention and treatment options rely heavily upon optimized pharmacologic access to the CNS. The major barriers between the systemic and CNS compartments are the blood-brain parenchyma barrier (microvessels) and blood-cerebrospinal fluid barrier (choroid plexus). Throughout this project, we refer to both collectively as the blood brain barrier (BBB) unless directing a specific research topic at only one of these interfaces.

**MAIN MENU**

- Home
- About NEUROBID
  - Concept
  - Objectives
- News
- Publications
- Contact

**CONSORTIUM MENU**

- MHH
- INSERM
- UMC Utrecht
- UOXFAV
- Simcyp
- UNIMELB
- UGOT
- Brain-i
- ABT
- UKW

**LOGIN FORM**

Username

Password

Remember Me

**LOGIN**

- [Forgot your password?](#)
- [Forgot your username?](#)

**WHO'S ONLINE**

We have 1 guest online

**COOPERATION**

# Not preserved website

Making SME meat processors more competitive by using automatic in-line fat analysis in meat trimmings



Home

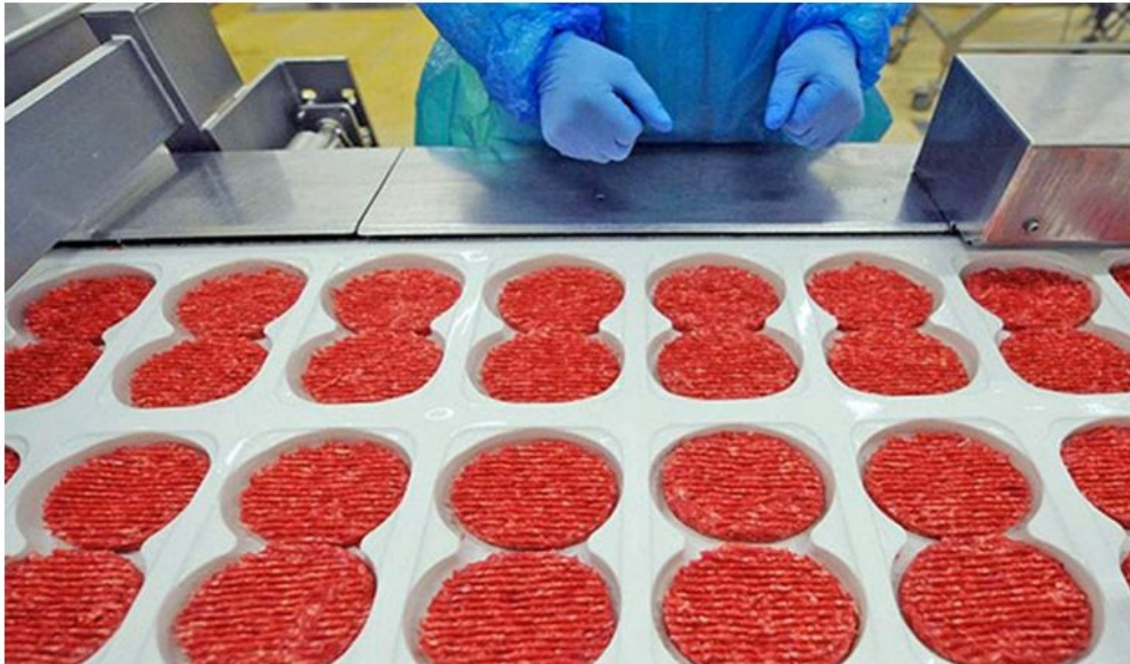
Project

News

Partners

Contact

Partner's area



## Automatic in-line fat analysis in meat trimmings

The **TRIMSCAN** solution consists in the development of an in-line fat analyser to scan pork and beef trimmings contained in plastic bins. The solution uses a contactless system to predict the fat content with an expected accuracy below 1.5 %.

**TRIMSCAN** will provide SME meat processors with a practical tool at an affordable cost to make them more competitive by reducing lean meat over-use and to avoid the need of inaccurate and time consuming manual sampling and recipe reformulation.

