



Fedora as EU Law repository

Fulgencio.Sanmartin@publications.europa.eu

Helsinki, OR2014, 13.06.2014

Ex: the right to be forgotten by Google

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style | Travel | Environment | T

News > Technology > Google

EU court backs 'right to be forgotten': Google must amend results on request

Individuals have right to control their data and can ask search engines to remove results, says European court

Alan Travis and Charles Arthur
theguardian.com, Tuesday 13 May 2014 14.06 BST

 Jump to comments (689)



Technology
Google · Search engines
· Internet · Data protection

World news
Privacy · European Union
· Europe

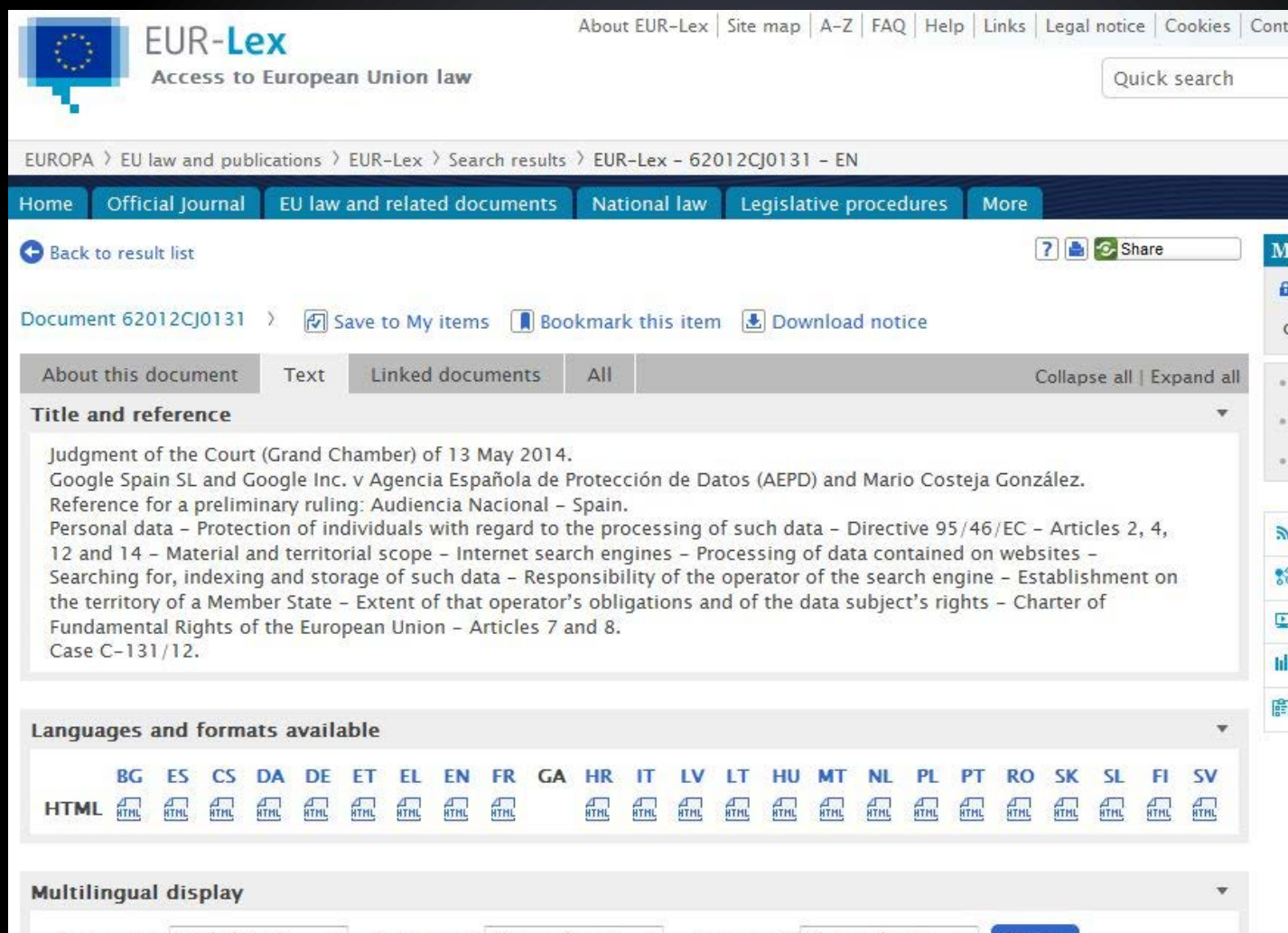
Share 2361
Tweet 1,217
g+1 495
Pin it 10
in Share 528
Email

