

Preservation of web content that documents Horizon 2020

Pedro.Gomes@fccn.pt



Arquivo.pt preserves national and international historical web content



Janeiro 1993

Manual de utilização

e

Manual: Elisa Prata Pina e M. Augusta Patrício
sugestões de exploração

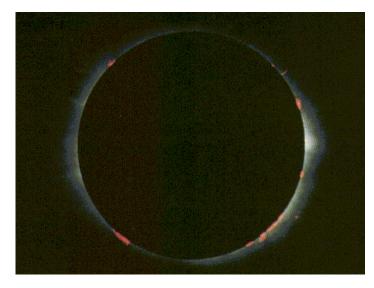
Orientação: Carlos Fiolhais

Departamento de Física da Universidade de Coimbra

ÍNDICE

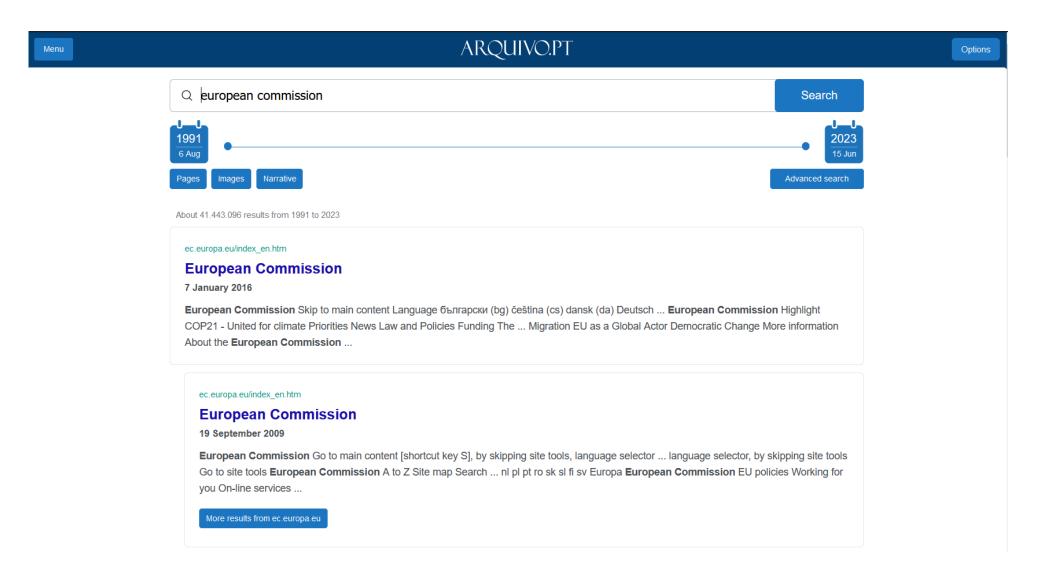
- 1. Apresentação sumária do programa
- 2. Características do equipamento
- 3. Configuração mínima do computador
- 4. Ficheiros do programa

nautilus.fis.uc.pt- 1993



spacelink.nasa.gov - 1992

Our services are publicly available



Increase our **collections** & users **community**

Acronym	Title	ProjectUrl
MARmaED	MARine MAnagement and Ecosystem Dynamics under climate change	https://www.marmaed.uio.no/
ENIW	Services to enhance innovation management capacities of SMEs in Engla	http://www.enterprise-europe.co.uk/
ARCADE	Aligning Research & Innovation for Connected and Automated Driving in	
QMiCS	Quantum Microwave Communcation and Sensing	https://qmics.wmi.badw.de
RenalToolBox	Developing novel tools and technologies to assess the safety and efficacy	
REDIFUEL	Robust and Efficient processes and technologies for Drop In renewable F	http://www.redifuel.eu
ROSSINI	RObot enhanced SenSing, INtelligence and actuation to Improve job qual	http://www.rossini-project.com
TRA VISIONS 2020	TRA VISIONS 2020	https://www.travisions.eu
NONTOX	Removing hazardous substances to increase recycling rates of WEEE, E	
ProtoMet	Protometabolic pathways: exploring the chemical roots of systems biology	
PEQEM	Photonics for engineered quantum enhanced measurement	http://www.bristol.ac.uk/physics/people/jonathar
I-MOVE-plus	I-MOVE+ Integrated Monitoring of Vaccines Effects in Europe: a platform	https://sites.google.com/a/epiconcept.fr/imovep
IMMUcan	Integrated IMMUnoprofiling of large adaptive CANcer patients cohorts	https://immucan.eu
PHOENIX	Electrical Power System □s Shield against complex incidents and extension	
AMR-TB	Theoretical and computational investigation of tuberculosis antimicrobial r	
CHINAWHITE	The Reconfiguration of Whiteness in China: Privileges, Precariousness, a	
EMP	European Microkelvin Platform	
JERICO-NEXT	Joint European Research Infrastructure network for Coastal Observatory	http://www.jerico-ri.eu
IPAD-MD	Research Infrastructures for Phenotyping, Archiving and Distribution of M	http://www.infrafrontier.eu/infrafrontier-research
ELSE - SIM	Enterprise Lazio and Sardegna for Europe Smes Innovation Manageme	http://www.eenelse.it

