



Setting up an EU web preservation service for the long-term

Tales of a (sometimes) bumpy road

Els Breedstraet

Team leader archives and long-term preservation

Publications Office of the European Union

A word about the Publications Office

- Inter-institutional service provider, evolving from traditional publisher to provider of information management services.
- We publish EU law and other information from EU institutions, and we make it available for easy long-term access and reuse.
- Three pillars of activities: production, access and reuse, long-term preservation.
- We work for around 150 'author services' from EU institutions, agencies and bodies.
- We maintain several websites:



Why?

Facts:

- More and more EU information is only available on the web.
- Web content has often a short life-span.

➔ **Risk: loss of information!**

Solution: web preservation!

➔ Contents stay available for future generations, even if it's no longer on the live website (or if the live website doesn't exist anymore).



What (not)?

Aim

Preserving EU online heritage

Scope

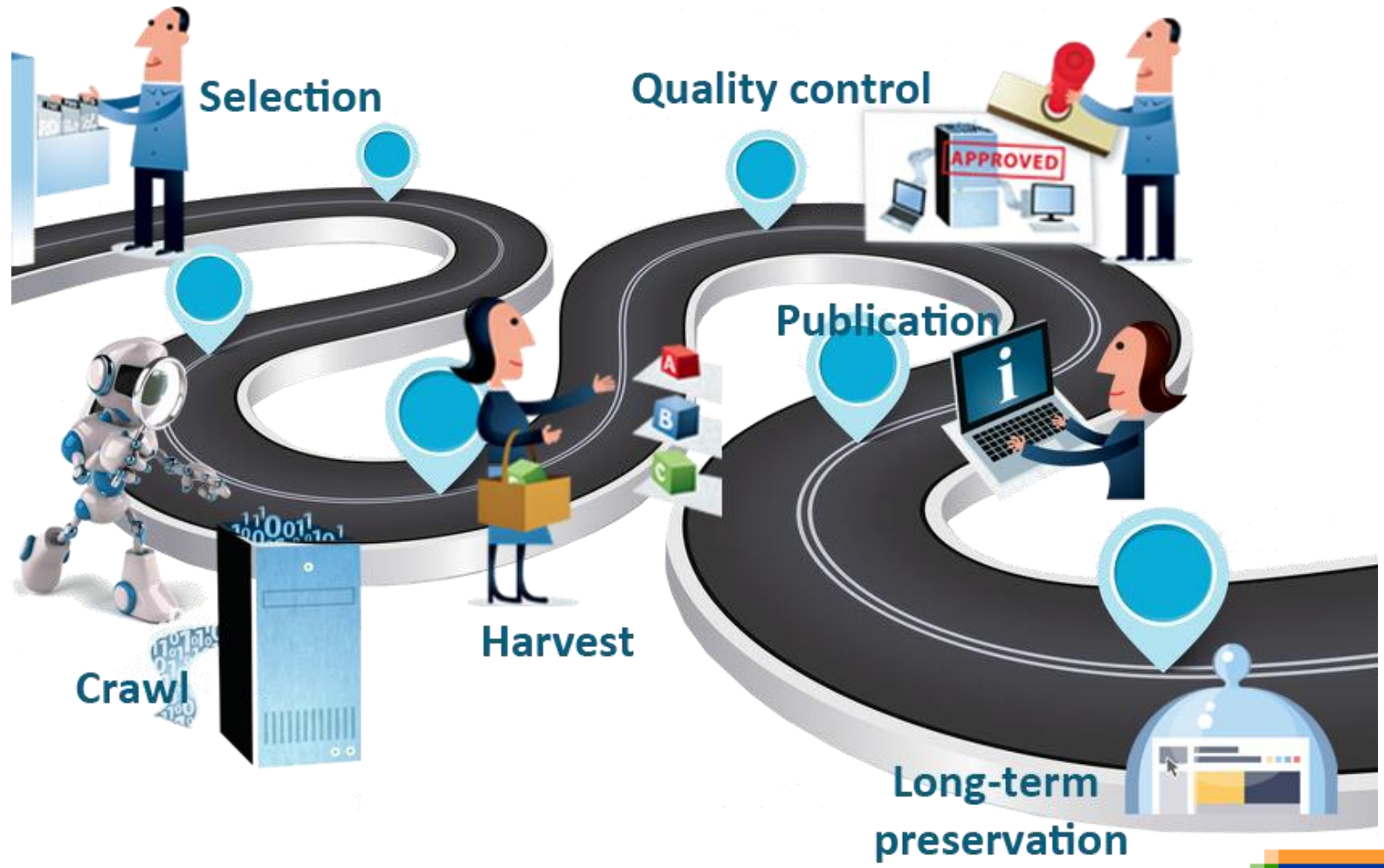
- Quarterly crawling of europa.eu domain and subdomains
- Ad hoc crawling of sites that will considerably change or that will be taken offline

Out of scope

- External links (redirection page to live)
- Social media
- Dynamic content / databases



Who and how?



