

Dibab

Manual del usuario

Clasificar los papeles administrativos de un hogar, en sus propias máquinas

Versión 0.9.0-alpha

2026-11-26

Traducción. Prevalece la versión en inglés.

`dibab.davalan.fr`

Índice general

Introducción **3**

Qué hace Dibab . . . 3
Qué no hace . . . 3
Una herramienta de hogar . . . 3
Cómo se organiza este manual . . . 4
Convenciones . . . 4

Instalación **5**

Lo que necesita . . . 5
Los modelos, por rol . . . 5
El árbol final . . . 5
Instalación desde el paquete Debian . . . 6
Instalación en Fedora, openSUSE Leap y Arch . . . 7
Instalar desde las fuentes . . . 7
Ejecutar las pruebas (opcional) . . . 8
Dónde se busca la configuración . . . 8

Configuración **9**

Las claves que debe escribir . . . 9
`destinations`: la bóveda de cada ámbito . . . 9
`destination`, `arborescence`, `executer_sur` . . . 10
`suivi`: estado, diario e informes . . . 10
`ollama`: los modelos . . . 11
`extraction` y `classification` . . . 12
`categories`, `personnes`, `emetteurs_connus`, `pieces_communes` . . . 13
`nommage`: el nombre de un documento clasificado . . . 13
`extensions_cibles` y `scan` . . . 14
`sources_locales` . . . 14
`ssh`, `sources_distantes`, `imap`, `oauth2`, `sources_imap` . . . 15
`mesure` e `inventaire` . . . 15

Primera pasada **16**

Simular: `--dry-run` . . . 16
Comprobar el encaminamiento sin los modelos . . . 18
Clasificar: `--execute` . . . 19
Deshacer: `--rollback` . . . 21
Códigos de salida . . . 21

Bóveda y cifrado **22**

Qué es una bóveda . . . 22
La comprobación del montaje . . . 22
Un ejemplo con gocryptfs . . . 23
Dónde vive el estado . . . 23
Espacio en el disco . . . 24

Fuentes por ssh **25**
Lo que debe ofrecer la máquina remota . . . 25
Declarar la máquina . . . 25
Una primera pasada: listar y luego recuperar . . . 26
Volver a ejecutar . . . 27
Cuando falla la conexión . . . 28

Buzones IMAP y cuentas de Google **29**
Lo que se pide al servidor . . . 29
Declarar un buzón . . . 29
La contraseña . . . 30
Una primera pasada: listar y luego depositar . . . 30
Cuando falla la conexión . . . 32
Cuentas de Google con OAuth2 . . . 33

Inventario y rotación **35**
Leer lo que una bóveda ya contiene: `--inventaire` . . . 35
Mantener corto el diario: la rotación . . . 36

Diagnóstico **37**
Los estados . . . 37
Las comprobaciones, en orden . . . 38
Leer un fallo . . . 39
Una instalación recién hecha . . . 40
En otro idioma . . . 40

Soberanía y límites **41**
Lo que la herramienta garantiza . . . 41
Respeto del hogar . . . 41
Los límites . . . 41
Informar de un problema . . . 42

Introducción

Dibab («elegir» en bretón) recoge, analiza y clasifica los documentos administrativos personales de un hogar: facturas, cartas, contratos, avisos de organismos. Mueve cada documento a un árbol de directorios normalizado, bajo un nombre que dice la fecha, el remitente y la categoría, y lleva un diario de todo lo que hace, de modo que cualquier clasificación pueda deshacerse. Este manual describe la versión 0.9.0-alpha.

Funciona en modo soberano. Ningún dato sale de su red local: ni servicio de terceros, ni modelo remoto, ni nube. Los modelos de lenguaje son los de Ollama, que se ejecuta en su propia máquina.

Qué hace Dibab

- Escanea directorios locales, máquinas remotas por `ssh` y buzones de correo por IMAP. Un archivo remoto se copia y se verifica por huella, y el original queda intacto. Un buzón se lee sin modificarse: solo se recuperan los adjuntos, y ningún mensaje se marca como leído.
- Extrae el texto de cada archivo (la capa de texto de un PDF o, si no la hay, la lectura de las imágenes de las páginas por un modelo de visión), aparta lo que no es administrativo y clasifica el resto por categoría y por ámbito, personal o profesional, con el acuerdo de varios modelos.
- Renombra y mueve cada documento a la bóveda de su ámbito, escribe un diario con marca de tiempo y puede deshacer toda una sesión con `--rollback`.

Qué no hace

- No mueve un archivo del que no está seguro. La duda lleva a una carpeta llamada `POR_VERIFICAR/`, nunca a una clasificación arriesgada.
- Nunca borra un documento. Copia, verifica y solo entonces retira el original, cuando la copia queda establecida como idéntica.
- No modifica ni un buzón de correo ni una máquina remota.
- No se ejecuta en segundo plano. Usted lo lanza a mano, y su modo por defecto es la simulación, `--dry-run`.
- No monta, no cifra ni sincroniza ninguna bóveda. Eso corre por su cuenta, y el capítulo 5 explica cómo lo comprueba la herramienta.

Una herramienta de hogar

Dibab está pensado para la vida administrativa de un hogar, por el hogar. Una máquina, un directorio o una cuenta de correo solo entra en su ámbito de aplicación con el acuerdo, explícito o implícito, de la persona a la que concierne; en caso de duda, déjelo fuera. El programa no puede garantizar ese respeto: es el administrador de la instancia quien se compromete a ello.

El diario contiene nombres de archivos y remitentes, que ya revelan mucho (basta una carta de un organismo de salud). Por eso su lugar está dentro de la bóveda, no a su lado: ponga las claves `suivi.*` dentro de la bóveda; si no, se escribe en claro bajo `~/.local/state/dibab/`.

Cómo se organiza este manual

1. Introducción (este capítulo).
2. Instalación: qué instalar, dónde va todo y una primera comprobación.
3. Configuración: el archivo de configuración, clave por clave.
4. Primera pasada: simular, leer el informe, clasificar de verdad, deshacer.
5. Bóveda y cifrado: cómo la herramienta establece que una bóveda está montada, y dónde guardar el estado.
6. Fuentes por ssh: las máquinas remotas.
7. Buzones IMAP y cuentas de Google.
8. Inventario y rotación: leer lo que una bóveda ya contiene y mantener corto el diario.
9. Diagnóstico: `--diagnose`, estado por estado.
10. Soberanía y límites: lo que la herramienta garantiza y lo que no.

Convenciones

Los comandos se muestran tal como se teclean, en un terminal, para el usuario de la instalación descrita en el capítulo 2. El comando es `dibab`. El texto que sigue a un comando es su salida real, producida al repetir el comando en un área de prueba hecha de documentos ficticios; los modelos nombrados en esas salidas (`essai-ocr`, `essai-tri`, `essai-classification`) son los sustitutos de esa área de prueba, y los suyos son los que usted nombra en su configuración. Todos los documentos, personas, remitentes y empresas de este manual son inventados.

El idioma de los mensajes sigue a su sesión: Dibab lee la primera variable no vacía entre `LC_ALL`, `LC_MESSAGES` y `LANG`, y habla inglés, francés, alemán, español o chino simplificado según corresponda (cualquier otro valor da inglés). Los nombres de las opciones, los términos de estado como `OK` y `ÉCHEC`, los identificadores, los diarios y los archivos de plan no cambian nunca con el idioma, de modo que un diario escrito en un idioma lo reproduce `--rollback` en cualquier otro. Las salidas de este manual están en inglés.

Instalación

Lo que necesita

- Un sistema Linux con Python 3.10 o posterior. Dibab usa solo la biblioteca estándar de Python, así que no hay nada más que instalar para Python.
- `findmnt` (paquete `util-linux`), que establece que una bóveda está montada.
- `pdftotext` y `pdftoppm` (paquete `poppler-utils`), que leen los archivos PDF.
- `ssh` (paquete `openssh-client`), solo si escanea máquinas remotas, con una clave que abra una sesión sin contraseña.
- Ollama, instalado y en marcha, con al menos un modelo para cada rol (más abajo). Dibab nunca instala Ollama ni descarga un modelo.
- Un directorio de destino por ámbito, personal y profesional, de preferencia en un volumen cifrado (capítulo 5).
- `git`, para obtener las fuentes.

En Debian o Ubuntu, los paquetes se instalan con un solo comando:

```
sudo apt install python3 git util-linux poppler-utils openssh-client
```

Los modelos, por rol

Tres roles usan un modelo cada uno. El rol `ocr` lee las páginas escaneadas, `tri_rapide` («triaje rápido») aparta lo que no es administrativo y `classification` clasifica. Los modelos del rol `classification` votan: un documento solo se clasifica si su acuerdo alcanza `ollama.seuil_confiance` (capítulo 3). Dos roles opcionales, `tri_decision` y `objet`, confían la clasificación a un modelo `decision` en una sola pasada y el objeto a un modelo pequeño; sin ellos, clasifica `tri_rapide` (capítulo 3).

```
ollama pull glm-ocr
ollama pull mistral-nemo:12b-instruct-2407-q6_K
ollama pull gemma4
ollama pull gemma3:4b
ollama pull qwen2.5
ollama pull granite4.2:8b-Q5_K_M
```

Son los modelos de la instancia de referencia; sirven otros cualesquiera, siempre que los nombre en `ollama.roles`. Lo que está instalado se lista con `ollama list`.

El árbol final

~ es su directorio personal. Adapte las rutas a su instalación.

```
~/applications/dibab/              el repositorio clonado
├── dibab.py                       el punto de entrada
├── modules/                            un archivo por responsabilidad
│   └── langues/                        los diccionarios de mensajes: en, fr, de, es, zh
├── docs/                               la página de manual y los README traducidos
├── tests/                              las pruebas
```

```
└── config.exemple.json                la configuración de ejemplo
~/.local/bin/dibab                  un enlace a dibab.py: el comando
~/.config/dibab/config.json         su configuración, nunca versionada
~/.local/state/dibab/               el estado, el diario y los informes de sesión (lugar por de
 ↪  fecto)
~/Classement/Personnel/                 la bóveda del ámbito personal
~/Classement/Professionnel/             la bóveda del ámbito profesional
```

Instalación desde el paquete Debian

En Debian 12 y 13 y en Ubuntu 22.04 y 24.04, un solo archivo instala el programa y sus dependencias (`python3`, `poppler-utils`, `util-linux`, `openssh-client`). No contiene ni Ollama, ni un modelo, ni una configuración. Descargue `dibab_0.9.0-alpha_all.deb` desde la página de descarga de dibab.davalan.fr y, desde la carpeta que lo contiene, ejecute:

```
sudo apt install ./dibab_0.9.0-alpha_all.deb
```

Escriba la configuración en el idioma de su sesión y siga luego los pasos 4 y 5 de la instalación desde las fuentes que figura más abajo (la bóveda de cada dominio, las claves de la configuración):

```
dibab --initialiser
```

`--initialiser` escribe la configuración en el idioma de su sesión: la carpeta de verificación, su prefijo y la carpeta de duplicados toman los nombres de ese idioma. Crea el archivo con modo `600` en una carpeta con modo `700` y nunca reescribe un archivo que ya existe; una sesión posterior en otro idioma no los renombra.