H2020 - cordis.europa.eu - Dataset

Previous work on preserving EU web content



A first attempt to archive the .EU domain

August 13, 2015



Arquivo.pt preserved websites about Research & Development projects funded by the EU

April 10, 2017



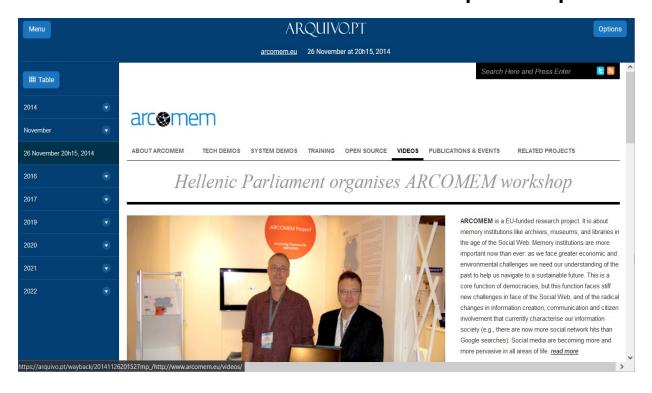
sobre.arquivo.pt > News > Cross-lingual collection about the 2019 European Elections is available

Cross-lingual collection about the 2019 European Elections is available

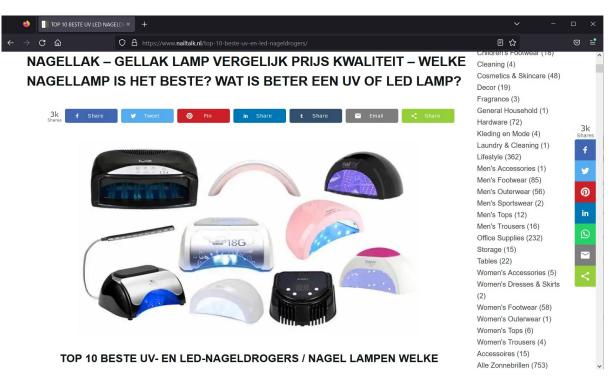
September 4, 2020

Motivation

arcomem.eu , 2014, Arquivo.pt



arcomem.eu, Live Web



H2020

Open data sets available at EU Data Portal used to identify URLs to be archived

- cordis-h2020organizations.xlsx
- cordis-h2020projectDeliverables.xlsx
- cordis-h2020reports.xlsx
- cordis-h2020projects.xlsx



cordis-h2020organizations.xlsx

projectID	projectAcronyn	role	name	shortName	country	organizationUrl
675997	MARmaED	partner	HAVFORSKNINGSINSTITUTTET	IMR	NO	www.imr.no
675997	MARmaED	partner	DANMARKS PELAGISKE PRODUCENTORGANISA	DANISH PELAG	DK	
675997	MARmaED	participant	HELSINGIN YLIOPISTO	UH	FI	www.helsinki.fi
675997	MARmaED	partner	Marine Stewardship Council		UK	
675997	MARmaED	partner	WWF DEUTSCHLAND	WWF DEUTSCH	DE	
675997	MARmaED	participant	UNIVERSITAET HAMBURG	UHAM	DE	
675997	MARmaED	partner	THE ROYAL MINISTRY OF FISHERIES AND COAS	FHF	NO	http://www.regjeringen.no/nb/dep/fkd
675997	MARmaED	partner	INTERNATIONAL COUNCIL FOR THE EXPLORATI	ICES	DK	
675997	MARmaED	partner	WWF Finland		FI	
675997	MARmaED	coordinator	UNIVERSITETET I OSLO	UNIVERSITY OF	NO	www.uio.no
675997	MARmaED	participant	DANMARKS TEKNISKE UNIVERSITET	DTU	DK	www.dtu.dk
675997	MARmaED	participant	METEO-FRANCE	MF-CNRM	FR	www.meteo.fr

