



Stop Paving the Cowpath

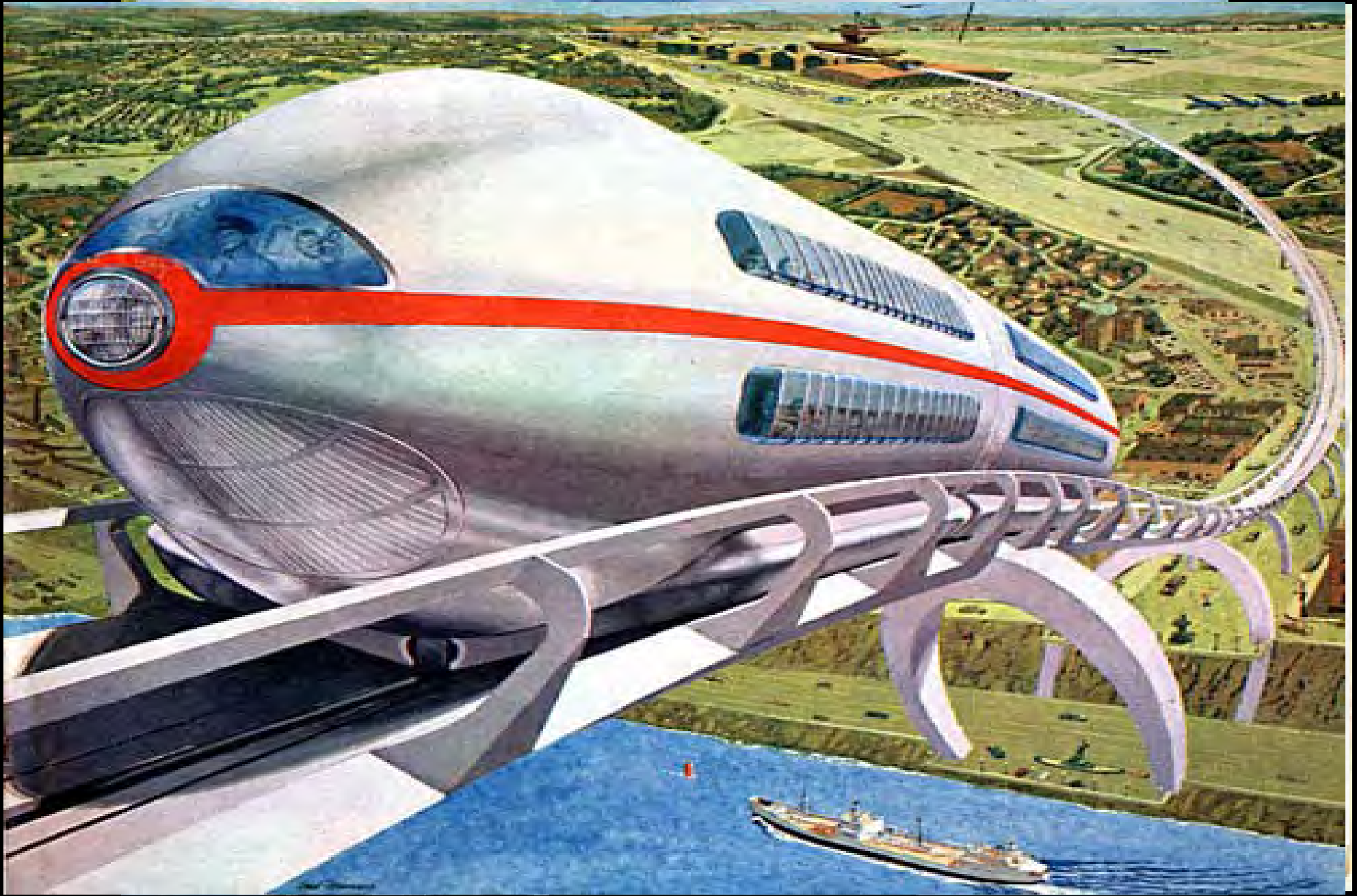
Avoid Building BI for Yesterday's Problems