Article history

<<http://goo.gl/Qdjp1X>>



Ex: EUR-Lex - Judgement of the Court C-131/12



The screenshot displays the EUR-Lex website interface. At the top left is the EUR-Lex logo with the text "Access to European Union law". To the right are navigation links: "About EUR-Lex", "Site map", "A-Z", "FAQ", "Help", "Links", "Legal notice", "Cookies", and "Contact". A "Quick search" box is located on the right. Below the header is a breadcrumb trail: "EUROPA > EU law and publications > EUR-Lex > Search results > EUR-Lex - 62012CJ0131 - EN". A navigation bar contains buttons for "Home", "Official Journal", "EU law and related documents", "National law", "Legislative procedures", and "More". Below this is a "Back to result list" link and a "Share" button. The main content area shows the document identifier "Document 62012CJ0131" and options to "Save to My items", "Bookmark this item", and "Download notice". A tabbed interface is visible with "About this document", "Text", "Linked documents", and "All" tabs. The "Text" tab is active, displaying the "Title and reference" section. The text in this section reads: "Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González. Reference for a preliminary ruling: Audiencia Nacional - Spain. Personal data - Protection of individuals with regard to the processing of such data - Directive 95/46/EC - Articles 2, 4, 12 and 14 - Material and territorial scope - Internet search engines - Processing of data contained on websites - Searching for, indexing and storage of such data - Responsibility of the operator of the search engine - Establishment on the territory of a Member State - Extent of that operator's obligations and of the data subject's rights - Charter of Fundamental Rights of the European Union - Articles 7 and 8. Case C-131/12." Below this is a "Languages and formats available" section with a grid of language codes (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV) and HTML icons. At the bottom is a "Multilingual display" section.

<<http://goo.gl/MwUkfq>>



Fedora data model in Cellar

work: DC, MD

```
cellar:09eb0861-da7a-11e3-8cd4-01aa75ed71a1  
http://publications.europa.eu/resource/celex/62012CJ0131
```

expression (24 languages): + RELS-EXT

```
cellar:09eb0861-da7a-11e3-8cd4-01aa75ed71a1.0006  
http://publications.europa.eu/resource/celex/62012CJ0131.ENG
```