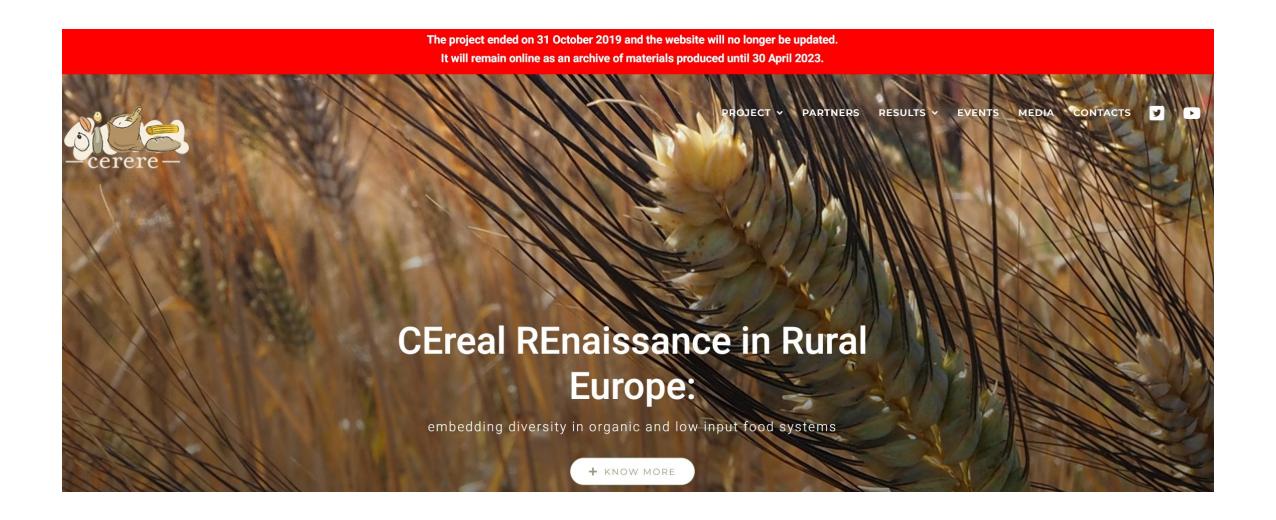
tinyurl.com/cordis-h2020organizations

cordis-h2020reports.xlsx

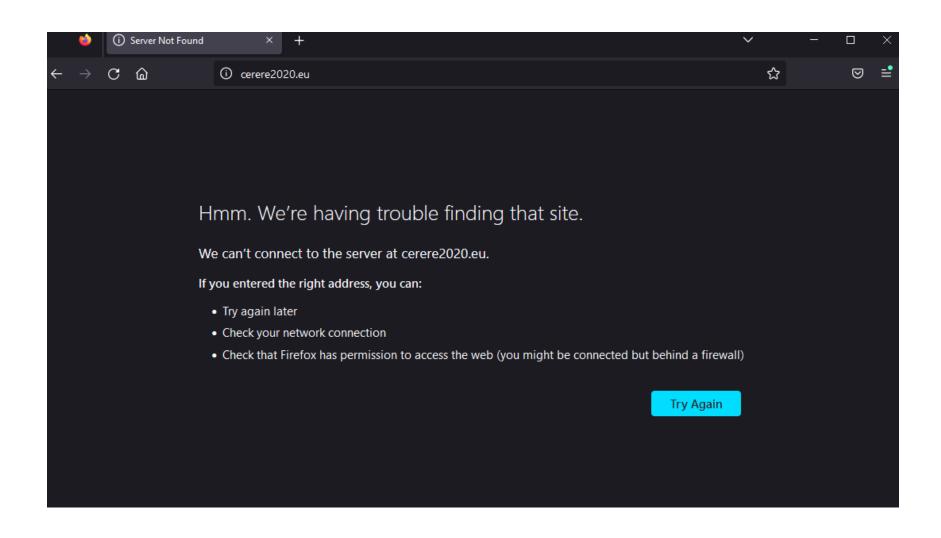
rcn	teaser	summary	workPerformed	projectAcronym	url
302075	The SWAMI project shall aim	The SWAMI project	Inital extension of UM from a upper boundary of 85 km to	SWAMI	http://swami-h2020.eu/
302061	Terrestrial demands on space	Terrestrial demands	During the 1st reporting period, LEA achievements are:	LEA	http://www.welea.eu
420383	The CLINGLIO project addre	The CLINGLIO proje	The project is divided into seven work packages (WP 1 to	CLINGLIO	https://clinglio.eu/
475812	CO2 concentrations are stea	CO2 concentrations	The major achievements of this project are: (1) experimen	WU TANG	http://www.smithelectrolab.com
473137	This project aims to perform	This project aims to	The following deliverables planned for period 1 have been	IMPACT-2	https://projects.shift2rail.org/s2r_ipcc_n.aspx
407722	Secure and affordable energ	Secure and affordat	During the first 30 months of the project, Spine has design	Spine	http://www.spine-model.org/
475914	Problem/Issues being address	Problem/Issues beir	Over the course of the first part of the project, the following	PHiVe	http://phive-project.com
478111	Fuel cells can play a major ro	Fuel cells can play a	The RoRePower project aims at strengthening the Europe	RoRePower	http://ower.com
455222	Industry is the largest heat-c	Industry is the large:	"During this first phase of the project activities within Tech	Highlift	http://highlift.olvondotech.no
646301	Guidelines such as "check th				
646038	Cardiovascular (CV) disease	Cardiovascular (CV)	During the first year, the activity mainly focused on three in	LION-HEARTED	https://lion-hearted.eu/
453396	Forest Flux will change the p	Forest Flux will char	The first period of the project started in December 2018 at	ForestFlux	http://forestflux.eu

tinyurl.com/cordis-h2020reports

cerere2020.eu



cerere2020.eu



euh2020aniage.org



ANIAGE PROJECT

High Dimensional Heterogeneous Data based Animation Techniques for Southeast Asian Intangible Cultural Heritage Digital Content-AniAge