Compruebe la instalación, sin escribir nada:

```
dibab --version
dibab --diagnose
```

El paquete coloca el comando en `/usr/bin/dibab`, el programa en `/usr/lib/dibab/`, y los README, la configuración de ejemplo y la página de manual (`man dibab`, en cinco idiomas) en `/usr/share/`. La desinstalación conserva sus datos: `sudo apt remove dibab` y `sudo apt purge dibab` eliminan los archivos del paquete y nunca tocan su configuración, su estado, su diario ni sus documentos; la purga nombra lo que deja.

Nivel: instalación y diagnóstico validados en contenedor en Debian 12 y 13 y en Ubuntu 22.04 y 24.04 (instalación, `--version`, `--diagnose`, la negativa a escribir sin configuración y la de escribir en una bóveda no montada, una clasificación simulada, desinstalación y purga). Un contenedor no muestra ni el montaje de una bóveda cifrada ni una ejecución con su propio Ollama y sus propios documentos.

Instalación en Fedora, openSUSE Leap y Arch

En Fedora 44, openSUSE Leap 16.0 y Arch Linux, un solo archivo instala también el programa y sus dependencias. Descargue el archivo de su sistema desde la página de descarga de dibab.davalan.fr y, desde la carpeta que lo contiene, ejecute la línea de su sistema:

```
sudo dnf install ./dibab-*.fc44.noarch.rpm
sudo zypper install --allow-unsigned-rpm ./dibab-*.leap160.noarch.rpm
sudo pacman -U ./dibab-*-any.pkg.tar.zst
```

Los paquetes no están firmados. El resto es igual que con el paquete Debian: `dibab --initialiser`, después `dibab --version` y `dibab --diagnose`. La desinstalación del paquete (`sudo dnf remove dibab`, `sudo zypper remove dibab`, `sudo pacman -R dibab`) nunca toca su configuración, su estado, su diario ni sus documentos.

Nivel: instalación y diagnóstico validados en contenedor en Fedora 44, openSUSE Leap 16.0 y Arch Linux (instalación, `--version`, `--diagnose`, las dos negativas a escribir, una clasificación simulada, reinstalación sobre una instalación dañada, desinstalación). Un contenedor no muestra ni el montaje de una bóveda cifrada ni una ejecución con su propio Ollama y sus propios documentos.

Instalar desde las fuentes

1. Obtenga las fuentes:

```
mkdir -p ~/applications
git clone https://github.com/RonanDavalan/dibab.git ~/applications/dibab
```

2. Haga disponible el comando. El punto de entrada es ejecutable; un enlace en `~/.local/bin` le da el nombre `dibab` (en la mayoría de las distribuciones este directorio ya está en el `PATH` de un shell de inicio de sesión):

```
mkdir -p ~/.local/bin
ln -s ~/applications/dibab/dibab.py ~/.local/bin/dibab
```

3. Escriba la configuración en el idioma de su sesión (el capítulo 3 explica cada clave):

```
dibab --initialiser
```

`--initialiser` escribe la configuración en el idioma de su sesión: la carpeta de verificación, su prefijo y la carpeta de duplicados toman los nombres de ese idioma. Crea el archivo con modo `600` en una carpeta con modo `700` y nunca reescribe un archivo que ya existe; una sesión posterior en otro idioma no los renombra.

4. Cree la bóveda de cada ámbito y lea el tipo de su sistema de archivos (el capítulo 5 explica por qué):

```
mkdir -p ~/Classement/Personnel ~/Classement/Professionnel
findmnt -n -T ~/Classement/Personnel -o FSTYPE
```

```
ext4
```

El comando imprime el tipo del volumen que contiene el directorio, por ejemplo `ext4`. Escriba ese tipo en `destinations.<domain>.systemes_fichiers_acceptes` para un directorio ordinario, o abra primero su volumen cifrado y escriba el tipo que `findmnt` indica para él.

5. Edite `~/.config/dibab/config.json`, al menos las claves del capítulo 3 marcadas como obligatorias, y luego compruebe la instalación sin escribir nada:

```
dibab --version
```

```
dibab 0.9.0-alpha
```

El comando imprime una línea y termina. No lee ninguna configuración, no toca ningún archivo y no habla con ningún servicio.

Ejecutar las pruebas (opcional)

Las fuentes traen sus propias pruebas, que necesitan `pytest` en un entorno virtual. No hacen falta para usar Dibab; ejecútelas después de una actualización o si sospecha de la instalación:

```
cd ~/applications/dibab && \
  python3 -m venv venv && \
  venv/bin/pip install -r requirements-dev.txt && \
  venv/bin/python -m pytest -q tests
```

La última línea indica las pruebas superadas y las omitidas; las omitidas son los ensayos contra un Ollama en marcha, que el conjunto de pruebas no exige.

Dónde se busca la configuración

Dibab lee el primer archivo que encuentra entre, en este orden: la opción `--config <path>`, la variable de entorno `DIBAB_CONFIG`, `$XDG_CONFIG_HOME/dibab/config.json` (`~/.config/dibab/config.json` cuando la variable está vacía) y luego `config.json` junto a `dibab.py`. No hay ninguna fusión: el primer archivo encontrado es el único que se lee. Una opción o una variable que nombra un archivo inexistente detiene el programa con el código de salida 1. Nunca recurre a otro archivo, porque entonces funcionaría con una configuración que usted no designó, es decir, con otra bóveda.

Configuración

En Dibab no hay nada escrito a fuego. Cada ruta, nombre de modelo, umbral, plazo y regla de nombrado vive en un único archivo JSON, `config.json`. Cambiar el comportamiento de la herramienta es cambiar este archivo, no el código. Las fuentes traen un ejemplo completo y realista, `config.exemple.json`; este capítulo lo explica bloque por bloque. El capítulo 2 dice dónde busca el archivo el programa.

Las rutas escritas con ~ o `$HOME` se expanden. El ejemplo no contiene ninguna ruta absoluta. Nunca se escribe un secreto en este archivo: la contraseña del correo va en una variable de entorno, un token de Google en un archivo aparte con el modo `600` (capítulo 7).

Una clave ausente no se sustituye en silencio por una suposición. El programa se detiene con el código de salida 1 y nombra las claves que faltan; por eso `dibab --diagnose` (capítulo 9) es lo primero que se ejecuta después de una modificación.

Las claves que debe escribir

Abra el archivo que escribió `dibab --initialiser` y cambie al menos estas:

Clave	Qué escribir
`destinations.personnel.racine`, `destinations.professionnel.racine`	la raíz de la bóveda de cada ámbito
`destinations.*.systemes_fichiers_acceptes`	el tipo de sistema de archivos que establece que la bóveda está montada (capítulo 5)
`personnes`	los nombres de pila de los miembros del hogar; dan nombre a las carpetas del ámbito personal
`sources_locales[].chemin`	los directorios que se escanean
`ollama.roles`	para cada rol, los modelos preferidos, en orden (`config.exemple.json` deja vacías las tres listas)
`suivi.fichier_date`, `suivi.journal`, `suivi.sessions`	opcionales, pero recomendadas dentro de la bóveda (capítulo 5)

Elimine las entradas de ejemplo de `sources_distantes` y de `sources_imap` que no use: una fuente declarada es una fuente leída.

`destinations`: la bóveda de cada ámbito

Una entrada por ámbito, `personnel` y `professionnel`.

Clave	Significado
`racine`	la raíz de la bóveda de este ámbito
`systemes_fichiers_acceptes`	los tipos de sistema de archivos que establecen que la bóveda está montada; el ejemplo espera `fuse.gocryptfs`

Clave	Significado
sous_repertoire_classement	la carpeta, bajo la raíz, donde se clasifican los documentos de este ámbito (Administratif para el ámbito personal, Classement para el profesional)
sous_repertoire_a_verifier	la carpeta de los documentos de los que Dibab no está seguro (POR_VERIFICAR/)

El nombre de la carpeta por verificar y su prefijo se escriben en la inicialización en el idioma de la sesión: sous_repertoire_a_verifier fija la carpeta y nommage.prefixe_a_verifier el prefijo. La carpeta en la raíz de una bóveda que recibe los duplicados se fija con destination.repertoire_doublon.

destination, arborescence, executer_sur

Clave	Significado
destination.sous_repertoire_logs	la carpeta de los registros de negocio dentro de una bóveda (logs); los planes de inventario se escriben allí
destination.vide_poches_prefixe	el prefijo de las «carpetas de entrada» (vide-poches-), donde esperan los archivos recuperados
destination.repertoire_doublon	la carpeta, en la raíz de una bóveda, que recibe un documento cuya huella ya está clasificada (duplicado/)
arborescence.personnel	la subruta de un documento personal bajo su carpeta de clasificación: {personne}/{annee} da Alice/2026
arborescence.professionnel	lo mismo para uno profesional: {annee}/{categorie} da 2026/Factures
arborescence.personne_inconnue	la carpeta que se usa cuando la persona no está en personnes (Commun)
executer_sur	los nombres de máquina (hostname) donde se permite --execute; una lista vacía significa cualquier máquina, y --execute en otra se detiene con el código de salida 4

suivi: estado, diario e informes

Clave	Significado
fichier_date	el archivo dernier_scan.json, que recuerda la fecha del último escaneo de cada fuente
journal	el diario de acciones journal.jsonl, una línea JSON por acción, releído por --rollback; contiene rutas, nunca el texto de un documento
sessions	la carpeta de los informes de sesión
logs_systeme	reservado para un registro técnico; hoy nada lo escribe

Clave	Significado
securite_jours	cuántos días antes del último escaneo reanuda un nuevo escaneo (1), para que no se pase por alto un archivo añadido en paralelo
premier_scan_jours	hasta dónde retrocede el primerísimo escaneo de una fuente (90 días)
rotation.journal_taille_max_mo	el tamaño a partir del cual se rota el diario (20 MB)
rotation.journal_archives_max	cuántos diarios rotados se conservan (5)
rotation.sessions_max	cuántos archivos de sesión se conservan (100)

Las cuatro claves de ruta (fichier_date, journal, sessions, logs_systeme) son opcionales. Sin ellas, los archivos van bajo $XDG_STATE_HOME/dibab/, es decir ~/.local/state/dibab/, con los nombres dernier_scan.json, journal.jsonl, sessions/ y dibab.log, con el modo 600 para los archivos y 700 para las carpetas. Estos lugares por defecto están en claro en el disco: el capítulo 5 explica cómo trasladarlos a una bóveda. Otros tres archivos siguen a fichier_date, esté donde esté: rapatries.json, imap_deposes.json y dibab.verrou.

ollama: los modelos

Clave	Significado
url	la dirección de Ollama (http://localhost:11434); cada llamada es una simple petición HTTP hacia ella
roles.ocr	las preferencias ordenadas del modelo que lee las páginas escaneadas
roles.tri_rapide	las preferencias ordenadas del modelo que aparta lo que no es administrativo
roles.classification	los modelos que clasifican; todos se ejecutan con cada documento y votan
roles.tri_decision	opcional: las preferencias ordenadas de un modelo decision (ollama show indica la capacidad), que clasifica en una sola pasada, con umbrales, en lugar de tri_rapide
roles.objet	opcional: el pequeño modelo de generación que da el objeto de un documento cuando el modelo de decisión lo clasifica
seuil_confiance	entre 0 y 1; un ámbito solo se conserva si la proporción de modelos de clasificación que coinciden lo alcanza (0.4 en el ejemplo)
timeout_secondes	el plazo de cada petición a Ollama (120)
keep_alive_entre_pages	cuánto tiempo permanece cargado el modelo de OCR entre dos páginas de un documento (2m)

Al arrancar, Dibab pregunta a Ollama por los modelos instalados y conserva, para ocr y tri_rapide, la primera preferencia que está instalada; un nombre sin etiqueta equivale a :latest. Para classification,

conserva todas las preferencias instaladas, y votan. Un rol sin ningún modelo instalado detiene el programa con el código de salida 3, lista los modelos presentes y nombra la línea que hay que corregir, antes de que se escriba nada. Los dos roles opcionales nunca lo detienen: cuando `tri_decision` está declarado pero es inutilizable (ningún modelo instalado, sin capacidad `decision`, sin umbrales, sin modelo para `objet`, o un Ollama que no conoce `/v1/systemone`), la clasificación pasa por `tri_rapide` y la cabecera de la sesión dice por qué.