manifestation (3-4 file types): +RELS-INT, DOC_n, DOC-EXT

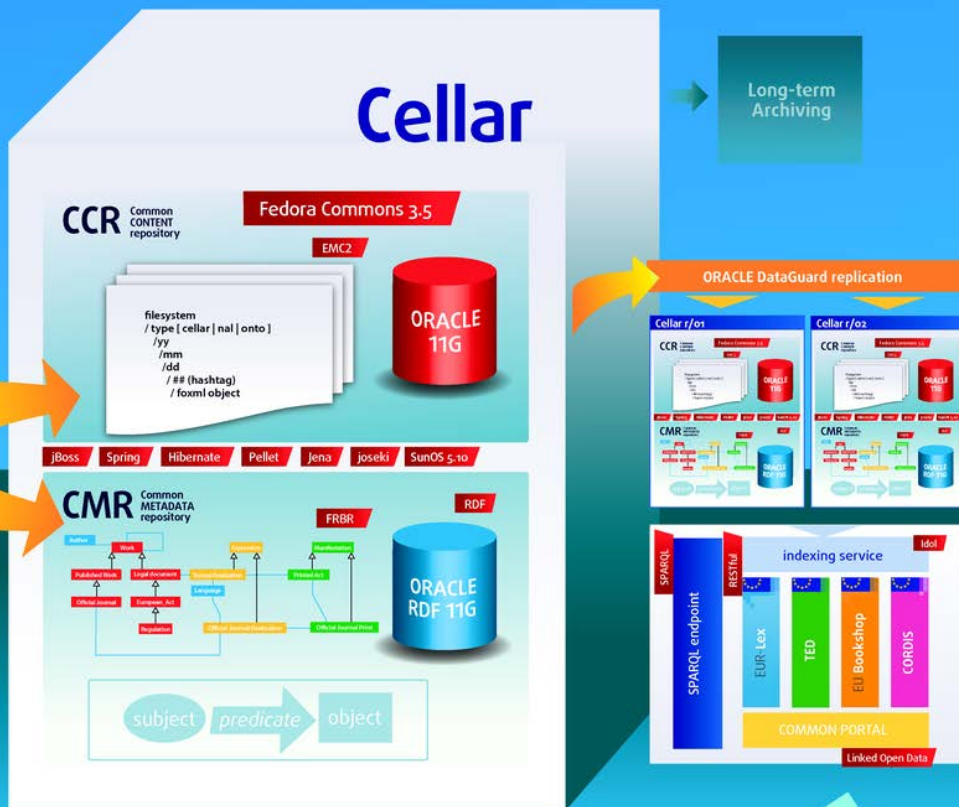
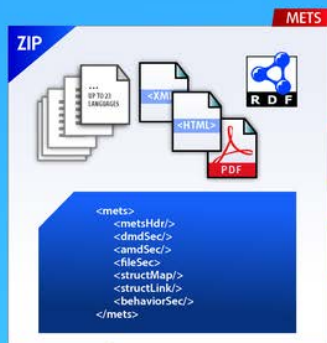
```
cellar:09eb0861-da7a-11e3-8cd4-01aa75ed71a1.0006.01  
http://publications.europa.eu/resource/celex/62012CJ0131.ENG.xhtml
```



Architecture

The metadata and content from daily production + backlog of the Publications Office represents over 43 million files,

43 million files,
over 13 TB and growing...



internet





Architecture

database Oracle

Nodes: 2

T4/Solaris 11, 128 cores, 256GB

Oracle 11g EE + RDF Store

Fedora v3.6.1

Nodes: 6 on NFS

RHEL 8 cores, 128GB, load balanced

Cache: Apache Traffic Server, 1TB



Volume of data

Fedora

Identifiers: 45million (+750K/month)

Files: 67million (+1.6mill/month)

Disk: 9.3TB (+1.2TB/month)

Requests: ~3.5million/day (peaks may double that)