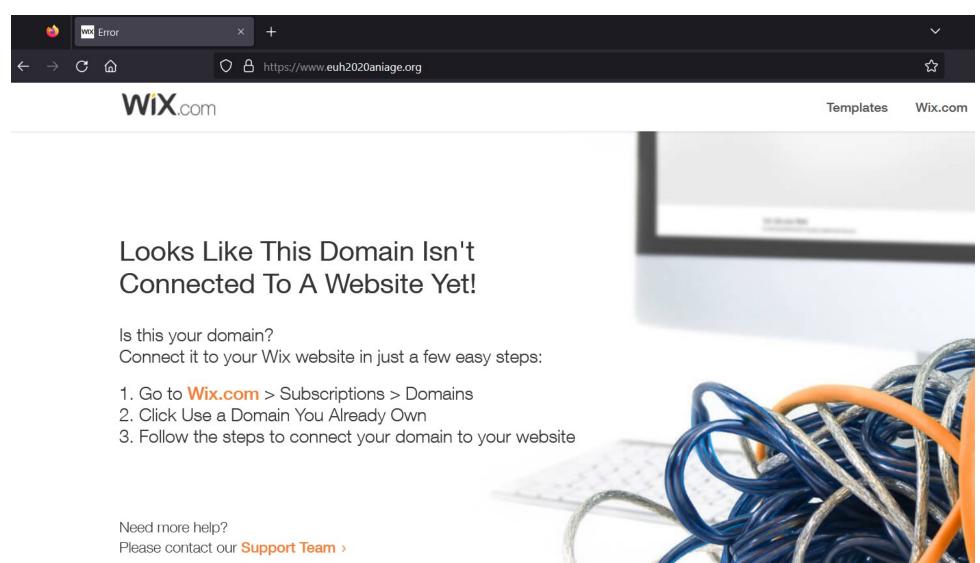
To develop novel techniques and tools to reduce the production costs and improve the level of automation without sacrificing the control from the artists, in order to preserve the performing art related ICHs of Southeast Asia.

HISTORY

Despite the advantages of 3D computer animation technology, due to the diversity of ICHs, digital preservation still faces enormous technical challenges

Read more

euh2020aniage.org

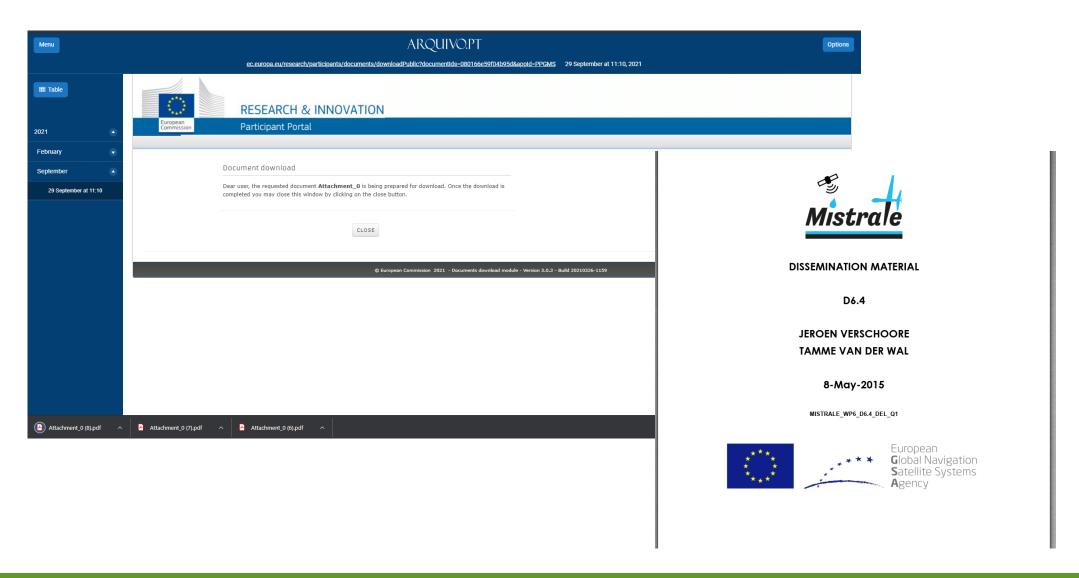


cordis-h2020projectDeliverables.xlsx

projectIE	projectAcronym	topics	description	deliverableType	url
764042	CHESTER	LCE-07-2016-2017	Detailed design of	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c7d5a799&appId=PPGMS
696129	GREENSOUL	EE-11-2015	This report will rep	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5abd1e095&appId=PPGMS
691246	QUANTUM DYNA	MSCA-RISE-2015	D6 Proceedings: E	Other	https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5cc70da01&appId=PPGMS
696129	GREENSOUL	EE-11-2015	This report will dea	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5abc145d8&appId=PPGMS
696129	GREENSOUL	EE-11-2015	This report will dea	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b3cdfe70&appId=PPGMS
690925	Phoenix	MSCA-RISE-2015	D7.8: Report on th	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c9cd177c&appId=PPGMS
690898	FORMILK	MSCA-RISE-2015	Reports on second	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5cb0161cd&appId=PPGMS
690898	FORMILK	MSCA-RISE-2015	Report on the real	i Open Research Da	https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5cab491b9&appId=PPGMS
689812	LANDSENSE	SC5-17-2015	Report on engage	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c2dc9f2c&appId=PPGMS
777365	eTRANSAFE	IMI2-2016-09-04	A roadmap of syne	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c9fdfe45&appId=PPGMS
693781	EULAC Focus	INT-12-2015	The paper will coll	Documents, reports	s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c11e7319&appId=PPGMS
801413	H-Reality				s https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c2f01005&appId=PPGM9

