Copper Canyon Dam: RCA example data set

> **Disclaimer.** This data set is part of the IT/OT Academy, offered by The IT/OT Insider. Everything in it (the dam, the company, the turbine supplier, the people, the data) is **fictional and generated**, made **only for demo and training purposes**. Do not use it for any real control, quality or maintenance decision.

This folder lets you try an AI-assisted **root cause analysis (RCA)** on a hydro turbine, with no prior experience. You give an AI assistant the folder, paste two prompts, and it works like a co-analyst: it reads the documents and the sensor data, tests possible causes against the evidence, and writes a draft RCA report for you to review.

Background, the full demo and the guide for engineers:

- Project: https://github.com/itotinsider/cromley-hydraulics-uns
- RCA guide for engineers: https://github.com/itotinsider/cromley-hydraulics-uns/wiki/RCA-guide-for-engineers

---

The story in 30 seconds

Copper Canyon Dam has three generating units with turbines from the (fictional) supplier Cromley Hydraulics. On **3 November 2026 at 09:00** the supplier's monitoring PLC on **Unit 1** (serial `CH-FR-1987-0042`) raised a *high-high* alarm on **pad 3 of the turbine guide bearing (TGB)**. The shift unloaded and stopped the unit, following procedure OP-07.

Your job: find out **why**. Was the bearing really overheating, or was it the measurement? What caused it, and what should be done so it does not happen again?

Step 1: Unzip the folder

1. Unzip the file you downloaded somewhere easy to find, for example your Desktop or Documents.
2. Open the unzipped folder and check that you see this:

```
Copper Canyon Dam RCA Example Data Set/
├── README.md        ← this file
├── AGENTS.md        ← instructions for the AI assistant (do not delete)
├── CLAUDE.md        ← points Claude to AGENTS.md (do not delete)
├── library/         ← documents: manuals, procedures, work orders, earlier RCA reports,
...
├── historian/       ← sensor data from the plant (compressed CSV files)
├── reports/         ← empty; the AI writes its RCA report here
└── work/            ← empty; the AI keeps its analysis scripts here
```

Do not rename or move the files inside. You do not need to unpack the `.csv.gz` files; the AI reads them as they are.

Important: always open the folder that *directly* contains `AGENTS.md`. If unzipping created a folder inside a folder, pick the inner one. Otherwise the assistant will not find its instructions.

Step 2: Pick an AI tool

You need an AI tool that can **(a) read the files in a folder on your computer**, **(b) read `AGENTS.md` (or `CLAUDE.md`) as its instructions**, and **(c) run small Python scripts** to analyse the data. Plain chat windows where you can only upload a few files will not work well: the data set is too big and the AI needs to compute on it.

Tool	Works?	Notes
Claude Code (Claude desktop app, *Code* tab, or the `claude` command in a terminal)	Yes, recommended	Reads `CLAUDE.md` / `AGENTS.md` automatically
OpenAI Codex (CLI or IDE extension)	Yes	Reads `AGENTS.md` automatically
GitHub Copilot in VS Code, *Agent* mode	Yes	Open the folder in VS Code; Copilot agent mode reads `AGENTS.md`
ChatGPT / Claude / Copilot web chat (upload only)	Not recommended	No folder access; see the tip below if this is all you have

You also need **Python 3** on your computer (the AI uses it to read the data; no extra packages are needed). On a Mac it is usually already installed. On Windows, install it from https://www.python.org/downloads/ and tick "*Add Python to PATH*" during installation. Not sure? Just start; the AI will tell you if Python is missing.

Using Claude (recommended)

Option A, Claude desktop app (easiest):

1. Install the Claude desktop app from https://claude.ai/download and sign in (a paid plan is needed for Claude Code).
2. Open the **Code** tab and choose this folder as the working folder.
3. Go to Step 3.

Option B, terminal:

1. Install Claude Code: see https://docs.claude.com/en/docs/claude-code/quickstart
2. Open a terminal *in this folder*. (Mac: right-click the folder in Finder → *Services* → *New Terminal at Folder*. Windows: open the folder in Explorer, type `cmd` in the address bar, press Enter.)
3. Type `claude` and press Enter.
4. Go to Step 3.