`extraction` y `classification`

Clave	Significado
`extraction.seuil_caracteres_alphanumeriques`	por debajo de esta cantidad de letras y cifras, un PDF lo lee el modelo de OCR en lugar de su capa de texto (`200`)
`extraction.pages_ocr`	cuántas páginas lee el OCR (`3`)
`extraction.resolution_ocr`	la resolución de las páginas generadas, en ppp (`200`)
`extraction.taille_max_texte`	el texto extraído se trunca a este número de caracteres (`20000`)
`extraction.repertoire_temporaire`	dónde se generan las páginas; manténgalo en memoria (`/dev/shm`)
`extraction.timeout_outils_secondes`	el plazo de `pdftotext` y de `pdftoppm` (`60`)
`extraction.consigne_ocr`	la instrucción que se envía al modelo de OCR con cada imagen
`classification.temperature`, `classification.num_ctx`	las opciones enviadas con cada petición (`0` y `8192`); un contexto demasiado corto corta el comienzo de la instrucción
`classification.caracteres_max_prompt`	qué parte del texto se envía a los modelos de clasificación (`12000`)
`classification.caracteres_max_tri`	qué parte del texto se envía al modelo de triaje (`3000`); el comienzo de un documento dice de qué se trata
`classification.keep_alive_entre_documents`	cuánto tiempo permanece cargado un modelo entre dos documentos (`2m`)
`classification.longueur_max_champ`, `longueur_max_citation`	la longitud máxima del remitente, del destinatario y del asunto conservados, y de un extracto citado (`80` y `120`)
`classification.marqueurs_mentions`	palabras cuya presencia es una señal profesional débil (`TVA`, `SIRET`, …)
`classification.seuils_tri_decision`	opcional: para cada modelo de decisión, `non_administratif` y `administratif` (entre 0 y 1) y `caracteres_min`; la probabilidad de no administrativo que da el modelo vale no administrativo desde el primero, administrativo por debajo del segundo, incierto entre ambos, y un texto de menos de `caracteres_min` letras y cifras es incierto. Los umbrales valen para un modelo y un `caracteres_max_tri`: vuelva a medirlos cuando cambie uno de los dos

El texto de un documento se trata como un dato, nunca como una instrucción: se delimita con una marca aleatoria, el modelo responde con un JSON cerrado que Dibab valida, y toda respuesta no válida cuenta como «incierta». Un modelo nunca puede desencadenar una acción.

categories, **personnes**, **emetteurs_connus**, **pieces_communes**

Clave	Significado
`categories`	la lista cerrada de categorías (`Factures`, `Aides`, `Impôts`, …); un modelo solo puede responder con una de ellas, o con «ninguna»
`personnes`	los miembros del hogar; un modelo solo puede responder con uno de ellos, o con «ninguno»
`emetteurs_connus`	los remitentes conocidos por categoría, cada uno `{"nom": ..., "domaine": "personnel" | "professionnel" | "mixte"}`; un remitente conocido es una señal de ámbito, y una lista vacía no da ninguna señal
`pieces_communes`	las categorías copiadas en la bóveda de cada ámbito, con el mismo nombre y la misma huella (`Identité`, `Domicile`, `RIB`), solo si todos los modelos de clasificación devuelven esa misma categoría

Las categorías y los remitentes conocidos del ejemplo son genéricos; adáptelos a su país.

nommage: el nombre de un documento clasificado

Clave	Significado
`format`	la plantilla de un documento clasificado, con `{date}`, `{emetteur}`, `{categorie}`, `{personne}`, `{objet}` (el asunto que devuelve el triaje), `{ext}` (`{date}_{emetteur}_{categorie}{ext}`)
`separateur`, `espaces_vers`	el separador y la sustitución de los espacios (_ y -)
`date_format`	el formato de la fecha (`%Y-%m-%d`)
`date_inconnue_prefixe`	el prefijo de la fecha de modificación que se usa cuando no se encuentra ninguna fecha en el documento (~)
`emetteur_inconnu`	el nombre que se usa cuando el remitente es desconocido (`Inconnu`)
`prefixe_a_verifier`	el prefijo de un documento colocado en la carpeta por verificar (`[POR_VERIFICAR]`)

Clave	Significado
format_a_verifier	la plantilla que sigue al prefijo, con las mismas variables y {origine}, el nombre de origen; una parte vacía o desconocida se retira con su separador, y el nombre entero no pasa de 180 caracteres ({date}_{emetteur}_{objet}_{origine}{ext}); sin esta clave, el documento conserva su nombre de origen
suffixe_collision	el sufijo de un nombre ya ocupado, a partir de 2 (_{n})

Con la configuración que escribe dibab --initialiser, un aviso de un organismo público fechado el 15 de marzo de 2026 se convierte en 2026-03-15_Organisme-public_Aides.pdf, y una carta sin membrete colocada en la carpeta por verificar se convierte en [POR_VERIFICAR]courrier-de-suivi-de-dossier_courrier-sans-en-tete.pdf.

extensions_cibles y scan

Clave	Significado
extensions_cibles	las extensiones que se toman (.pdf, .png, .jpg, .jpeg), comparadas sin distinguir mayúsculas de minúsculas
scan.recursif_par_defaut	escanear las subcarpetas (true); cada fuente puede modificarlo
scan.exclusions_automatiques	las subcarpetas que nunca se escanean, para todas las fuentes (rebut, y la carpeta nombrada por destination.repertoire_doublon)
scan.suivre_symlinks	seguir los enlaces simbólicos (false, para que el escaneo no pueda salir de su perímetro)
scan.ignorer_dotfiles	ignorar los archivos y carpetas ocultos (true)

sources_locales

Cada entrada es un directorio que se escanea.

Clave	Significado
id	un nombre único; es la clave de la fuente en dernier_scan.json
chemin	el directorio
date_reference	modification (recomendado) o creation; la fecha que se usa para conservar los archivos recientes
recursif	sustituye a scan.recursif_par_defaut para esta fuente
exclusions	nombres de subcarpetas que se omiten solo para esta fuente

El primer escaneo de una fuente retrocede `suivi.premier_scan_jours` días; los siguientes reanudan desde la fecha del último escaneo completo, menos `suivi.securite_jours`.

`ssh`, `sources_distantes`, `imap`, `oauth2`, `sources_imap`

Estos bloques se explican donde se usan: `ssh` y `sources_distantes` en el capítulo 6, `imap`, `oauth2` y `sources_imap` en el capítulo 7.

`mesure` e `inventaire`

Clave	Significado
`inventaire.exclusions`	los nombres de carpetas que `--inventaire` no lee (capítulo 8)
`mesure.graine`, `documents_par_coffre`, `exclusions`, `taux_erreur_max`	leídas solo por la herramienta de medición de las fuentes (`_outils/mesurer_domaines.py`), que toma una muestra de documentos clasificados y confronta los modelos con la bóveda; los valores del ejemplo son un punto de partida

Primera pasada

Una primera pasada tiene cuatro momentos: colocar los documentos en un directorio de origen, simular con `--dry-run`, leer lo que propone la simulación y luego clasificar de verdad con `--execute`. Toda clasificación puede deshacerse con `--rollback`. Empiece siempre por la simulación: no mueve nada, no escribe nada en ninguna bóveda y muestra exactamente adónde iría cada documento.

Este capítulo usa el directorio de origen `~/Classement/Entrant` de la configuración del área de prueba, que contiene cuatro documentos ficticios: un aviso de prestaciones familiares, una factura de alojamiento web, una receta de tarta y una carta sin ningún membrete.

```
ls ~/Classement/Entrant

courrier-sans-en-tete.pdf
facture-hebergement.pdf
notification-organisme.pdf
recette-gateau.pdf
```

Simular: `--dry-run`

`--dry-run` es el modo por defecto: `dibab` a secas hace lo mismo. Necesita Ollama, porque se ejecuta toda la clasificación, pero no mueve ningún archivo y no escribe ni `dernier_scan.json` ni ningún documento clasificado.

```
dibab --dry-run

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : dry-run
  Configuration      : ~/.config/dibab/config.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

Source: entrant (~/Classement/Entrant)
  - notification-organisme.pdf                        (928 B, mtime 2026-11-23)
  - facture-hebergement.pdf                           (922 B, mtime 2026-11-23)
  - recette-gateau.pdf                                (895 B, mtime 2026-11-23)
  - courrier-sans-en-tete.pdf                         (901 B, mtime 2026-11-23)
  Source total: 4 file(s), 3.6 KB

============================================================
Scan total: 4 candidate file(s), 3.6 KB
============================================================
Text extraction (the text is neither shown nor written)
  - notification-organisme.pdf: text layer, 258 character(s), 1 page(s), 0.02 s
  - facture-hebergement.pdf: text layer, 246 character(s), 1 page(s), 0.02 s
```

```
  - recette-gateau.pdf: text layer, 226 character(s), 1 page(s), 0.02 s
  - courrier-sans-en-tete.pdf: text layer, 236 character(s), 1 page(s), 0.02 s

Classification (dry-run, nothing is moved; the text is neither shown nor written)
model 1/1, document 1/3
model 1/1, document 2/3
model 1/1, document 3/3
  - notification-organisme.pdf: triage administrative, domain personnel, agreement 1/1
      subject: notification de paiement d'allocation
      -> vault personnel, classement: ~/Classement/Personnel/Administratif/Alice/2026/2026-03-1
       ↪  5_Organisme-public_Aides.pdf
  - facture-hebergement.pdf: triage administrative, domain professionnel, agreement 1/1
      subject: facture d'hébergement
      -> vault professionnel, classement: ~/Classement/Professionnel/Classement/2026/Factures/2
       ↪  026-03-15_Hebergeur-Exemple_Factures.pdf
  - recette-gateau.pdf: triage not administrative, domain not applicable
      subject: recette de gâteau
      -> stays in place (not administrative)
  - courrier-sans-en-tete.pdf: triage administrative, domain incertain, agreement 1/1
      subject: courrier de suivi de dossier
      reason: no recipient quoted and no known issuer
      -> vault personnel, a_verifier: ~/Classement/Personnel/POR_VERIFICAR/[POR_VERIFICAR]courr
       ↪  ier-de-suivi-de-dossier_courrier-sans-en-tete.pdf (not moved)
```

Las líneas de progreso (`model 1/1, document 1/3`) van a la salida de error, una por modelo y por documento. La primera pasada es lenta, ya que cada modelo de clasificación lee cada documento, y estas líneas le indican que el programa no está colgado.