**TRIMSCAN** will significantly improve the production of processed meat by homogenising the quality of the final product and by reducing the percentage of rejections due to products that do not meet the legislation and/or customer of specifications

[\[Read More\]](#)

<http://trimscan.eu>

# Not preserved website



`http://trimscan.eu/`

Latest

Show All

Hrm.

Wayback Machine doesn't have that page archived.

[https://web.archive.org/\\*/http://trimscan.eu/](https://web.archive.org/*/http://trimscan.eu/)

# Who preserves the .EU domain?

- National institutions preserve national domains.
- Which institution preserves the .EU domain?
- This can lead to the loss of European Cultural Heritage for future generations.

# RESAW

RESAW is an European network that aims to create a **Research Infrastructure** for the Study of Archived Web Materials established in late 2012.

**Arquivo.pt** is one of the Web Archiving initiatives participating in RESAW.

International and National Projects.

<http://resaw.eu>

# Archiving .EU project

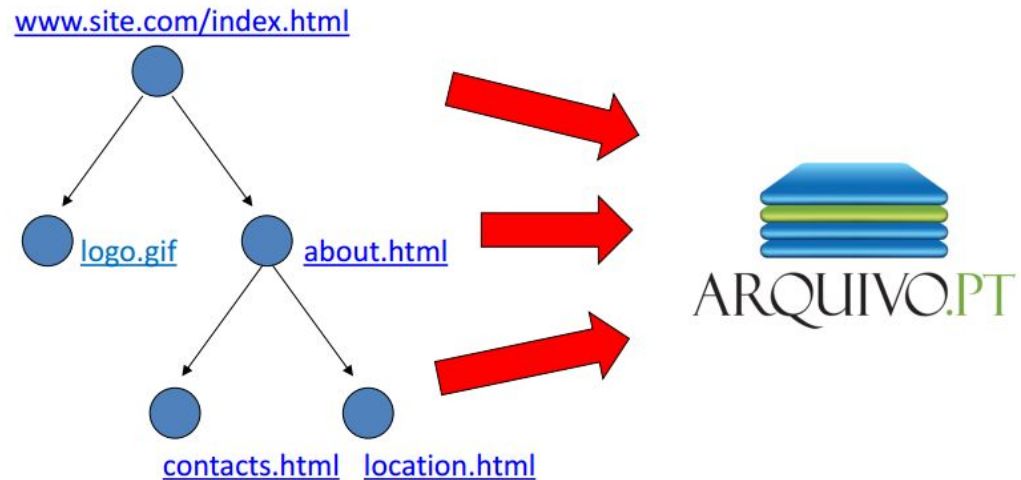
- A pilot exploratory crawl of the .EU domain.
- The objective was to gain insight about how to preserve the .EU domain.



# How a Crawler Works

Crawler is a **robot** that automatically navigates from page to page.

- Crawler gets a start page (seed URL).
- Extracts addresses available on the page.
- Follows the extracted addresses to new pages
- Repeat process for all sites.