TDWI Chapter Meeting
September, 2010

Mark R. Madsen
<http://ThirdNature.net>



BI: Building Yesterday's Future Today

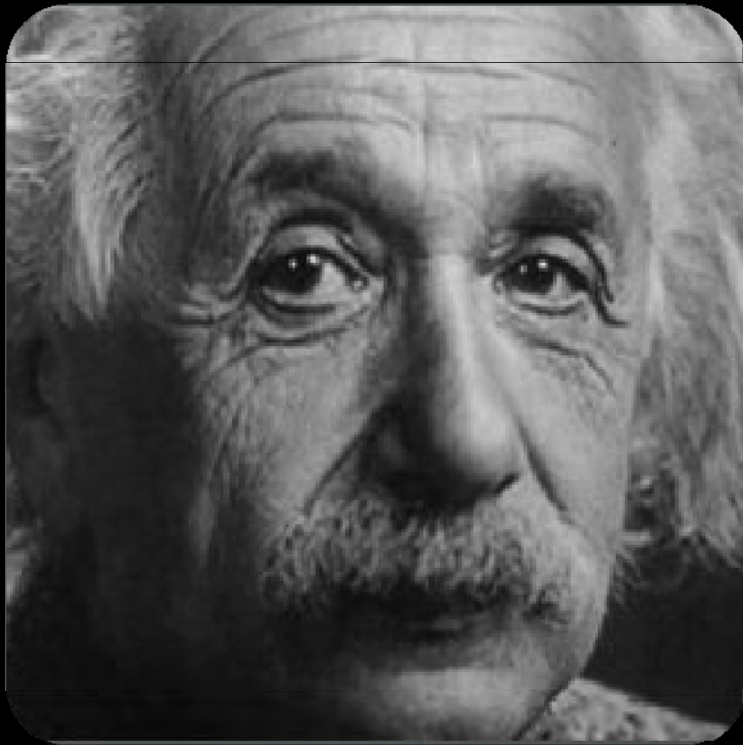




Sometimes the direction you should be moving is not the obvious direction everyone else is excited about.

“We can't solve problems by using the same kind of thinking we used when we created them.”

Albert Einstein



How did we get where we are today?

Information Management v1.0: Clay Tech



The first information explosion, the first metadata

Clay Tech has some familiar limitations



Information Management v2.0 Paper Tech



Lighter, denser, faster storage media

More information = need for new metadata techniques: content tagging, author catalogs



Discovery of one tradeoff between clay and paper...



Paper Tech v2.1: increased storage density, smaller form factor, random access, high res RGB graphics

Loquente
u petro
cedit sps sci
sup omi q au
diebant ubi
magnificabat
deum. et misit
petrus bapu
tari eos i noie
dni ihu xpi.



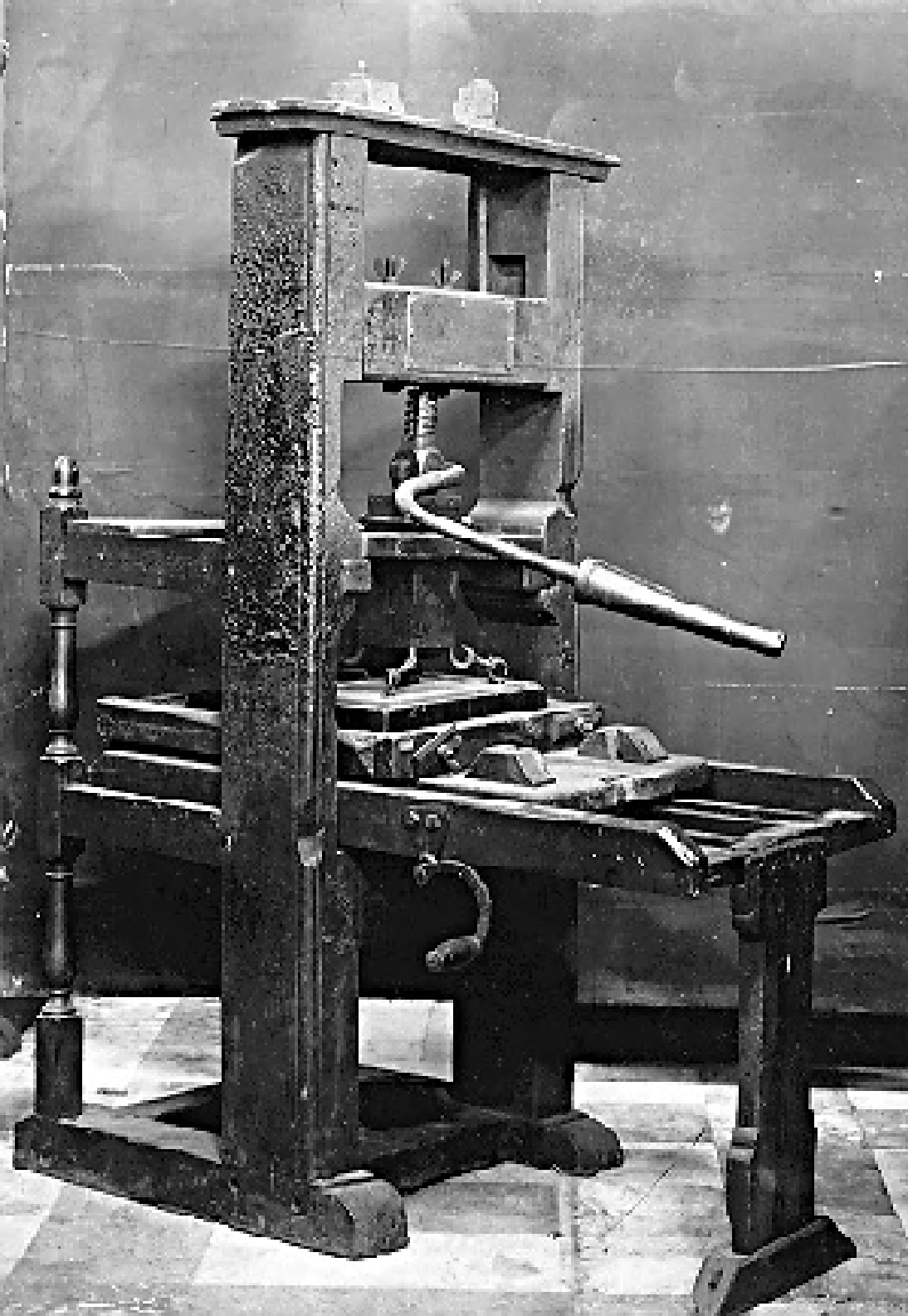
Quem tu
apphen
ditet
misit i cau
terem. uoles
produce eu
poplo. rora
tio fiebat fi
ne m m m m m
one ab ecclia ad
dm pro eo.