tinyurl.com/cordis-h2020projectDeliv

<u>ec.europa.eu/research/participants/documents/downloadPublic?</u> <u>documentIds=080166e59f04b95d&appId=PPGMS</u>



cordis-h2020projects.xlsx

47% of the projectUrls were empty

acronym	title	ojectUrl
MARmaED	MARine MAnagement and Ecosystem Dynamics under cl http	os://www.marmaed.uio.no/
ENIW	Services to enhance innovation management capacities o http	o://www.enterprise-europe.co.uk/
ARCADE	Aligning Research & Innovation for Connected and Autom	
QMiCS	Quantum Microwave Communcation and Sensing http	os://qmics.wmi.badw.de
RenalToolBox	Developing novel tools and technologies to assess the sa	
REDIFUEL	Robust and Efficient processes and technologies for Drop http	o://www.redifuel.eu
ROSSINI	RObot enhanced SenSing, INtelligence and actuation to I http	o://www.rossini-project.com
TRA VISIONS 2020	TRA VISIONS 2020 http	os://www.travisions.eu
NONTOX	Removing hazardous substances to increase recycling rate	
ProtoMet	Protometabolic pathways: exploring the chemical roots of	
PEQEM	Photonics for engineered quantum enhanced measureme http	o://www.bristol.ac.uk/physics/people/jonathan-c-matthews/
I-MOVE-plus	I-MOVE+ Integrated Monitoring of Vaccines Effects in Eur http	os://sites.google.com/a/epiconcept.fr/imoveplus/

Automated identification of URLs to complement EU cordis data set

Automatic search for meta-data on Bing API to find additional URLS

acronym	title	projectUrl
MARmaED	MARine MAnagement and Ecosystem Dynamics under cl	https://www.marmaed.uio.no/
ENIW	Services to enhance innovation management capacities of	http://www.enterprise-europe.co.uk/
ARCADE	Aligning Research & Innovation for Connected and Autom	
QMiCS	Quantum Microwave Communcation and Sensing	https://qmics.wmi.badw.de
RenalToolBox	Developing novel tools and technologies to assess the sa	
REDIFUEL	Robust and Efficient processes and technologies for Drop	http://www.redifuel.eu
ROSSINI	RObot enhanced SenSing, INtelligence and actuation to I	http://www.rossini-project.com
TRA VISIONS 2020	TRA VISIONS 2020	https://www.travisions.eu
NONTOX	Removing hazardous substances to increase recycling ra	
ProtoMet	Protometabolic pathways: exploring the chemical roots of	
PEQEM	Photonics for engineered quantum enhanced measureme	http://www.bristol.ac.uk/physics/people/jonathan-c-matthews/
I-MOVE-plus	I-MOVE+ Integrated Monitoring of Vaccines Effects in Eur	https://sites.google.com/a/epiconcept.fr/imoveplus/

+Acronym +Title +CommonTerm -Cordis -EC

tinyurl.com/preserving-websites-research

acronym	title projectUrl
MARmaED	MARine MAnagement and Ecosystem Dynamics under cl https://www.marmaed.uio.no/
ENIW	Services to enhance innovation management capacities o http://www.enterprise-europe.co.uk/
ARCADE	Aligning Research & Innovation for Connected and Autom
QMiCS	Quantum Microwave Communcation and Sensing https://qmics.wmi.badw.de
RenalToolBox	Developing novel tools and technologies to assess the sa
REDIFUEL	Robust and Efficient processes and technologies for Drop http://www.redifuel.eu
ROSSINI	RObot enhanced SenSing, INtelligence and actuation to I http://www.rossini-project.com
TRA VISIONS 2020	TRA VISIONS 2020 https://www.travisions.eu
NONTOX	Removing hazardous substances to increase recycling rate
ProtoMet	Protometabolic pathways: exploring the chemical roots of
PEQEM	Photonics for engineered quantum enhanced measureme http://www.bristol.ac.uk/physics/people/jonathan-c-matthews/
I-MOVE-plus	I-MOVE+ Integrated Monitoring of Vaccines Effects in Eur https://sites.google.com/a/epiconcept.fr/imoveplus/

ProtoMet

Protometabolic pathways: exploring the chemical roots of systems biology

project

-site:cordis.europa.eu

-site:ec.europa.eu

+Acronym

+Title

+CommonTerm

-Cordis

-EC

Microsoft Bing ProtoMet Protometabolic pathways: exploring the chemical re

IMAGES







34 Results

Protometabolic pathways: exploring the chemical roots of ...