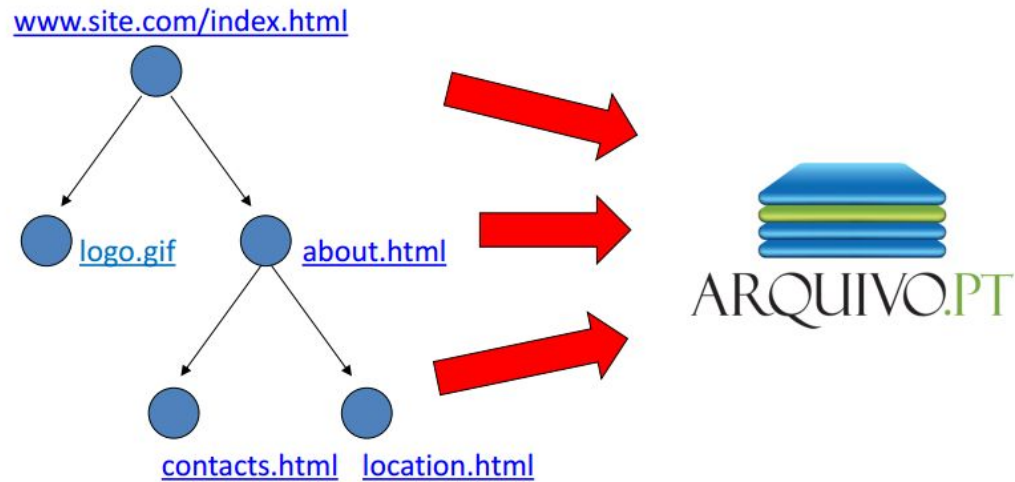


# URL Seeds

- We gathered seeds from several sources:
  - dmoz - <http://www.dmoz.org/docs/en/rdf.html>
  - Alexa Top Sites - <http://www.alexa.com/topsites>
  - HttpArchive - <http://www.httparchive.org>
  - Websiteisworth - <http://goo.gl/0if6pe>
  - Search our Crawl Logs for .eu URL addresses.
  - Google Search Queries for .eu URL addresses.
- Gathered a total of 34 thousand URL seeds.

# First crawl Configuration

- 5 links followed from homepage.
- 10 000 files per site.



# Information Crawled

- 1 million hosts.
- 250 million files.
- Crawler ran for 23 days.
- 5.8 Terabytes of **.eu** domain content stored.

# Second Crawl Configuration

- 5 links followed from homepage.
- 1 000 files per site.

# Information Crawled

- 72 thousand hosts.
- 14 million files.
- Crawler ran for 19 days.
- 3.1 Terabytes of **.eu** domain content stored.

# What we learned

- Redirects to other top level domains are common.
  - 9% of the URL seeds were redirected to other top level domains.
- Poorly designed online shops that generate infinite number of URLs. ( <http://autobazar.eu> )

# What we learned

- Lots of Web Spam.
  - Link farm sites. ( <http://in-links.eu/> )
- Sub-domains that referenced **multilingual versions**.



# Search Prototype

- We developed a URL and Full-Text Search prototype on the crawled content.
- Full-Text Search makes more easy the research for crawled contents on .EU domain.
- Publicly accessible through: <http://arquivo.pt/resawdev>

# Search Prototype

RESEARCH  
.EU

[Advanced search](#)

Search from the past

[Know the project](#)

<http://arquivo.pt/resawdev>

# Search Prototype



European Research Projects

between:   and:

[Advanced search](#)

Results 1 to 10 from 153,282

## [ERC: European Research Council | ERC: European Research Coun...](#)

**21 November, 2014** - other dates

ERC: **European Research** Council | ERC: **European Research** Council Jump to Navigation Advanced Search | Glossary / A-Z | Sitemap Home Funding and Grants **Projects** and Results Media and Events About ERC ... ERC Grants Open also to non-**European** researchers ERC News 17.11.14 ERC President speaks to **research** ... <http://erc.europa.eu/>

## [Research Projects / University of Ostrava](#)

**22 November, 2014** - other dates

**Research Projects** / University of Ostrava Home UO Portal FAQ Contacts UO Premises Site Map ... Art. Activity Main Areas of **Research** **Research Projects** Seventh Framework Programme Faculties and ... , Institute for **Research** and Applications of Fuzzy Modeling **European Research** Institute for Social Work ... <http://www.osu.eu/index.php?kategorie=671>

## [European Commission](#)

**21 November, 2014** - other dates

. **Research** and innovation funding Tenders Grants **European** structural and investment funds Microfinance ... **European** Commission Skip to main content Language български (bg) čeština (cs) dansk (da) Deutsch ... **European** Commission News Highlight 01/11/2014 The Juncker Commission A strong and experienced team ... [http://ec.europa.eu/index\\_en.htm](http://ec.europa.eu/index_en.htm)

<http://arquivo.pt/resawdev>

# Research Data

- Crawling logs and report at <http://arquivo.pt/crawlreport>.
- Full-text search .EU collection <http://arquivo.pt/resawdev>.
- Talk with us if you need more information for your research.

# Future/Current Work

- Preservation of R&D projects.
  - they publish valuable scientific outputs.
  - they vanish shortly after the project funding ends.
  - constitute a trans-national, multi-lingual and cross-field set of historical web data for researchers (e.g. social scientists).