Leer el informe

El informe tiene cuatro partes, en este orden.

1. La cabecera: la versión y la hora de la sesión, el modo, el archivo de configuración leído, cada bóveda con la comprobación de su montaje, la ruta del diario de sesión, el modelo conservado para cada rol y la regla de consenso.
2. El escaneo: para cada fuente, los archivos candidatos con su tamaño y su fecha, y luego los totales. Un archivo es candidato cuando su extensión figura en `extensions_cibles` y su fecha es suficientemente reciente.
3. La extracción del texto: para cada candidato, cómo se obtuvo su texto (`text layer`, u `OCR` con el modelo que lo leyó), cuántos caracteres y páginas, y cuánto tiempo tardó. El texto en sí no se muestra ni se escribe.
4. La clasificación: un bloque por documento. La primera línea da el triaje (administrativo, no administrativo o incierto), el ámbito que Dibab conserva y el acuerdo, que cuenta los modelos de clasificación que devuelven ese ámbito entre los consultados (`agreement 2/3`: dos modelos de tres). La línea `subject:` dice en pocas palabras qué es el documento, tal como lo leyó el modelo de triaje; nunca es un extracto del texto. Para un documento enviado a la carpeta por verificar, la línea `reason:` dice por qué, según una lista cerrada. La última línea, tras `->`, da el destino.

En la ejecución, una última línea nombra el informe de sesión.

Un destino se lee `-> vault <domain>, <zone>: <path>`. La zona indica lo que va a ocurrir:

Zona o texto	Significado
`classement`	el documento se clasifica en la zona de clasificación de esa bóveda, bajo su nombre normalizado
`a_verifier`	Dibab no está seguro: el documento va a la carpeta por verificar, `POR_VERIFICAR/`, bajo el prefijo `[POR_VERIFICAR]` seguido de su fecha, su remitente y su asunto cuando el documento los da, y de su nombre de origen; la simulación añade `(not moved)` porque no se clasifica nada en la zona de clasificación
`stays in place (not administrative)`	el triaje apartó el documento; no se mueve ni se renombra, y se lista de nuevo en cada pasada
`(duplicate: would go to duplicado/)`	la huella del documento ya está clasificada; va a la carpeta `duplicado/` en lugar de clasificarse dos veces

El ámbito `incertain` lleva siempre a la carpeta por verificar, y lo mismo ocurre con un documento sobre el que los modelos no coinciden. Una pieza común (un documento de identidad, un justificante de domicilio, un documento de identidad bancaria) solo se propone en la bóveda de cada ámbito cuando todos los modelos de clasificación devuelven la misma categoría; si no, va a la carpeta por verificar. En caso de duda, Dibab prefiere una decisión humana a una clasificación arriesgada.

Comprobar el encaminamiento sin los modelos

Para ver a qué destino lleva un veredicto sin esperar a ningún modelo, entregue a `--dry-run` un archivo JSON que asocie un nombre de archivo con un veredicto. Se le aplica la tabla de encaminamiento y se muestra el destino. La opción solo existe con `--dry-run`: el programa la rechaza, con el código de salida 2, junto con `--execute`.

```
cat > verdicts.json <<'EOF'
{
  "notification-organisme.pdf": {"administratif": true, "domaine": "personnel", "categorie": "A
   ↪  ides", "personne": "Alice", "date": "2026-03-15", "emetteur": "Organisme public"},
  "recette-gateau.pdf": {"administratif": false, "categorie": "Inconnue"}
}
EOF
dibab --dry-run --verdicts-simules verdicts.json | sed -n '/Simulated routing/,$p'

model 1/1, document 1/3
model 1/1, document 2/3
model 1/1, document 3/3
Simulated routing (verdicts supplied, nothing is moved)
  - notification-organisme.pdf: domain personnel
      -> vault personnel, classement: ~/Classement/Personnel/Administratif/Alice/2026/2026-03-1
       ↪  5_Organisme-public_Aides.pdf
```

```
  - facture-hebergement.pdf: no simulated verdict
  - recette-gateau.pdf: not administrative, stays in place
  - courrier-sans-en-tete.pdf: no simulated verdict
```

Clasificar: `--execute`

Cuando la simulación le convenga, ejecute el mismo comando con `--execute`. Cada documento se copia a su destino, la copia se verifica por su huella SHA256 y solo entonces se retira el original del origen. No se retira nada antes de que la copia quede establecida como idéntica.

```
dibab --execute

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : execute
  Configuration      : ~/.config/dibab/config.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

Source: entrant (~/Classement/Entrant)
  - notification-organisme.pdf                          (928 B, mtime 2026-11-23)
  - facture-hebergement.pdf                             (922 B, mtime 2026-11-23)
  - recette-gateau.pdf                                  (895 B, mtime 2026-11-23)
  - courrier-sans-en-tete.pdf                           (901 B, mtime 2026-11-23)
  Source total: 4 file(s), 3.6 KB

============================================================
Scan total: 4 candidate file(s), 3.6 KB
============================================================
Text extraction (the text is neither shown nor written)
  - notification-organisme.pdf: text layer, 258 character(s), 1 page(s), 0.02 s
  - facture-hebergement.pdf: text layer, 246 character(s), 1 page(s), 0.02 s
  - recette-gateau.pdf: text layer, 226 character(s), 1 page(s), 0.02 s
  - courrier-sans-en-tete.pdf: text layer, 236 character(s), 1 page(s), 0.02 s

Filing (execution: each document is copied, checked by SHA256, then removed from its source)
model 1/1, document 1/3
model 1/1, document 2/3
model 1/1, document 3/3
  - notification-organisme.pdf: classe
      -> ~/Classement/Personnel/Administratif/Alice/2026/2026-03-15_Organisme-public_Aides.pdf
  - facture-hebergement.pdf: classe
      -> ~/Classement/Professionnel/Classement/2026/Factures/2026-03-15_Hebergeur-Exemple_Factu
       ↪  res.pdf
  - recette-gateau.pdf: not administrative, stays in place
  - courrier-sans-en-tete.pdf: a_verifier
```

```
      -> ~/Classement/Personnel/POR_VERIFICAR/[POR_VERIFICAR]courrier-de-suivi-de-dossier_courr⌋
       ↪  ier-sans-en-tete.pdf

Session report: ~/.local/state/dibab/sessions/rapport_2026-11-23T12:09:03.json
```

En la ejecución, cada línea termina con la acción realizada: `classe` (clasificado en la zona de clasificación), `a_verifier` (colocado en la carpeta por verificar) o `doublon` (colocado en la carpeta `duplicado/`). Estas palabras son las mismas en todos los idiomas, como las acciones escritas en el diario.

Los documentos están ahora donde decía el informe:

```
find ~/Classement -type f | sort

~/Classement/Entrant/recette-gateau.pdf
~/Classement/Personnel/Administratif/Alice/2026/2026-03-15_Organisme-public_Aides.pdf
~/Classement/Personnel/POR_VERIFICAR/[POR_VERIFICAR]courrier-de-suivi-de-dossier_courrier-sans-⌋
 ↪  en-tete.pdf
~/Classement/Professionnel/Classement/2026/Factures/2026-03-15_Hebergeur-Exemple_Factures.pdf
```

Tres archivos han salido de `Entrant`. La receta se quedó: no es administrativa, y Dibab no toca lo que no se le ha pedido clasificar.

Lo que escribe `--execute`

Cada movimiento se escribe en el diario de acciones, `suivi.journal`, una línea JSON por acción:

```
head -n 1 ~/.local/state/dibab/journal.jsonl

{"schema_version": 1, "session": "2026-11-23T12:09:03", "timestamp": "2026-11-23T12:09:03",
 ↪  "evenement": "action", "id_action": "<identifiant>", "action": "classe", "domaine": "person⌋
 ↪  nel", "origine": "~/Classement/Entrant/notification-organisme.pdf", "empreinte_avant":
 ↪  "6850987e722babfbca9f474ce8971afc532762ac07cd767308c8294fc1383d28", "destinations":
 ↪  [{"chemin": "~/Classement/Personnel/Administratif/Alice/2026/2026-03-15_Organisme-public_Ai⌋
 ↪  des.pdf", "empreinte_apres":
 ↪  "6850987e722babfbca9f474ce8971afc532762ac07cd767308c8294fc1383d28"}]}
```

La línea contiene la acción y su identificador, el ámbito, las rutas y las huellas antes y después, y ningún texto del documento. Cada sesión deja además dos archivos en la carpeta `suivi.sessions`: un diario de sesión (`session_<time>.jsonl`, los mismos eventos que la pantalla) y un informe (`rapport_<time>.json`) con el recuento de cada acción, los fallos y la duración, sin ningún nombre de documento. La fecha del escaneo de cada fuente solo avanza en `dernier_scan.json` para las fuentes leídas sin fallo. El capítulo 8 explica cómo se mantienen cortos estos archivos.

Solo se ejecuta una a la vez. Un segundo `--execute` o `--rollback` que encuentre tomado el cerrojo `dibab .verrou` se detiene con el código de salida 5 y no toca nada.

Deshacer: `--rollback`

`--rollback` deshace la última sesión de clasificación. `--rollback --session <time>` deshace una sesión dada, cuya hora es la impresa en su cabecera y empleada en los nombres de sus archivos.

```
dibab --rollback
```

```
Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : rollback
  Configuration      : ~/.config/dibab/config.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl

Rollback of session 2026-11-23T12:09:03: 3 undone, 0 failed
```

Cada documento vuelve a su lugar de origen, con la copia verificada por huella, y se retira su destino. Se añade una línea `annulation` al diario, que nunca se reescribe. Una pieza común se retira a la vez de la bóveda de cada ámbito. Una acción ya deshecha no se repite dos veces.

```
find ~/Classement/Entrant -type f | sort
```

```
~/Classement/Entrant/courrier-sans-en-tete.pdf
~/Classement/Entrant/facture-hebergement.pdf
~/Classement/Entrant/notification-organisme.pdf
~/Classement/Entrant/recette-gateau.pdf
```

Cuando una acción no puede deshacerse, no se retira nada por ella y se muestra el motivo, con el código de salida 1. El motivo es `destination modified since the filing` o `destination not found` cuando el documento se modificó o se movió a mano (devuélvalo a su sitio o déjelo), o `source occupied by another file` cuando otro archivo ocupó el lugar de origen (muévalo y vuelva a ejecutar el comando).