VIDEOS

https://protomet-etn.eu ▼

ProtoMet attempts to understand how prebiotic chemistry gave rise to dissipative **chemical systems** that supported early cellular life. Previously deciphered prebiotic versions of triose glycolysis and the citric acid cycle will be integrated, thus reconstructing large fragments of a **chemical** network, i.e. metabolism, capable of supporting protocellular life.

NEWS

ProtoMeT Protometabolic pathways: exploring the chemical ...

https://www.ikertzaileengaua-ehu.org/rincon-europeo/protomet... •

ProtoMeT Protometabolic pathways: exploring the chemical roots of systems biology Proyecto presentado por Nino Lauber Understanding how prebiotic chemistry gave rise to life as we know it represents one of the greatest enduring mysteries.

EU H2020 Project "PROTOMET (Protometabolic pathways ...

https://www.fabiodisconzi.com/open-h2020/projects/218166/index.html ▼ Information about the Horizon 2020 project "Protometabolic pathways: exploring the chemical roots of systems biology (2018)" - open-H2020 observatory.

We complement the EU data sets with archived versions and the additional URLs identified

acronym	title	projectUrl	URLsBing Search	ArchivedProjectURLs
MARmaED	MARine MAnagement and Ecosystem Dynamics under cl	https://www.marmaed.uio.no/		https://arquivo.pt/wayback/https://www.marmaed.uio.no/
ENIW	Services to enhance innovation management capacities of	http://www.enterprise-europe.co.uk/		https://arquivo.pt/wayback/http://www.enterprise-europe.co.uk/
ARCADE	Aligning Research & Innovation for Connected and Autom		https://www.rijkswaterstaat	https://arquivo.pt/wayback/https://www.rijkswaterstaat.nl/en/mobility/projects/ar
QMiCS	Quantum Microwave Communcation and Sensing	https://qmics.wmi.badw.de		https://arquivo.pt/wayback/https://qmics.wmi.badw.de
RenalToolBox	Developing novel tools and technologies to assess the sa		https://www.findaphd.com/	https://arquivo.pt/wayback/https://www.findaphd.com/phds/project/assess-the-s
REDIFUEL	Robust and Efficient processes and technologies for Drop	http://www.redifuel.eu		https://arquivo.pt/wayback/http://www.redifuel.eu
ROSSINI	RObot enhanced SenSing, INtelligence and actuation to I	http://www.rossini-project.com		https://arquivo.pt/wayback/http://www.rossini-project.com
TRA VISIONS 2020	TRA VISIONS 2020	https://www.travisions.eu		https://arquivo.pt/wayback/https://www.travisions.eu
NONTOX	Removing hazardous substances to increase recycling ra		https://www.energy.imdea.o	https://arquivo.pt/wayback/https://www.energy.imdea.org/research/projects/non
ProtoMet	Protometabolic pathways: exploring the chemical roots of		https://protomet-etn.eu/, htt	https://arquivo.pt/wayback/https://protomet-etn.eu/, https://arquivo.pt/wayback/l
PEQEM	Photonics for engineered quantum enhanced measureme	http://www.bristol.ac.uk/physics/people/jonathan-c-	na	https://arquivo.pt/wayback/http://www.bristol.ac.uk/physics/people/jonathan-c-n
I-MOVE-plus	I-MOVE+ Integrated Monitoring of Vaccines Effects in Eur	https://sites.google.com/a/epiconcept.fr/imoveplus/		https://arquivo.pt/wayback/https://sites.google.com/a/epiconcept.fr/imoveplus/

tinyurl.com/cordis-h2020projects

protomet-etn.eu, Arquivo.pt

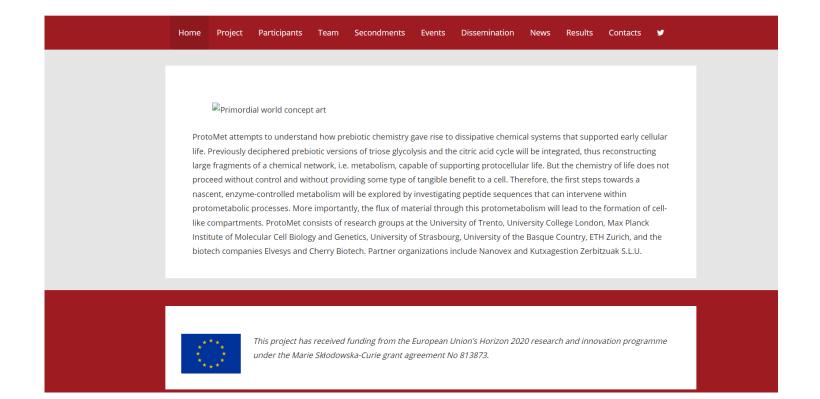
ARQUIVO.PT

protomet-etn.eu 1 October at 18h00, 2020



Protometabolic pathways: exploring the chemical roots of systems biology

PROTOMET MSCA ITN ETN 2018 PROJECT (H2020 EU PROJECT MSCA GA NO 813873)



