

# How to lure Google Inc. to work for your personal ends

Eli Brandt  
[eub@google.com](mailto:eub@google.com)

generate and

"Google's mission is to organize the world's information and make it universally accessible and useful."

You have data.

And ways of gathering data, and ideas for that -- putting a huge effort into getting data is within bounds.

Context: why does Google do Maps?

You, too, can work in the almost-altruistic bubble that's supported by the giant money spigot.

# What you and Google might do

Import data.

Collect new data.

Process data.

Visualize data.

Our biases:

Tools "in the cloud".

Integrate data, rather than yet another item on [labs.google.com](https://labs.google.com).

# What Google likes, #0: okay, yes, more giant money spigots

If your bioinformatics data can make Google billions of dollars, do let me know. (But talk to your lawyer first. Seriously, don't tell us anything in which you have a proprietary interest unless we've done the whole legal dance.)

# What Google likes, #1: lots of users

Google Sky, for example.

I don't know the exact story, but basically there was a contact between astronomers who said "look at all this data", and some Googlers in the Pittsburgh office.

Or Street View.

# What Google likes, #2: do what won't otherwise get done

Get data online that was offline.  
Data that was too big to deal with.  
Data that was "orphaned".  
(But it's got to have value.)

Make data searchable that was opaque.  
Better yet, build tools for you to make data searchable.  
We do have an interest in scientific data sets.

# What to ask for

*Ask big.* It's really a much easier sell.

Google has a ton of computing resources per engineer.

Can Google's work be leveraged by users?

We like: providing APIs, building platforms.

Not so much: writing one data importer per project.

# How Google works

Most decisions are bottom-up.

Get a Googler excited about what you want to have happen.

An engineer can devote 20% time with essentially no oversight.

Google is surprisingly secretive.



# Grab bag of ideas

MapReduce: high-performance parallel processing on clusters.

Open-source implementation: Hadoop.

It might be tough for a Googler to work with you on this.

Are there computations you'd like to do but they're too big?

General-purpose tools for interconverting data.

Visualization jumping off from Trendalyzer?

Data-mining scholarly literature. (What can we get at?)

Sharing datasets

Search for "palimpsest sneakernet", "freeing the dark data"