

# Cancer data not readily available for future research

July 13 2011

---

A new study finds that -- even in a field with clear standards and online databases -- the rate of public data archiving in cancer research is increasing only slowly. Furthermore, research studies in cancer and human subjects are less likely than other research studies to make their datasets available for reuse.

The results come from a study of patterns of research data availability conducted by Dr Heather Piwowar of the National [Evolutionary Synthesis](#) Center.

Data collected in scientific research is often useful for future studies by other investigators, but scientists have rarely made their raw research data widely available. Tools and initiatives are underway to encourage scientists to publicly archive their data. This analysis confirms there is still much room for improvement.

By querying the full text of the scientific literature through websites like Google Scholar and PubMed Central, Piwowar identified eleven thousand studies that collected a particular type of data about [cellular activity](#), called [gene expression](#) microarray data. Only 45% of recent gene expression studies were found to have deposited their data in the public databases developed for this purpose. The rate of data publication has increased only slightly from 2007 to 2009. Data is shared least often from studies on cancer and human subjects: [cancer studies](#) make their data available for wide reuse half as often as similar studies outside cancer.

"It was disheartening to discover that studies on [cancer](#) and human subjects were least likely to make their data available. These data are surely some of the most valuable for reuse, to confirm, refute, inform and advance bench-to-bedside translational research," Piwowar said.

"We want as much scientific progress as we can get from our tax and charity dollars. This requires increased access to data resources. Data can be shared while maintaining [patient privacy](#)," Piwowar added, noting that patient re-identification is rarely an issue for gene expression microarray studies.

Most likely to share their data in public databases were investigators from Stanford University and those who published in the journal *Physiological Genomics*.

Scientists sometimes email each other to request datasets that aren't available online, but these requests often go unanswered or are denied by the original investigators. Publishing data in online data repositories is considered the best way to share data for future reuse.

Recent policies by the NSF seek to increase the amount of data disseminated from federally-funded research by requiring data management and dissemination plans in all new grant applications.

**More information:** Piwowar, H. (2011). "Who shares? Who doesn't? Factors associated with openly archiving raw research data." *PLoS ONE* 6(7): e18657. doi:10.1371/journal.pone.0018657

Provided by National Evolutionary Synthesis Center

Citation: Cancer data not readily available for future research (2011, July 13) retrieved 24 April

2024 from <https://medicalxpress.com/news/2011-07-cancer-readily-future.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.