Collections metrics



Collection

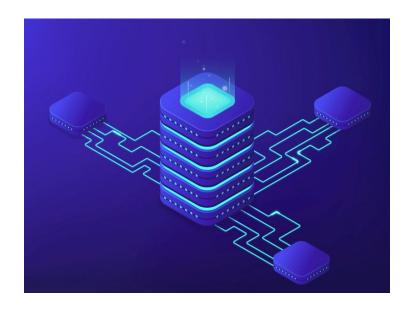
• Date: 1 full year (end of 2020 - end of 2021)

Number of crawls: 2

• URLs: 118 326

• Disk space: 17,3 TB

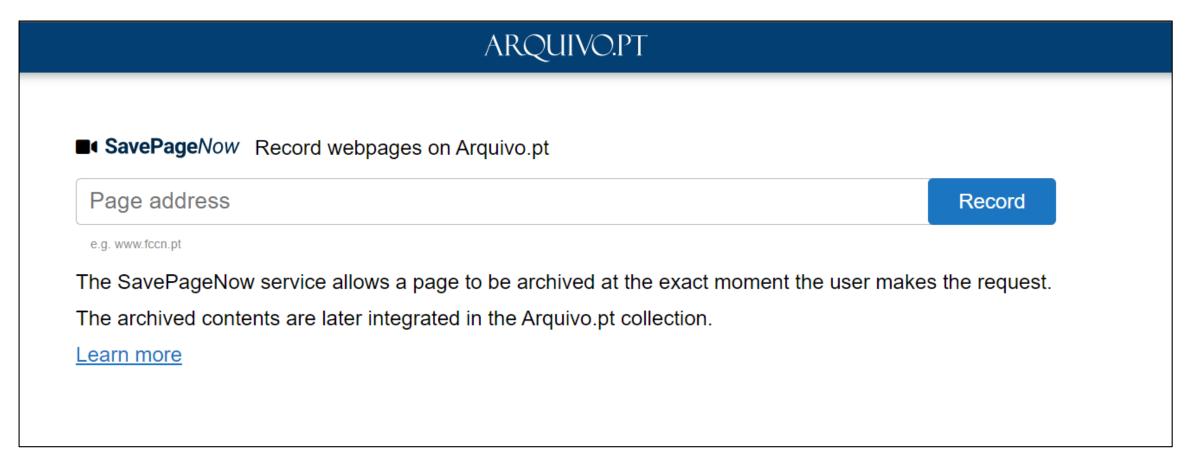
• Files: 197 millions collected files



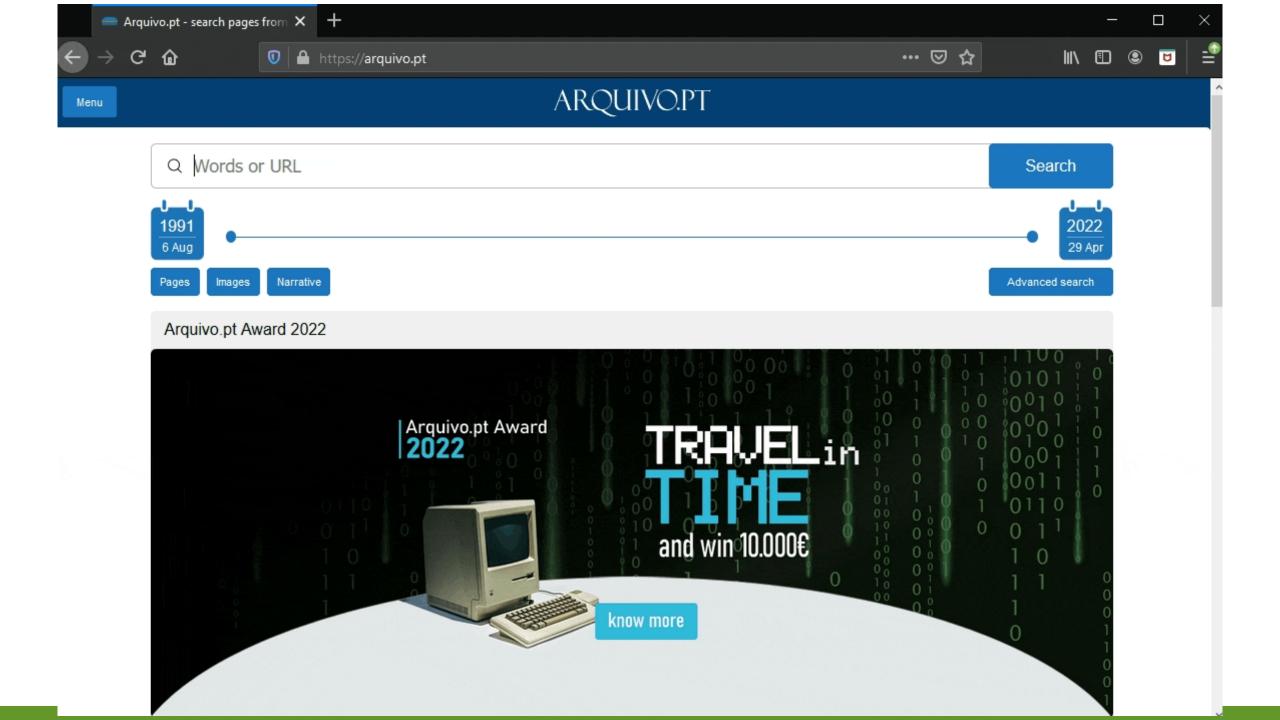
Collection

- Languages: All EU languages
- Scope: 5 depth (similar to domain crawl)
- Resources (People):
 - > 10 days digital curator quality control
 - > 10 days engineer run and quality control
- Resources (Machine):
 - > 1 computer (Run)
 - ➤ 1 computer (Store)