Códigos de salida

Código	Significado
`0`	éxito
`1`	configuración no encontrada, no válida o sin una clave; al menos una fuente o un documento con fallo; deshacer solo en parte; consentimiento OAuth2 fallido; `--diagnose` con al menos un `ÉCHEC`
`2`	montaje de la bóveda no comprobado antes de un modo que escribe; línea de comandos no válida
`3`	ningún modelo instalado para un rol, u Ollama no responde
`4`	`--execute` rechazado en esta máquina (clave `executer_sur`)
`5`	otra ejecución tiene el cerrojo `dibab.verrou`

`--version` devuelve `0`, o `2` para una línea de comandos no válida.

Bóveda y cifrado

Qué es una bóveda

Una bóveda es el directorio que recibe los documentos de un ámbito: una para el ámbito personal, otra para el profesional (`destinations.personnel.racine` y `destinations.professionnel.racine`). Dibab no la crea, no la monta, no la cifra ni la sincroniza. Espera que el directorio esté ahí y, como los documentos son sensibles, espera que se encuentre en un volumen cifrado que usted abre antes de la pasada.

El peligro es fácil de describir. Cuando un volumen cifrado está cerrado, su punto de montaje es solo un directorio vacío en el disco sin cifrar. Un programa que escribe ahí escribe los documentos en claro, por debajo del cifrado, sin ningún error. Por eso Dibab establece, antes de cada escritura, que cada bóveda está realmente montada.

La comprobación del montaje

Antes de `--execute`, `--rollback`, `--inventaire` o `--autoriser-google`, Dibab ejecuta `findmnt -n -T <root> -o FSTYPE` para la raíz de cada bóveda y compara el tipo de sistema de archivos con la lista `systemes_fichiers_acceptes` de esa bóveda. Comprueba también que la raíz admite escritura. Si el tipo no está en la lista, el montaje de la bóveda no queda establecido y el programa se detiene con el código de salida 2 sin tocar nada.

La configuración de ejemplo espera `fuse.gocryptfs`, el tipo que `findmnt` indica para un volumen gocryptfs montado. Con un directorio ordinario en su lugar, que es el aspecto de una bóveda cerrada, el rechazo es inmediato:

```
dibab --execute --config ~/.config/dibab/config-ferme.json; echo "exit code: $?"

ERROR: vault personnel: vault not mounted (type ext4, expected fuse.gocryptfs) (~/Classement/Pe⌋
↪  rsonnel)
ERROR: vault professionnel: vault not mounted (type ext4, expected fuse.gocryptfs) (~/Classemen⌋
↪  t/Professionnel)
        Nothing was touched. Mount the destination directory, then run again.
exit code: 2
```

La simulación no se detiene ante una bóveda cuyo montaje no está establecido: lo señala con `[!]` en la cabecera y continúa, para que pueda leer las propuestas antes de abrir la bóveda. No escribe su diario de sesión en una ruta que la bóveda contenga, porque eso dejaría el diario en claro en el disco.

Elegir los tipos que se aceptan

Lea el tipo de su propio volumen una vez abierto, con el comando del capítulo 2, y escríbalo en la lista:

```
"systemes_fichiers_acceptes": ["fuse.gocryptfs"]
```

Si cifra de otra manera (LUKS, eCryptfs, un volumen de otra herramienta), abra el volumen y escriba el tipo que `findmnt` devuelve para él. En un volumen sin cifrar, la comprobación ya no distingue un montaje ausente de uno presente: renuncia entonces a esta protección, y el diagnóstico del capítulo 9 le recuerda que el estado está en claro.

Un ejemplo con gocryptfs

gocryptfs no forma parte de Dibab, y nada en Dibab lo exige. Es una manera de construir la bóveda de cada ámbito. Cree cada directorio cifrado una vez, eligiendo una frase de contraseña que guarde fuera de la máquina; el comando se la pide.

```
mkdir -p ~/Chiffre/Personnel ~/Chiffre/Professionnel
gocryptfs -init ~/Chiffre/Personnel
gocryptfs -init ~/Chiffre/Professionnel
```

Abra cada bóveda antes de una pasada. El comando vuelve a pedir la frase de contraseña y monta la vista en claro en el directorio nombrado en la configuración:

```
gocryptfs ~/Chiffre/Personnel ~/Classement/Personnel
gocryptfs ~/Chiffre/Professionnel ~/Classement/Professionnel
```

Ciérrelas después de la pasada:

```
fusermount3 -u ~/Classement/Personnel
fusermount3 -u ~/Classement/Professionnel
```

Los directorios cifrados `~/Chiffre/...` son lo que usted respalda y sincroniza. Dibab no sabe nada de ellos.

Dónde vive el estado

Junto a los documentos, Dibab mantiene un estado: la fecha del último escaneo de cada fuente, el diario de acciones, los diarios e informes de sesión, el registro de los archivos recuperados y el cerrojo. Sin ninguna ruta `suivi.*` en la configuración, todo ello va bajo `~/.local/state/dibab/` (o `$XDG_STATE_HOME/dibab/`), en claro, con el modo `700` para las carpetas y `600` para los archivos. El diario contiene nombres de archivos y remitentes; eso ya es revelador. El diagnóstico lo señala como advertencia: `AVERTISSEMENT emplacement _etat`.

Para cifrar el estado, colóquelo dentro de una bóveda, con tres claves del bloque `suivi`:

```
"suivi": {
  "fichier_date": "~/Classement/Personnel/logs/dernier_scan.json",
  "journal": "~/Classement/Personnel/logs/journal.jsonl",
  "sessions": "~/Classement/Personnel/logs/sessions"
}
```

Los demás archivos del estado (`rapatries.json`, `imap_deposes.json`, `dibab.verrou`) siguen a `fichier_ date`. Con este ajuste, el diagnóstico ya no señala ninguna advertencia.

Espacio en el disco

Dibab mueve y copia, y nunca borra. Un documento clasificado se copia a su destino, se verifica y luego se retira del origen, de modo que el espacio en el disco no crece con una clasificación. Un documento que es un duplicado de otro ya clasificado va a la carpeta `duplicado/` de la bóveda en lugar de clasificarse dos veces; vaciar esa carpeta es decisión suya, nunca de la herramienta.

Fuentes por ssh

Una máquina remota del hogar se lee por `ssh`. Dibab lista los archivos de los directorios que usted nombra, los copia en una carpeta de entrada de la bóveda personal, verifica cada copia por su huella y luego los clasifica como documentos locales. Los archivos originales quedan intactos en la máquina remota: Dibab nunca retira ni modifica nada allí.

Lo que debe ofrecer la máquina remota

- Un sistema Linux con GNU `find` (el listado usa `-printf`), `sha256sum` y `cat`.
- Un alias `ssh` que abra una sesión sin contraseña. Todo lo relativo a la conexión, la dirección, el usuario y la clave, vive en el `~/.ssh/config` de la máquina local, bajo el alias; Dibab no sabe nada de ello. Por ejemplo, con una máquina inventada:

```
Host machine1
    HostName 192.0.2.10
    User archive
    IdentityFile ~/.ssh/id_ed25519
```

Compruebe el alias a mano antes de la primera pasada. El comando no debe imprimir nada y debe devolver 0:

```
ssh machine1 true
```

Declarar la máquina

Intervienen dos bloques de `config.json`: `ssh`, que da los plazos, y `sources_distantes`, que lista las máquinas.

```
"ssh": {
  "delai_connexion_secondes": 10,
  "delai_commande_secondes": 120,
  "delai_copie_secondes": 900
},

"sources_distantes": [
  {
    "id": "ssh-machine1",
    "alias_ssh": "machine1",
    "vide_poches_destination": "vide-poches-machine1",
    "repertoires": ["~/Téléchargements"],
    "date_reference": "modification",
    "recursif": true,
    "exclusions": []
  }
]
```

Clave	Significado
`ssh.delai_connexion_secondes`	el plazo para abrir la conexión (`10`)

Clave	Significado
`ssh.delai_commande_secondes`	el plazo de un listado o de una huella (`120`)
`ssh.delai_copie_secondes`	el plazo para copiar un archivo (`900`)
`id`	un nombre único; es la clave de la máquina en `dernier_scan.json`
`alias_ssh`	el alias de `~/.ssh/config`
`vide_poches_destination`	el nombre de la carpeta de entrada, sin ningún separador, creada bajo la raíz de la bóveda personal; los archivos recuperados llegan allí
`repertoires`	los directorios que se leen en la máquina remota; ~ es el directorio personal del usuario remoto
`date_reference`, `recursif`, `exclusions`	como en una fuente local (capítulo 3)

Los tres plazos de `ssh` son obligatorios en cuanto se declara una fuente remota.

Una primera pasada: listar y luego recuperar

La configuración de este capítulo declara la máquina y nada más. La simulación lista los archivos sin copiar nada:

```
dibab --dry-run --config ~/.config/dibab/config-ssh.json

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : dry-run
  Configuration      : ~/.config/dibab/config-ssh.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

Remote source: ssh-machine1 (nothing is fetched in dry-run)
  - releve-bancaire.pdf                              (remote, 920 B, mtime 2026-11-23)

============================================================
Scan total: 0 candidate file(s), 0 B
============================================================
Text extraction (the text is neither shown nor written)

Classification (dry-run, nothing is moved; the text is neither shown nor written)
```

El bloque `Remote source` muestra cada archivo con la palabra `remote`. No se escribe nada en la carpeta de entrada. En un primer escaneo, solo se listan los archivos modificados durante los últimos `suivi.premier_scan_jours` días (90 por defecto), y solo aquellos cuya extensión figura en `extensions_cibles`.

La ejecución recupera los archivos y luego los clasifica:

```
dibab --execute --config ~/.config/dibab/config-ssh.json

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : execute
  Configuration      : ~/.config/dibab/config-ssh.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

Transfer of remote sources (copy checked by fingerprint, nothing is removed on the remote side)
  ssh-machine1: 1 candidate(s)

Source: ssh-machine1-vide-poches (~/Classement/Personnel/vide-poches-machine1)
  - releve-bancaire.pdf                              (920 B, mtime 2026-11-23)
  Source total: 1 file(s), 920 B

============================================================
Scan total: 1 candidate file(s), 920 B
============================================================
Text extraction (the text is neither shown nor written)
  - releve-bancaire.pdf: text layer, 242 character(s), 1 page(s), 0.02 s

Filing (execution: each document is copied, checked by SHA256, then removed from its source)
model 1/1, document 1/1
  - releve-bancaire.pdf: classe
      -> ~/Classement/Personnel/Administratif/Bob/2026/2026-03-15_Banque-Exemple_Banques.pdf

Session report: ~/.local/state/dibab/sessions/rapport_2026-11-23T12:09:03.json
```

Cada archivo se copia en `vide-poches-machine1`, se relee y se compara por SHA256 con la huella calculada en la máquina remota. Una copia que no coincide no se conserva. El escaneo trata entonces la carpeta de entrada como una fuente local, sin ningún filtro de fecha, porque una carpeta de entrada se vacía a medida que avanza la clasificación: todo lo que contiene está esperando. El registro `rapatries.json`, junto a `dernier_scan.json`, recuerda los archivos recuperados por huella.

