The democratization of data for the improvement of the failure data management industry
Agenda

- Open Data
- Who is data.world and why do we care about open data
- Why is a shared model good for the industry
- The HDD Reliability Project
- Where will the HDD Reliability Project take us
If the universe of data were suddenly made available, it would unleash the creativity of problem-solvers to combine different data sets – public and private – to develop innovative solutions to innumerable challenges.

Mikael Hagstrom, former Chair of the World Economic Forum’s Global Agenda Council on Data-Driven Development
5 Star Open Data

- ★★★ make your stuff available on the Web (whatever format) under an open license\(^1\)
- ★★★★ make it available as structured data (e.g., Excel instead of image scan of a table)\(^2\)
- ★★★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)\(^3\)
- ★★★★★★ use URIs to denote things, so that people can point at your stuff\(^4\)
- ★★★★★★ link your data to other data to provide context\(^5\)

http://5stardata.info/en/
Who is data.world?

We are building the most meaningful, collaborative, and abundant data resource in the world.
Linked data realizes the vision of evolving the Web into a global data commons, allowing applications to operate on top of an unbounded set of data sources, via standardized access mechanisms.

I expect that Linked Data will enable a significant evolutionary step in leading the web to it's full potential.

Tim Berners-Lee
TED Talk 2009
Shared Model
“We don’t have better algorithms. We just have more data”

- Peter Norvig, Director of research at Google Inc
What is the HDD Reliability Project?

**Partners**

**Backblaze**
Training data set, Open source code

**Prophetstor**
Open sourced collection agent

**Red Hat**
Will integrate collection as an optional add on to Ceph deployment

**CROSS**
Providing development resource for collection agent development

**data.world**
Platform of record

**Project/Program management**

**Cisco**
Will be primary alpha/beta tester

All UCS (data center server) hardware will be added to the collection
Next Steps

- Alpha: simple collection, integration with data.world API
- Beta: deployment/configuration automation && aggregation on d.w platform
- Release: deployable agent with dynamic aggregation
Future

Post-release:

- Advanced modeling / machine learning
- Automated response to submission with % chance failure
For Questions or to learn how to get involved

Contact
Patrick McGarry, Head of Community at data.world
patrick.mcgarry@data.world
data.world/scuttlemonkey
Thank you 🦉

selene@data.world
data.world/sya