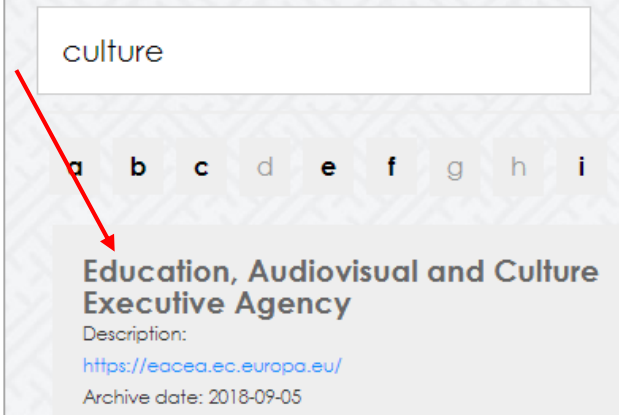
How to access it?

Access via data.europa.eu/webarchive



EU Web Archive overview
Publications Office

This web archive contains the websites of, mainly, EU institutions and agencies. Most of these sites are ho




culture

a b c d e f g h i


Education, Audiovisual and Culture Executive Agency
Description:
<https://eacea.ec.europa.eu/>
Archive date: 2018-09-05



EU web archive 

September 2018

Su	Mo	Tu	We	Th	Fr	Sa
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	1	2	3	4	5	6

Erasmus|  advanced search



EACEA
Education,
European Commission

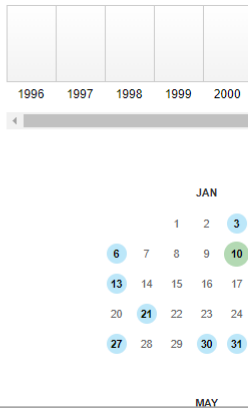
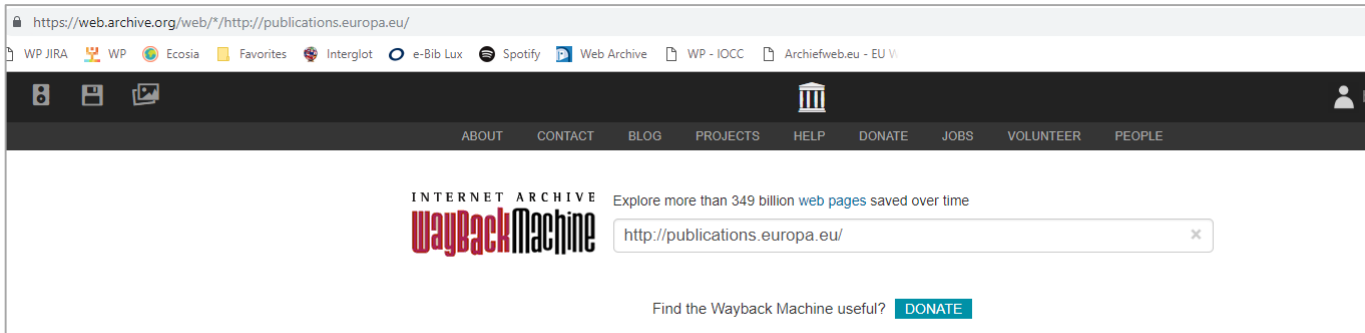
European Commission > EACEA > Homepage

PROGRAMMES MANAGED



How else to access it?

Access via <https://archive.org/web/>

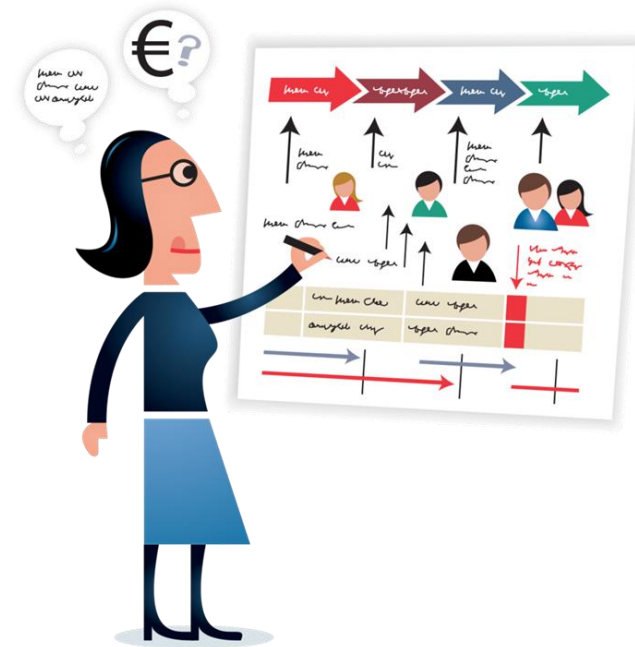


Staff	635 officials
Administrative budget	81.07 million euros
— Consolidation	1.99 million euros



Challenges

- Improving the user interface / usability of the archive
- Outsourcing vs doing it ourselves?
- Quality control: how good is good enough?
- Scope: setting the boundaries
- Distributed harvesting and central access?
Collaborations?
- ...



More information



- Find out more about OP:
<https://publications.europa.eu>
- Main contacts:
 - op-web-preservation@publications.europa.eu
 - els.breedstraet@publications.europa.eu
(head of sector and project manager)



Thanks for your attention!

Questions?