Volver a ejecutar

Una segunda ejecución no recupera nada nuevo. El archivo remoto se lista de nuevo, pero su huella está en el registro: por eso no se copia ni se clasifica una segunda vez:

```
dibab --execute --config ~/.config/dibab/config-ssh.json

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : execute
  Configuration      : ~/.config/dibab/config-ssh.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
```

```
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

Transfer of remote sources (copy checked by fingerprint, nothing is removed on the remote side)
  ssh-machine1: 1 candidate(s)

Source: ssh-machine1-vide-poches (~/Classement/Personnel/vide-poches-machine1)
  (no candidate file)
  Source total: 0 file(s), 0 B

============================================================
Scan total: 0 candidate file(s), 0 B
============================================================
Text extraction (the text is neither shown nor written)

Filing (execution: each document is copied, checked by SHA256, then removed from its source)

Session report: ~/.local/state/dibab/sessions/rapport_2026-11-23T12:09:03.json
```

La fecha de la máquina en `dernier_scan.json` solo avanza tras un listado completo en el que se verificó cada copia. Un archivo fallido se recupera de nuevo en la siguiente ejecución, sin duplicado. Un archivo que existe a la vez en la máquina remota y en un escaneo local va a la carpeta `duplicado/`: se clasifica una sola vez.

Cuando falla la conexión

Un rechazo detiene la máquina afectada, y solo esa; las demás fuentes continúan. Dibab no reintenta en bucle, y usted tampoco debería hacerlo.

Mensaje	Causa probable
`Could not resolve hostname` o `Permission denied`	la máquina está apagada, o el alias o la clave son erróneos; corríjalo y vuelva a ejecutar una sola vez
un plazo superado	aumente `ssh.delai_commande_secondes` (listado) o `ssh.delai_copie_secondes` (copia)
`remote directory partly readable or absent`	corrija `repertoires`
ningún candidato aunque la máquina tiene archivos	los archivos son más antiguos que `suivi.premier_scan_jours`, o su extensión no está en `extensions_cibles`

Buzones IMAP y cuentas de Google

Dibab lee los adjuntos de un buzón de correo por IMAP, con una contraseña o, para una cuenta de Google, con un token OAuth2. Lee sin modificar: el buzón se abre en solo lectura, ningún mensaje se marca como leído, no se pone ningún indicador y no se copia, mueve ni borra nada en el servidor.

Lo que se pide al servidor

El programa solo envía al servidor `EXAMINE` (la apertura de una carpeta en solo lectura), `UID SEARCH ... SINCE` (los mensajes recibidos desde una fecha) y `UID FETCH` de la fecha de recepción, de la estructura del mensaje y de un adjunto cada vez con `BODY.PEEK`, la forma que no marca el mensaje como leído. El cuerpo de un mensaje no se pide nunca. La conexión es siempre TLS, y el certificado se verifica contra el almacén del sistema: una fuente con `"ssl": false` se rechaza.

Un adjunto es una parte del mensaje que lleva un nombre y cuya disposición es `attachment`, o `inline` sin `Content-ID` (las imágenes de una firma llevan uno y se quedan en el mensaje). Su extensión debe figurar en `extensions_cibles`. Su nombre es un dato que aporta el remitente: se decodifica y se limpia como el nombre de un archivo recuperado (sin separador, sin carácter de control, nunca oculto, longitud acotada), y no entra ni en el diario de sesión ni en el informe.

Declarar un buzón

```
"imap": {
  "delai_secondes": 60,
  "taille_max_piece_mo": 25
},

"sources_imap": [
  {
    "id": "imap-foyer",
    "nom": "foyer",
    "serveur": "imap.example.com",
    "port": 993,
    "ssl": true,
    "authentification": "login",
    "utilisateur": "utilisateur@example.com",
    "mot_de_passe_env": "DIBAB_IMAP_FOYER",
    "dossiers": ["INBOX"],
    "vide_poches_destination": "vide-poches-foyer",
    "date_reference": "reception"
  }
]
```

Clave	Significado
`imap.delai_secondes`	el plazo de cada intercambio de red, conexión incluida (`60`)
`imap.taille_max_piece_mo`	un adjunto mayor que esto, según el tamaño que anuncia el servidor, no se recupera (`25`)

Clave	Significado
`id`	un nombre único; es la clave de la fuente en `dernier_scan.json`
`nom`	una etiqueta
`serveur`, `port`, `utilisateur`	la cuenta
`ssl`	debe ser `true`
`authentification`	`login` (contraseña en una variable de entorno) u `oauth2` (cuenta de Google); cualquier otro valor detiene la fuente
`mot_de_passe_env`	el nombre de la variable de entorno que contiene la contraseña (solo `login`)
`dossiers`	las carpetas que se leen, cada una abierta en solo lectura
`vide_poches_destination`	la carpeta de entrada, un nombre único, bajo la raíz de la bóveda personal
`date_reference`	`reception`: la fecha de recepción registrada por el servidor, nunca la cabecera `Date`, que escribe el remitente

Los dos bloques `imap` y `sources_imap` son obligatorios a la vez; una clave ausente detiene el programa con el código de salida 1 y la nombra.

La contraseña

La contraseña nunca está en `config.json`. Dibab la lee de la variable de entorno que nombra `mot_de_passe_env`, y nunca la escribe, la registra ni la muestra. Defínala en el terminal que lanza Dibab, sin escribirla en ninguna parte:

```
read -rs DIBAB_IMAP_FOYER && export DIBAB_IMAP_FOYER
```

Una fuente cuya variable no está definida se omite con una advertencia, sin ninguna conexión. Es también lo que le ocurre a la entrada de ejemplo de `config.exemple.json` mientras la deje declarada.

Una primera pasada: listar y luego depositar

La configuración de este capítulo declara el buzón y nada más. La simulación lista los adjuntos que se recuperarían, con la palabra `imap`, y no descarga ninguno:

```
dibab --dry-run --config ~/.config/dibab/config-imap.json

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : dry-run
  Configuration      : ~/.config/dibab/config-imap.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4
```

```
IMAP source: imap-foyer (nothing is downloaded in dry-run)
  - attestation-assurance.pdf                          (imap, 1274 B, received 2026-11-23)

============================================================
Scan total: 0 candidate file(s), 0 B
============================================================
Text extraction (the text is neither shown nor written)

Classification (dry-run, nothing is moved; the text is neither shown nor written)
```

La ejecución descarga cada adjunto, verifica su huella, lo deposita en la carpeta de entrada y lo clasifica como un documento local:

```
dibab --execute --config ~/.config/dibab/config-imap.json

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : execute
  Configuration      : ~/.config/dibab/config-imap.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

IMAP sources (read-only: nothing is modified or deleted in the mailbox)
IMAP source: imap-foyer
  - attestation-assurance.pdf                          (imap, 1274 B, received 2026-11-23)

Source: imap-foyer-vide-poches (~/Classement/Personnel/vide-poches-foyer)
  - attestation-assurance.pdf                          (928 B, mtime 2026-11-23)
  Source total: 1 file(s), 928 B

============================================================
Scan total: 1 candidate file(s), 928 B
============================================================
Text extraction (the text is neither shown nor written)
  - attestation-assurance.pdf: text layer, 258 character(s), 1 page(s), 0.02 s

Filing (execution: each document is copied, checked by SHA256, then removed from its source)
model 1/1, document 1/1
  - attestation-assurance.pdf: classe
      -> ~/Classement/Personnel/Administratif/Alice/2026/2026-03-15_Assurance-Exemple_Assurance
       ↪  s.pdf

Session report: ~/.local/state/dibab/sessions/rapport_2026-11-23T12:09:03.json
```

El archivo `imap_deposes.json`, junto a `dernier_scan.json`, recuerda lo que se depositó (carpeta, `UIDVALIDITY`, `UID`, parte y huella). Un adjunto cuya huella ya figura en él, o en el diario de acciones, no se

deposita de nuevo:

```
dibab --execute --config ~/.config/dibab/config-imap.json

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : execute
  Configuration      : ~/.config/dibab/config-imap.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl
  Model ocr           : essai-ocr:latest
  Model tri_rapide    : essai-tri:latest
  Model classification: essai-classification:latest
  Consensus          : 1 classification model(s), threshold 0.4

IMAP sources (read-only: nothing is modified or deleted in the mailbox)
IMAP source: imap-foyer
  - attestation-assurance.pdf                          (imap, 1274 B, received 2026-11-23)

Source: imap-foyer-vide-poches (~/Classement/Personnel/vide-poches-foyer)
  (no candidate file)
  Source total: 0 file(s), 0 B

============================================================
Scan total: 0 candidate file(s), 0 B
============================================================
Text extraction (the text is neither shown nor written)

Filing (execution: each document is copied, checked by SHA256, then removed from its source)

Session report: ~/.local/state/dibab/sessions/rapport_2026-11-23T12:09:03.json
```

La fecha de la fuente en `dernier_scan.json` solo avanza tras una lectura completa en la que cada depósito tuvo éxito.

Cuando falla la conexión

Un rechazo de autenticación o una conexión fallida detiene la fuente, sin segundo intento. Las demás fuentes continúan.

Mensaje	Causa probable
`environment variable ... not defined, source skipped`	exporte la contraseña en el terminal que lanza Dibab
`authentication refused by the server`	la contraseña o el usuario son erróneos; corríjalo y vuelva a ejecutar una vez, no en bucle
`connection impossible`	el servidor, el puerto o la red
`folder ... not found`	corrija `dossiers`
`attachment(s) above imap.taille_max_piece_mo`	aumente este valor

Cuentas de Google con OAuth2

Una cuenta de Google se lee con un token OAuth2 en lugar de una contraseña. Los pasos en la consola de Google le corresponden a usted, con la cuenta de Google que se va a leer, y las etiquetas de la consola cambian con el tiempo: las nociones son las de su documentación OAuth2.