Database

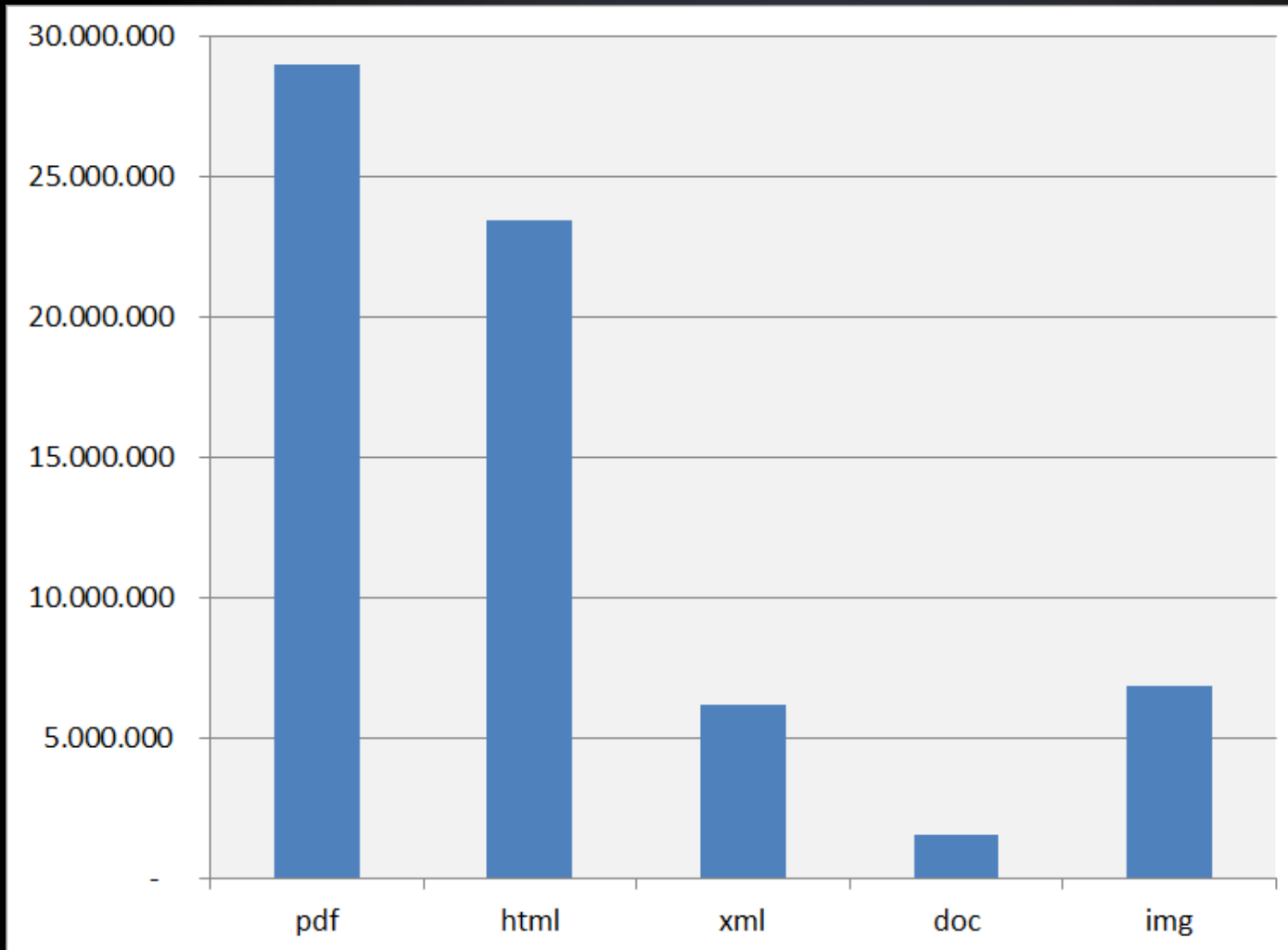
Business identifiers: 103million (+1.6mill/month)

Disk: 3.15TB (+60GB/month)

Triples: 1.08billion (+10mill/month)



Volume of data



Consistency

Problem: FO vs RDF Store triples

Transactions simulated in Cellar, eventually consistent

Solution and lessons learned

Keep concurrency in mind (fs and db)

Time to write in fs/db can be *very* different

Tool to compare FO vs triples

Future

Transactional Fedora



Storing policies

Problem: daily vs archive objects

Simultaneous load

Legal constraints, mandate

Solution and lessons learned

Separate ingestion threads and priorities

Store /yy/mm/dd versus /uu/id/31

Future

Fedora statistics for popular FO

Policies to move less popular FO to lower QoS stores



Performance

Problems with FO

+500 updates >> try disabling versioning, AUDIT DS

Blocking, rare errors >> check lib versions and compatib

Lessons learned

Performance is half software (YourKit, your friend), half hw

Be aware of redundancies: triples, resource index, audit





Future

Cellar

RESTful versioning

Public SPARQL endpoint

We would like

Transactions + storing policies and stats: Fedora

Export tool for FO 3.x