# Open Data Portal

The screenshot shows the European Union Open Data Portal interface. At the top, there is a navigation bar with 'Sitemap | Legal notice | Contact' and a language dropdown set to 'English (en)'. Below this is the portal's logo and name. A breadcrumb trail reads 'EUROPA > Open Data Portal > Data > Search'. A secondary navigation bar contains 'Data', 'Applications', 'Linked Data', 'Developers' corner', and 'About', along with a 'Data provider's area' link and a 'Share' button.

The search results are displayed for the query 'cordis'. The search bar includes a magnifying glass icon and a search button. Below the search bar, there are radio buttons for 'Show results with: all of these words | any of these words | the exact phrase'. The results are filtered by 'Publishers: Publications Office x' and sorted by 'Most viewed'. A total of '9 datasets found' is indicated.

Three dataset entries are listed:

- CORDIS – EU research projects under Horizon 2020 (2014–2020)**  
This dataset contains projects funded by the European Union under the Horizon 2020 framework programme for research and innovation (H2020) from 2014 to 2020. Grant information is provided...  
Available in: CSV, ZIP, XLSX (9247 views)
- CORDIS – EU research projects under FP7 (2007–2013)**  
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...  
Available in: CSV, EXCEL, XML, XLSX, ZIP (5474 views)
- CORDIS – EU research projects under FP6 (2002–2006)**  
This dataset contains all projects funded by the European Union under the sixth framework programme for research and technological development (FP6) from 2002 to 2006. Grant information...  
Available in: CSV, EXCEL, ZIP (704 views)

On the right side of the page, there are three panels: 'Suggest a dataset' (with a question about missing datasets), 'EuroVoc domains' (listing categories like Production, technology and research), and 'EuroVoc concepts' (listing terms like technical cooperation).

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

# FP7 Projects List

	C	H	K	L
1	acronym	title	projectUrl	objective
2	PROVIPS	Development of an automated classification system for polished stone slabs and tiles	<a href="http://88.32.124.85/PROVIPS/">http://88.32.124.85/PROVIPS/</a>	'Limitations in quality control performance repr
3	ABC4EU	ABC GATES FOR EUROPE	<a href="http://abc4eu.com/">http://abc4eu.com/</a>	'During the last years, many ABC Gates have bee
4	ACOBAR	Acoustic Technology for observing the interior of the Arctic Ocean	<a href="http://acobar.nersc.no/">http://acobar.nersc.no/</a>	'ACOBAR will develop an observing system for t
5	ACROPOLIS	Aggregate and Cumulative Risk Of Pesticides: an On-Line Integrated Strategy	<a href="http://acropolis-eu.com/">http://acropolis-eu.com/</a>	'EU Directive 396/2005 states that methods are
6	ADAPOND	"Development of an automatic process of in-house collection, storage and application of adaptive bacteria culture for fish farms."	<a href="http://adapond.eii.ee/">http://adapond.eii.ee/</a>	'Evidence from recirculation systems shows tha
7	NANOFOODS	Development of foods containing nanoencapsulated ingredient	<a href="http://allegati.unina.it/ricerca/progr">http://allegati.unina.it/ricerca/progr</a>	'The preservation of bioactive food ingredients
8	ANTICARB	Monoclonal ANTIBody-targeted CARBOn nanobues against cancer	<a href="http://anticarb.org/">http://anticarb.org/</a>	'ANTICARB attempts to exploit the advantages ,
9	APARET	African Programme for Advanced Research Epidemiology Training	<a href="http://aparet.org/index.php?option=">http://aparet.org/index.php?option=</a>	'The APARET fellowship programme will catalys
10	AQUAVALENS	Protecting the health of Europeans by improving methods for the detection of pathogens in drinking water and water used in food preparation	<a href="http://aquavalens.org/">http://aquavalens.org/</a>	'The AQUAVALENS consortium has brought tog
11	ARCH	Agenda for research on chernobyl health (ARCH): to develop of a strategic plan for research on the health consequences of radiation from the chernobyl accident	<a href="http://arch.iarc.fr">http://arch.iarc.fr</a>	The Chernobyl accident led to the most serious

# Recovering project URLs

- 25 608 R&D projects funded by FP7 listed by the EU Open Data Portal.
- Only 7.9% had an associated project URL.
- Researching techniques to automatically identify this project URLs.