1. En la consola de Google Cloud, cree un proyecto, active la API de Gmail y configure la pantalla de consentimiento de OAuth (tipo «externo», su dirección como contacto). Después publique la aplicación, para que su estado sea «En producción». Con el estado «Prueba», la autorización caduca a los siete días.
2. Cree una credencial OAuth del tipo «Aplicación de escritorio» y descargue su archivo JSON.
3. Guarde ese archivo en la bóveda, en una carpeta con el modo `700`, con el modo `600`. Dibab nunca lo modifica. Es el `client_fichier`.
4. Declare la fuente con `"authentification": "oauth2"`, `client_fichier` y `token_fichier` en lugar de `mot_de_passe_env`, y añada el bloque `oauth2`:

```
"oauth2": {
  "url_autorisation": "https://accounts.google.com/o/oauth2/v2/auth",
  "url_jeton": "https://oauth2.googleapis.com/token",
  "portee": "https://mail.google.com/",
  "marge_renouvellement_secondes": 300,
  "delai_secondes": 30,
  "delai_consentement_secondes": 300,
  "chemins_interdits": ["~/Documents"]
},

"sources_imap": [
  {
    "id": "imap-google",
    "nom": "foyer-google",
    "serveur": "imap.gmail.com",
    "port": 993,
    "ssl": true,
    "authentification": "oauth2",
    "utilisateur": "utilisateur@example.com",
    "client_fichier": "~/Classement/Personnel/logs/client-google.json",
    "token_fichier": "~/Classement/Personnel/logs/jeton-google.json",
    "dossiers": ["INBOX"],
    "vide_poches_destination": "vide-poches-google",
    "date_reference": "reception"
  }
]
```

Clave de `oauth2`	Significado
`url_autorisation`, `url_jeton`	los puntos de acceso de Google; deben ser `https`, salvo en el bucle local
`portee`	el alcance OAuth (`https://mail.google.com/`)

Clave de `oauth2`	Significado
`marge_renouvellement_secondes`	el token de acceso se renueva cuando le queda menos que esto (`300`)
`delai_secondes`	el plazo de cada intercambio con el servidor de autorización (`30`)
`delai_consentement_secondes`	cuánto espera `--autoriser-google` su respuesta (`300`)
`chemins_interdits`	carpetas donde se rechaza un token (`~/Documents`)

5. Dé su consentimiento una vez, desde un terminal de la máquina, con el `id` de la fuente:

```
dibab --autoriser-google imap-google
```

Dibab imprime una dirección que empieza por `https://accounts.google.com/`. Ábrala en un navegador. Acepte la advertencia «aplicación no verificada», que aparece porque la aplicación es suya, y luego el consentimiento. El programa recibe la respuesta en el bucle local, la comprueba, la canjea por un token y escribe el archivo de token con el modo `600` en una carpeta con el modo `700`. El archivo de token debe estar fuera de cualquier repositorio git y fuera de `chemins_interdits`; un archivo de token legible por el grupo o por otros detiene la fuente. Si se rechaza el consentimiento, si no llega ninguna respuesta a tiempo o si Google no devuelve ningún token de renovación, no se escribe nada; vuelva a ejecutar el comando.

A partir de ahí, el escaneo funciona como con una fuente de contraseña, sin ninguna variable de entorno: el token de acceso se renueva solo, y el token no aparece ni en el diario ni en pantalla. Un archivo de token que no existe no es un error: la fuente se omite, con el comando que hay que ejecutar. Si Google revoca la autorización (usted retiró el acceso de la aplicación, cambió la contraseña de la cuenta de Google o pasaron seis meses sin un escaneo), Dibab detiene la fuente con un mensaje explícito, y usted vuelve a ejecutar `--autoriser-google`.

El alcance `https://mail.google.com/` es el único que Google documenta para IMAP, y permite leer, escribir y borrar. Que el buzón permanezca sin modificar lo garantiza lo que Dibab envía, que se reduce a los comandos enumerados al principio de este capítulo, no el alcance. El capítulo 10 vuelve sobre ello.

Inventario y rotación

Leer lo que una bóveda ya contiene: `--inventaire`

Si ya tiene documentos en una bóveda, clasificados a mano o por otra herramienta, `--inventaire` los lee y escribe un plan de migración, sin mover, renombrar ni borrar nada. Es de solo lectura: antes y después, cada bóveda es idéntica. Necesita la comprobación del montaje del capítulo 5, ya que cada plan se escribe en su bóveda.

Para tener algo que leer, vuelva a clasificar los cuatro documentos del área de prueba; la reversión del capítulo 4 los devolvió a su origen:

```
dibab --execute | tail -n 9

model 1/1, document 1/3
model 1/1, document 2/3
model 1/1, document 3/3
  - notification-organisme.pdf: classe
      -> ~/Classement/Personnel/Administratif/Alice/2026/2026-03-15_Organisme-public_Aides.pdf
  - facture-hebergement.pdf: classe
      -> ~/Classement/Professionnel/Classement/2026/Factures/2026-03-15_Hebergeur-Exemple_Factu⌋
       ↪  res.pdf
  - recette-gateau.pdf: not administrative, stays in place
  - courrier-sans-en-tete.pdf: a_verifier
      -> ~/Classement/Personnel/POR_VERIFICAR/[POR_VERIFICAR]courrier-de-suivi-de-dossier_courr⌋
       ↪  ier-sans-en-tete.pdf

Session report: ~/.local/state/dibab/sessions/rapport_2026-11-23T12:09:03.json
```

Lea después la bóveda de cada ámbito:

```
dibab --inventaire

Dibab 0.9.0-alpha — session 2026-11-23T12:09:03
  Mode               : inventaire
  Configuration      : ~/.config/dibab/config.json
  Vault personnel     : ~/Classement/Personnel — mounted
  Vault professionnel : ~/Classement/Professionnel — mounted
  Session journal    : ~/.local/state/dibab/sessions/session_2026-11-23T12:09:03.jsonl

Inventory of what exists (read-only, nothing is moved)
  Vault personnel     : 3 listed document(s), 3 in the filed zone, 0 to file, 0 duplicate(s), 0
   ↪  present in the other vault, 1 in POR_VERIFICAR (not listed)
      plan written: ~/Classement/Personnel/logs/inventaire_2026-11-23T12:09:03.json
  Vault professionnel : 1 listed document(s), 1 in the filed zone, 0 to file, 0 duplicate(s), 0
   ↪  present in the other vault, 0 in POR_VERIFICAR (not listed)
      plan written: ~/Classement/Professionnel/logs/inventaire_2026-11-23T12:09:03.json
```

Para cada bóveda, el informe da el número de documentos leídos, cuántos están ya en la zona de clasificación, cuántos están fuera y deberían clasificarse, cuántos son duplicados de otro documento de la misma bóveda y cuántos están también presentes en la otra bóveda. El plan es un archivo JSON, `logs/inventaire_<time>`

`.json`, escrito en la bóveda que describe:

```
ls ~/Classement/Personnel/logs

inventaire_2026-11-23T12:09:03.json
```

El plan de cada bóveda nombra solo los archivos de esa bóveda. De la otra bóveda conserva un único dato, «esta huella también está allí», de modo que el plan de una bóveda nunca contiene la ruta de un archivo de la otra.

Cada documento del plan recibe una propuesta:

Propuesta	Significado
nada que hacer	el documento está en la zona de clasificación
ponerlo en una carpeta de entrada y luego leer el dry-run	el documento está fuera de la zona de clasificación: deposítelo en una fuente y deje que la siguiente pasada lo clasifique
idéntico a otro documento: se clasificará en `duplicado/` tras comprobarlo	la huella ya está presente

Las carpetas enumeradas en `inventaire.exclusions` (`logs`, `Archive`, `rebut`, y la carpeta nombrada por `destination.repertoire_doublon`, en el ejemplo), la carpeta por verificar, las carpetas de entrada y las carpetas ocultas no se leen.

Mantener corto el diario: la rotación

Dos archivos crecen con el uso: el diario de acciones, `suivi.journal`, y la carpeta de las sesiones, `suivi.sessions`. La rotación aplica los límites de `suivi.rotation` al arrancar `--execute` o `--rollback`:

Clave	Efecto
`journal_taille_max_mo`	cuando el diario supera este tamaño, se renombra `journal.1.jsonl`, los archivos más antiguos suben un número y empieza un diario nuevo
`journal_archives_max`	el número de archivos que se conservan; el más antiguo se retira cuando uno nuevo superaría ese número
`sessions_max`	el número de diarios e informes de sesión que se conservan; los más antiguos se retiran

Es el único borrado que realiza Dibab, y afecta a sus propios diarios, nunca a un documento. Un archivo que aún contiene una acción no deshecha no se retira jamás: la rotación espera antes que perder lo que `--rollback` debe poder leer. Deshaga una clasificación de la que dude antes de que su diario envejezca.

Los archivos del estado se escriben con el modo `600` y su carpeta con el modo `700`. Dónde se guardan es la elección descrita en el capítulo 5.

Diagnóstico

`--diagnose` comprueba la instalación, la configuración, Ollama y la bóveda de cada ámbito, una línea por comprobación, y no escribe nada: no se crea ningún archivo, no se mueve nada, no se toma ningún cerrojo. Abre una sola conexión, hacia Ollama, para leer la lista de modelos. No abre ninguna hacia una máquina remota ni hacia un buzón de correo, de modo que puede ejecutarse en cualquier momento sin riesgo. Realiza cada comprobación hasta el final, sea cual sea el resultado de las anteriores.

```
dibab --diagnose

OK python : 3.13.5
OK outil_pdftotext : /usr/bin/pdftotext
OK outil_pdftoppm : /usr/bin/pdftoppm
OK outil_findmnt : /usr/bin/findmnt
OK outil_ssh : /usr/bin/ssh
OK configuration : ~/.config/dibab/config.json
AVERTISSEMENT emplacement_etat : the state and the journal are in clear text, outside any vault:
 ↪  ~/.local/state/dibab
OK ollama : http://127.0.0.1:18434 answers, 3 model(s) installed
OK modele_ocr : essai-ocr:latest
OK modele_tri_rapide : essai-tri:latest
OK modele_classification : essai-classification:latest, 1 model(s) in the consensus
NON_CONFIGURÉ modele_tri_decision : ollama.roles.tri_decision absent or empty: the triage goes
 ↪  through ollama.roles.tri_rapide
OK destination_personnel : ~/Classement/Personnel
OK destination_professionnel : ~/Classement/Professionnel
Diagnostic: 12 OK, 1 warning(s), 0 failure(s), 1 not configured, 0 not checked
```

Cada línea se lee `<STATE> <identifier> : <message>`, y un resumen cierra el informe. El mensaje está en el idioma de su sesión; los estados y los identificadores no cambian nunca con el idioma, para que un script pueda compararlos.

Los estados

Estado	Significado
`OK`	la comprobación ha tenido éxito
`AVERTISSEMENT`	un defecto que no impide ningún modo, o un defecto de un componente que esta instancia no usa
`ÉCHEC`	la comprobación ha fallado en algo que exige la instalación o la configuración encontrada
`NON_CONFIGURÉ`	la configuración que lee la comprobación todavía no existe
`NON_VÉRIFIÉ`	la comprobación depende de otra que no ha tenido éxito, o de una configuración ausente

El código de salida es `0` cuando no hay ningún `ÉCHEC` y `1` cuando hay al menos uno. Una advertencia, una configuración ausente y un punto sin comprobar no lo cambian.

