HHOF – Streamlit Web App (Explanation-Only)

Goal: operator pastes **JORF email** + a **EUR-Lex daily URL** (auto-crawl subpages) → parse, classify (prompt-only), export **CSV**.

Operator Flow

1 Open the Streamlit app. Sidebar: date, "Include subpages (~10)", keywords, language (FR/EN), API toggle.

2 **Paste JORF email** into the text area → click *Parse JORF* → table shows items (ministry, title, link).

3 **Paste EUR-Lex Daily View URL** (L or C series) → check "Include subpages" → click *Parse EUR-Lex* → items appear (type, date, title, link).

4 Click *Classify* → zero-shot labels for **Theme** and **Applicability** (Info / Regulation / Directive / Decree).

5 Optionally set **Filters** (keywords, ministries, types). Review merged table.

6 Click *Export CSV* → Google-Sheets-compatible CSV downloads.

No training. Public sources only. Operator-triggered (no scheduler).

CSV Schema (Output)

Column	Description
Date	Publication date (ISO, e.g., 2025-09-17)
Source	"EUR-Lex L", "EUR-Lex C", "JORF", "Red On Line"
Title	Official title / email line
Summary	Short subject/abstract (if available)
Link	Canonical URL
Theme	Zero-shot topic (Education, Health, Interior/Security, …)
Applicability	Info vs Directive/Regulation/Decree/Circular

Column	Description
Keywords (opt.)	Matched operator filters

UTF-8 CSV; opens directly in Sheets/Excel.

Sizing – Roles (Integer FTE) & Aggressive Days

Component	Description	FTEs	Days	Skills & Roles	Azure	Notes
Streamlit App (UI + API)	Two inputs, previews, CSV export	1	3	Full-Stack Engineer, Streamlit, UX	App Service / Container Apps	Single-page app
EUR-Lex Parser	Daily view + ~10 subpages; titles/links/dates	1	3	Data Engineer, httpx, BeautifulSoup	Runs in app	Robust to HTML drift
JORF Parser	Paste email → extract sections/items/URLs	1	2	Data Engineer, Regex, FR text	Runs in app	Handles rectificatifs
Standardize & Merge	Unify schema; combine sources	1	1	Data Engineer, pandas	In-memory	DF → CSV
Zero-Shot Classification	Theme & Applicability via prompts	1	2	ML Engineer, Azure OpenAI, Prompting	Azure OpenAI + Key Vault	No training
Error Handling & Perf	Timeouts, retries, rate limiting	1	1	Software Engineer, Logging	App Insights	Graceful failures
Infra & CI/CD	Provision Azure, secrets, pipeline	1	2	DevOps Engineer, Key Vault, IaC	App Service, KV, OpenAI	Staging → Prod
QA & Fixtures	Golden samples; unit tests	1	1	QA Engineer, pytest	App Insights	Smoke + regression

Component	Description	FTEs	Days	Skills & Roles	Azure	Notes
Total		—	**15**	Python NLP Azure DevOps	As listed	Aggressive (≈3–4 weeks elapsed)

Azure Minimal Stack

- **App Service (Linux)** or **Container Apps** to host Streamlit.
- **Azure OpenAI** for zero-shot labels (Theme, Applicability).
- **Key Vault** for API keys; **App Insights** for logging/metrics.
- **Blob Storage** (optional) to archive daily CSVs.

Manual runs via the UI; no background scheduler.

Contact: stephanie.garrigues.external@airbus.com • Call 17/09/2025 • Inputs: EUR-Lex (L/C daily) & JORF daily email.