Importance of Measuring MT Output

You need a metric to be able to check your progress.
An example from the history:

- Manual judgement at Euratom (Ispra) of a Systran system (Russian→English) in 1972 revealed huge differences in judging; (Blanchon et al., 2004):
    - 1/5 (D–) for output quality (evaluated by teachers of language),
    - 4.5/5 (A+) for usability (evaluated by nuclear physicists).

Metrics can drive the research for the topics they evaluate.

- Some measured improvement required by sponsors: NIST MT Eval, DARPA, TC-STAR, EuroMatrix+.
- BLEU has lead to a focus on phrase-based MT.
- Other metrics may similarly change the community's focus.