Las comprobaciones, en orden

Identificador	Qué se comprueba	Si no es OK
python	el intérprete es 3.10 o posterior	ÉCHEC por debajo
outil_pdftotext, outil_pdftoppm, outil_findmnt	la herramienta es ejecutable en el PATH	ÉCHEC si falta
outil_ssh	ssh es ejecutable en el PATH	ÉCHEC si la configuración declara sources_distantes, AVERTISSEMENT en caso contrario
configuration	se encuentra una configuración (capítulo 2) y es válida: JSON legible, las claves que exige cada modo	ninguna encontrada: NON_CONFIGURÉ, y el mensaje lista los cuatro lugares probados; encontrada pero no válida: ÉCHEC, con las claves que faltan
emplacement_etat	el estado y el diario están bajo la raíz de una bóveda declarada	fuera de cualquier bóveda: AVERTISSEMENT (el estado está en claro)
ollama	GET <ollama.url>/api/tags responde dentro de ollama.timeout_secondes	sin respuesta: ÉCHEC
modele_ocr, modele_tri_rapide, modele_classification	el rol tiene al menos un modelo instalado entre sus preferencias	ninguno instalado o lista vacía: ÉCHEC, y el mensaje nombra ollama.roles.<role>
modele_tri_decision	la clasificación por decisión, opcional (capítulo 3): primer modelo instalado de ollama.roles.tri_decision, capacidad decision, umbrales válidos en classification.seuils_tri_decision, un modelo instalado para ollama.roles.objet	rol ausente o vacío: NON_CONFIGURÉ; declarado pero inutilizable: AVERTISSEMENT, el mensaje nombra la clave que corregir, y la clasificación pasa por tri_rapide
destination_personnel, destination_professionnel	la comprobación del montaje del capítulo 5: el tipo de sistema de archivos de la raíz es aceptado y la raíz admite escritura	no establecida: ÉCHEC, y el mensaje da el motivo

Una fuente con contraseña cuya variable de entorno no está definida se señala con una advertencia en la comprobación configuration, ya que el programa omite una fuente así en la ejecución. Ninguna comprobación escribe un secreto, un nombre de documento o el contenido de una variable de entorno.

Leer un fallo

Una configuración del área de prueba declara para cada ámbito una bóveda que debe ser `fuse.gocryptfs`, mientras que los directorios son ordinarios, que es el aspecto de una bóveda cerrada:

```
dibab --diagnose --config ~/.config/dibab/config-ferme.json; echo "exit code: $?"

OK python : 3.13.5
OK outil_pdftotext : /usr/bin/pdftotext
OK outil_pdftoppm : /usr/bin/pdftoppm
OK outil_findmnt : /usr/bin/findmnt
OK outil_ssh : /usr/bin/ssh
OK configuration : ~/.config/dibab/config-ferme.json
AVERTISSEMENT emplacement_etat : the state and the journal are in clear text, outside any vault:
 ↪  ~/.local/state/dibab
OK ollama : http://127.0.0.1:18434 answers, 3 model(s) installed
OK modele_ocr : essai-ocr:latest
OK modele_tri_rapide : essai-tri:latest
OK modele_classification : essai-classification:latest, 1 model(s) in the consensus
NON_CONFIGURÉ modele_tri_decision : ollama.roles.tri_decision absent or empty: the triage goes
 ↪  through ollama.roles.tri_rapide
ÉCHEC destination_personnel : vault not mounted (type ext4, expected fuse.gocryptfs) (~/Classem
 ↪  ent/Personnel)
ÉCHEC destination_professionnel : vault not mounted (type ext4, expected fuse.gocryptfs) (~/Cla
 ↪  ssement/Professionnel)
Diagnostic: 10 OK, 1 warning(s), 2 failure(s), 1 not configured, 0 not checked
exit code: 1
```

Las dos comprobaciones `destination_*` fallan y el código es `1`, mientras que Ollama, los modelos y las herramientas están bien. El mensaje dice qué hacer: abrir la bóveda, o escribir el tipo correcto en `systemes_fichiers_acceptes`.

Otros fallos frecuentes:

Línea	Qué hacer
`ÉCHEC ollama`	el servicio no responde: compruebe `ollama.url` y luego que Ollama se está ejecutando (`systemctl status ollama`)
`ÉCHEC modele_<role>`	corrija `ollama.roles.<role>` según `ollama list`
`AVERTISSEMENT modele_tri_decision`	instale el modelo de decisión o el modelo de objeto nombrado, o escriba sus umbrales; mientras tanto, la clasificación pasa por `tri_rapide`
`ÉCHEC destination_<domain>`	abra la bóveda nombrada, o corrija los tipos aceptados
`ÉCHEC configuration`	añada las claves nombradas, copiando el bloque de `config.exemple.json`

Una instalación recién hecha

Justo después de la instalación, sin configuración y sin Ollama, el diagnóstico no es un fallo: una configuración ausente es el estado normal de una instalación recién hecha y se llama `NON_CONFIGURÉ`, sin confundirse con una avería. En un directorio personal vacío:

```
HOME="$(mktemp -d)" dibab --diagnose; echo "exit code: $?"

OK python : 3.13.5
OK outil_pdftotext : /usr/bin/pdftotext
OK outil_pdftoppm : /usr/bin/pdftoppm
OK outil_findmnt : /usr/bin/findmnt
OK outil_ssh : /usr/bin/ssh
NON_CONFIGURÉ configuration : no configuration found (the --config option: not given; the DIBAB
↪  _CONFIG variable: not given; the XDG location: ~/.config/dibab/config.json; the script
↪  folder: ~/applications/dibab/config.json)
NON_VÉRIFIÉ emplacement_etat : depends on the configuration
NON_VÉRIFIÉ ollama : depends on the configuration
NON_VÉRIFIÉ modele_ocr : depends on the configuration
NON_VÉRIFIÉ modele_tri_rapide : depends on the configuration
NON_VÉRIFIÉ modele_classification : depends on the configuration
NON_VÉRIFIÉ modele_tri_decision : depends on the configuration
NON_VÉRIFIÉ destination_personnel : depends on the configuration
NON_VÉRIFIÉ destination_professionnel : depends on the configuration
Diagnostic: 5 OK, 0 warning(s), 0 failure(s), 1 not configured, 8 not checked
exit code: 0
```

Las herramientas están `OK`, la configuración está `NON_CONFIGURÉ` y las ocho comprobaciones que dependen de ella están `NON_VÉRIFIÉ`. El código de salida es `0`. En cuanto existe una configuración, todo lo que declara y la máquina no tiene es un `ÉCHEC`: un diagnóstico que devolviera `0` con Ollama detenido ocultaría justamente la avería que usted ha venido a buscar.

En otro idioma

El idioma sigue a `LC_ALL`, luego a `LC_MESSAGES`, luego a `LANG`:

```
LC_ALL=fr_FR.UTF-8 dibab --diagnose | tail -n 1

Diagnostic : 12 OK, 1 avertissement(s), 0 échec(s), 1 non configuré(s), 0 non vérifié(s)
```

Soberanía y límites

Lo que la herramienta garantiza

- **Ningún documento sale de su red**. Las llamadas a los modelos de lenguaje son simples peticiones HTTP a `ollama.url` (por defecto `http://localhost:11434`). Ningún documento ni nombre de archivo va a ninguna otra parte, y no hay telemetría.
- **Los únicos servicios externos que el código puede alcanzar son los que usted declara**: sus máquinas por `ssh`, sus buzones IMAP y, si activa OAuth2, los puntos de acceso de autorización de Google.
- **Ningún secreto se guarda en `config.json`**. La contraseña del correo vive en una variable de entorno, un token de Google en un archivo con el modo `600` fuera de cualquier repositorio. Un token nunca se escribe en un diario, ni se muestra en pantalla, ni se pone en un mensaje de error.
- **Una clasificación nunca pierde un documento**. Se copia, la copia se verifica por SHA256 y solo entonces se retira el original. Cada clasificación se escribe en el diario de acciones y puede deshacerse con `--rollback`.
- **Un buzón de correo y una máquina remota quedan intactos**. Los comandos IMAP que se envían son los enumerados en el capítulo 7; una máquina remota solo se lee, y allí no se retira nada.
- **El texto de un documento es un dato, nunca una instrucción**. Se delimita con una marca aleatoria y el modelo responde con un JSON cerrado que Dibab valida. Una respuesta que no se ajusta al esquema cuenta como «incierta» y lleva a la carpeta por verificar. Un modelo nunca puede desencadenar una acción.
- **La duda lleva a una decisión humana**. Un ámbito incierto, modelos en desacuerdo o una pieza común sin categoría unánime colocan el documento en `POR_VERIFICAR/`, no en un lugar arriesgado.
- **El diario contiene rutas y categorías, nunca texto**. Sí contiene nombres de archivos y remitentes, que ya revelan mucho; guárdelo en una bóveda (capítulo 5).

El código se publica bajo la GNU Affero General Public License, versión 3. Usa solo la biblioteca estándar de Python, así que no hay ningún paquete de terceros que auditar.

Respeto del hogar

Una máquina, un directorio o una cuenta de correo solo entra en el ámbito de aplicación de la herramienta con el acuerdo, explícito o implícito, de la persona a la que concierne. El programa no puede comprobarlo: es el administrador de la instancia quien se compromete a ello. No apunte Dibab al buzón ni a los archivos de una persona que ignora que se leen.

Los límites

- **Las fuentes remotas solo se han ensayado contra servidores falsos**. La fuente `ssh` y la fuente IMAP, con contraseña o con OAuth2, están cubiertas por pruebas y por las repeticiones de este manual, que emplean un `ssh` falso, un servidor IMAP falso y un servidor de autorización falso. Su uso en una máquina real del hogar y en un buzón real, incluido el consentimiento con Google, está por hacer: mientras no se haga ese ensayo, considere estas fuentes como no validadas en condiciones reales y empiece por `--dry-run`.
- **La calidad de la clasificación es la de sus modelos**. En la instancia de referencia, un primer ensayo con 15 documentos reales clasificó 2, apartó 3 como no administrativos y puso 10 en `POR_VERIFICAR/`, por prudencia. Mida antes de confiar; la herramienta de medición de las fuentes, `_outils/mesurer_domaines`

`.py`, toma una muestra de documentos clasificados y compara el veredicto de los modelos con la bóveda en la que está cada uno.

- **Solo Linux**. La comprobación del montaje se apoya en `findmnt`, que pertenece a `util-linux`.
- **Dibab no monta, no cifra ni sincroniza nada**, y no gestiona ni el espacio en disco ni las copias de seguridad: eso corre por su cuenta.
- **El alcance de Google es más amplio que el uso**. El alcance `https://mail.google.com/` es el único que Google documenta para IMAP. Permite leer, escribir y borrar. Que el buzón permanezca sin modificar descansa en los comandos que Dibab envía, enumerados en el capítulo 7, no en el alcance.
- **Los valores por defecto son ejemplos**. Las categorías, los remitentes conocidos y la plantilla de nombrado de `config.exemple.json` son genéricos; adáptelos a su país.
- **Los pasos en la consola de Google dependen de la interfaz de Google**, que cambia. Las nociones siguen siendo las de su documentación OAuth2.
- **El proyecto es joven**. La versión es 0.9.0-alpha; no hay integración continua. Las pruebas de las fuentes se ejecutan a mano.
- **Ollama no forma parte de Dibab**. Dibab nunca lo instala ni lo actualiza; un modelo que se renombra o se retira aguas arriba es una línea que corregir en `ollama.roles`.

Informar de un problema

Una vulnerabilidad se comunica en privado, según el archivo `SECURITY.md` de las fuentes, nunca en una incidencia pública. Para una contribución, lea `CONTRIBUTING.md` en el mismo lugar.