Improvements in Build Reproducibility

Why yes, it is mainly SOURCE_DATE_EPOCH.

What changes when you build the same image twice?

- Datestamps!
  - “org.opencontainers.image.created” annotation in the image manifest (OCI only)
  - “created” field in the image configuration JSON
  - “created” field in history entries in image configuration JSON
  - mtime on files and directories created at build-time in layers
- Randomly-generated things!
  - “container” ID field in the image configuration JSON (Docker only)
    - Truncated and used for the default hostname for RUN instructions
- Not random things!
  - Buildah vanity label

Previous effort: –timestamp

- Force-set creation timestamps to supplied value
- Force-set history timestamps to supplied value
- Force-set timestamps on content in layers to supplied value
- No hint or help for commands run using RUN instructions
- Net result: not what we needed
  - Timestamps recorded *in* files still varied, mismatches broke python byte-compile cache
  - Random hostname would pop up in logs

New features: –source-date-epoch, –rewrite-timestamp

- –source-date-epoch
  - Modeled after reproducible-builds.org's SOURCE_DATE_EPOCH spec
  - Like –timestamp, affects creation and history dates in image metadata
  - Can be set using CLI flag, environment variable, or as a build-arg
  - When set, provides a default value for ARG SOURCE_DATE_EPOCH
    - Declared ARGs are exposed in the environment for RUN instructions
    - Many (but not all) tools we use pick up on the environment variable and do their part
  - When set, RUN instructions get a static hostname
  - When set, the container ID field is cleared in the committed image
- –rewrite-timestamp
  - "Clamps" timestamps on contents of layers to be no later than the –source-date-epoch

Before

https://asciinema.org/a/nyLgfZaVprJpX0txwZyvPZgUj

After

https://asciinema.org/a/5Eg0LHa9HX1ilTpqUs4GQw3Am