Using OpenAI Codex

1. Install Codex (see https://developers.openai.com/codex) and sign in with your ChatGPT account.
2. Open a terminal in this folder (see above), type `codex` and press Enter.
3. Go to Step 3.

Using GitHub Copilot (VS Code)

1. Install VS Code and the GitHub Copilot extension, and sign in.
2. *File → Open Folder…* and choose this folder.
3. Open the Copilot Chat panel and switch the mode to **Agent**.
4. Go to Step 3.

Tip if you only have a web chat

Start the chat by pasting the full contents of `AGENTS.md`, then upload as many files from `library/` and `historian/` as the tool allows. Expect a weaker result: the assistant cannot see everything and may not be able to run the data analysis.

Step 3: Run the demo (two prompts)

Copy each prompt below **exactly**, paste it into the AI tool, and press Enter. Wait until the AI has completely finished before you paste the next one.

Prompt 1: get to know the data

```
Read library/README.md and historian/README.md. Tell me what documents and data you have, which unit and period the data covers, and what you think is missing for an RCA.
```

What to expect (1–3 minutes): a short overview of the documents (supplier manuals, procedures, work orders, oil analysis, shift log, earlier RCA reports), the sensor data (three units, about three weeks, with 1-second detail around the event), and a list of things that would help but are not in the folder.

Prompt 2: start the RCA

```
Start the Copper Canyon Dam RCA process for Unit 1 (CH-FR-1987-0042): CH PLC alarm TGB pad 3 high-high at 3 Nov 2026 09:00, unit unloaded and stopped by the shift per OP-07.
```

What to expect (roughly 10–30 minutes, depending on the tool): the AI follows the site's RCA procedure. It lists possible causes from the failure-mode table (FMECA), checks each one against the documents and the sensor data, writes small Python scripts in `work/`, and finally writes a draft report to `reports/RCA-2026-<number>.md`. In the chat it gives a short summary: the problem, the physical / human / latent root causes, the top three actions and open questions for the team.

Step 4: Approve, wait, read

- **The AI will ask permission** to read files, run Python or write the report. That is normal. Allow these actions; everything stays inside this folder. (In Claude Code you can choose "*Yes, and don't ask again*" to reduce the number of questions.)

- **Let it work.** Long pauses are normal while it analyses data.
- **Open the report** in `reports/` when it is done. It is a Markdown (`.md`) file: open it in VS Code, any text editor, or drag it into your browser.
- **Review it like an engineer.** The AI is a co-analyst, not the final judge. Every claim in the report should point to evidence (a work order, a log line, a data file and time window). Check a few of them yourself. Anything without evidence should be listed as open, not as a conclusion.

Good follow-up questions to try

- "*Show me the evidence that the pad 3 temperature was real and not a sensor fault.*"
- "*Compare this event with the earlier RCA reports in library/drwpa/rca-reports.*"
- "*Which hypotheses are still open, and what evidence would close them?*"
- "*Write the interview questions for the shift team.*"

Troubleshooting

Problem	What to do
The AI does not seem to follow the RCA procedure or writes no report	It probably did not read `AGENTS.md`. Check that you opened the folder that directly contains `AGENTS.md`, or start with: "*Read AGENTS.md and follow it.*"
"python3: command not found" or similar	Install Python 3 (see Step 2). On Windows the command may be `python` or `py`; tell the AI that.
The AI says it cannot open `.gz` files	Tell it: "*Use python3 with gzip.open(path, 'rt') as described in AGENTS.md.*"
It stopped halfway	Type "*continue*".
You want to start over	Delete the contents of `reports/` and `work/`, then start a new chat.
Results differ between runs or tools	That is normal for AI tools. The evidence in the files does not change, so compare the citations, not the wording.

About your own data

Everything here is generated, so you can store and use it anywhere. **If you later try this with your own company's documents or data, only use an AI environment your company has approved.**

Licence

See `LICENSE`.