SavePageNow: Archive a webpage immediately (dpa German Press Agency is using it)



arquivo.pt/savepagenow



Submit our Web page

ARQUIVO.PT

■ SavePageNow Record webpages on Arquivo.pt

commission.europa.eu

Record

e.g. www.fccn.pt

The SavePageNow service allows a page to be archived at the exact moment the user makes the request.

The archived contents are later integrated in the Arquivo.pt collection.

Learn more

arquivo.pt/services/savepagenow?l=en

■ SavePage/				QUIVO.PT ission.europa.eu		Saving	Conclude
An official web		n How do you know? ✓ opean nmission	E	nglish		Search	
		J: affordable e energy for l	Europe			© damianobuff	o/Adobe Stoo
			• 0	0 0		1 of 4	
A	This site uses cookie	s to offer you a better brow	sing experience. Find ou	t more on <u>how we</u>	Accept all cookies	Accept only essential cookies	

use cookies.

Memorial

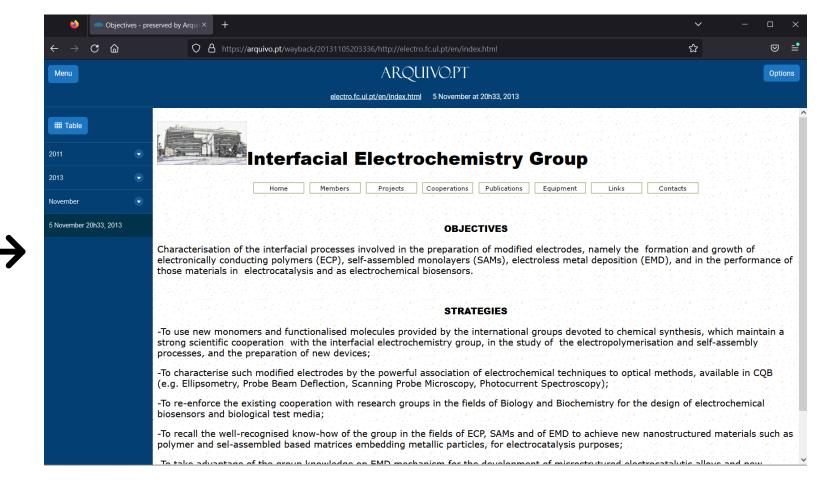
ARQUIVO.PT The site has been disabled.

ne one nao been aloubica.

Arquivo.pt Memorial preserved its content.

Browee in Arquivo of

electro.fc.ul.pt



190 Website Active in Memorial

1	Domain of the Original Website - Auto generated by W column (fill i		URL of destiny version (do not edit this column)	Date of activation i
175	giardiarisk.fc.ul.pt	20120126142552	https://arquivo.pt/wayback/2012012614255	2023-05-24
176	habcol.fc.ul.pt	20100610035458	https://arquivo.pt/wayback/2010061003545	2023-05-24
177	ica.fc.ul.pt	20131105230359	https://arquivo.pt/wayback/2013110523035	2023-05-24
178	if.fc.ul.pt	20030325143605	https://arquivo.pt/wayback/2003032514360	2023-05-24
179	insectamadeira.fc.ul.pt	20150416142932	https://arquivo.pt/wayback/2015041614293	2023-05-24
180	learnc3s.fc.ul.pt	20230316213813	https://arquivo.pt/wayback/2023031621381	2023-05-24
181	lisboncisa.fc.ul.pt	20230320094332	https://arquivo.pt/wayback/2023032009433	2023-05-24
182	liscos.fc.ul.pt	20030720100054	https://arquivo.pt/wayback/2003072010005	2023-05-24
183	matematicaabrincar.fc.ul.pt	20091218222739	https://arquivo.pt/wayback/2009121822273	2023-05-24
184	megahazards.fc.ul.pt	20160208095626	https://arquivo.pt/wayback/2016020809562	2023-05-24
185	nurseries.fc.ul.pt	20230320103744	https://arquivo.pt/wayback/2023032010374	2023-05-24
186	obrascompletas-dm.fc.ul.pt	20120123224303	https://arquivo.pt/wayback/2012012322430	2023-05-24
187	oincendio.fc.ul.pt	20230320104540	https://arquivo.pt/wayback/2023032010454	2023-05-24
188	parquedeestufas.fc.ul.pt	20131115044517	https://arquivo.pt/wayback/2013111504451	2023-05-24
189	perarocha.fc.ul.pt	20230319132800	https://arquivo.pt/wayback/2023031913280	2023-05-24
190	medidasjustica.portugal.gov.pt	20190424170717	https://arquivo.pt/wayback/2019042417071	2023-05-24