Hoc fige
qd qn x
aliqui

Hoc fige
qd qn x
amul m
truduit in au

Better embedded metadata: title page, colophon, ToC





Paper Tech v2.2

Huge information explosion: 8M books in 1500, 200M by 1600

More data management innovations:

- Perfect copies
- Indices
- Topical catalogs
- First real encyclopedia
- Font standardization

Taxonomy ascends

Linnaeus



Top down orientation

Static structure

Descriptive rather than explanatory

Much like a schema...

Buffon



Bottom up orientation

Flexible structure

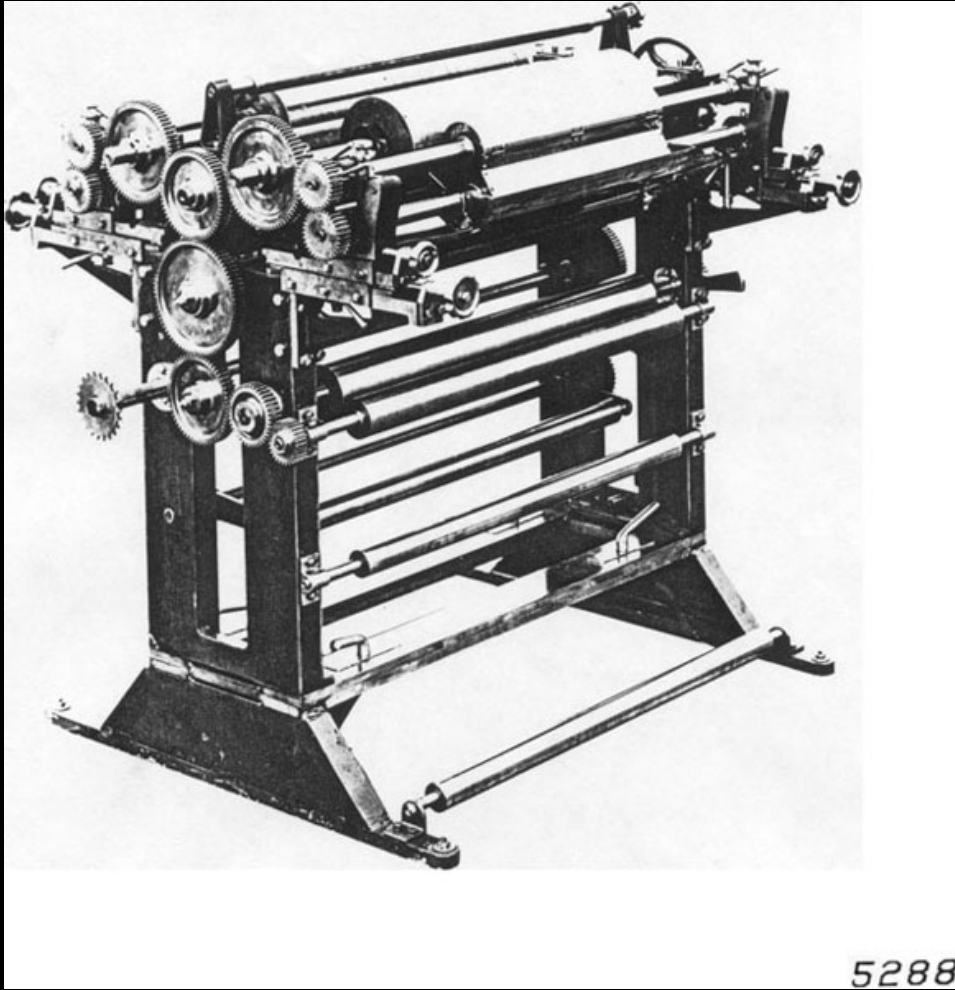
Explanatory, descriptive

Much like faceted search

Paper Tech v2.3

Second printing information explosion:

- Universal classifications
- Card catalogs, cross-referencing, random access metadata
- Trading effort and flexibility for storage and retrieval
- Dewey vs. Cutter



Core Principle Throughout Human History:

New technology development

generates

New information scale and availability

generates

New methods to cope

rinse, repeat

“Everything made now is either a replica or a variant of something made a little time ago and so on back without break to the first morning of human time.”

George Kubler, *The Shape of Time: Remarks on the History of Things*, 1962

Nothing We Do is Original: the Mundaneum



Paul Otlet at his desk

19th Century ETL



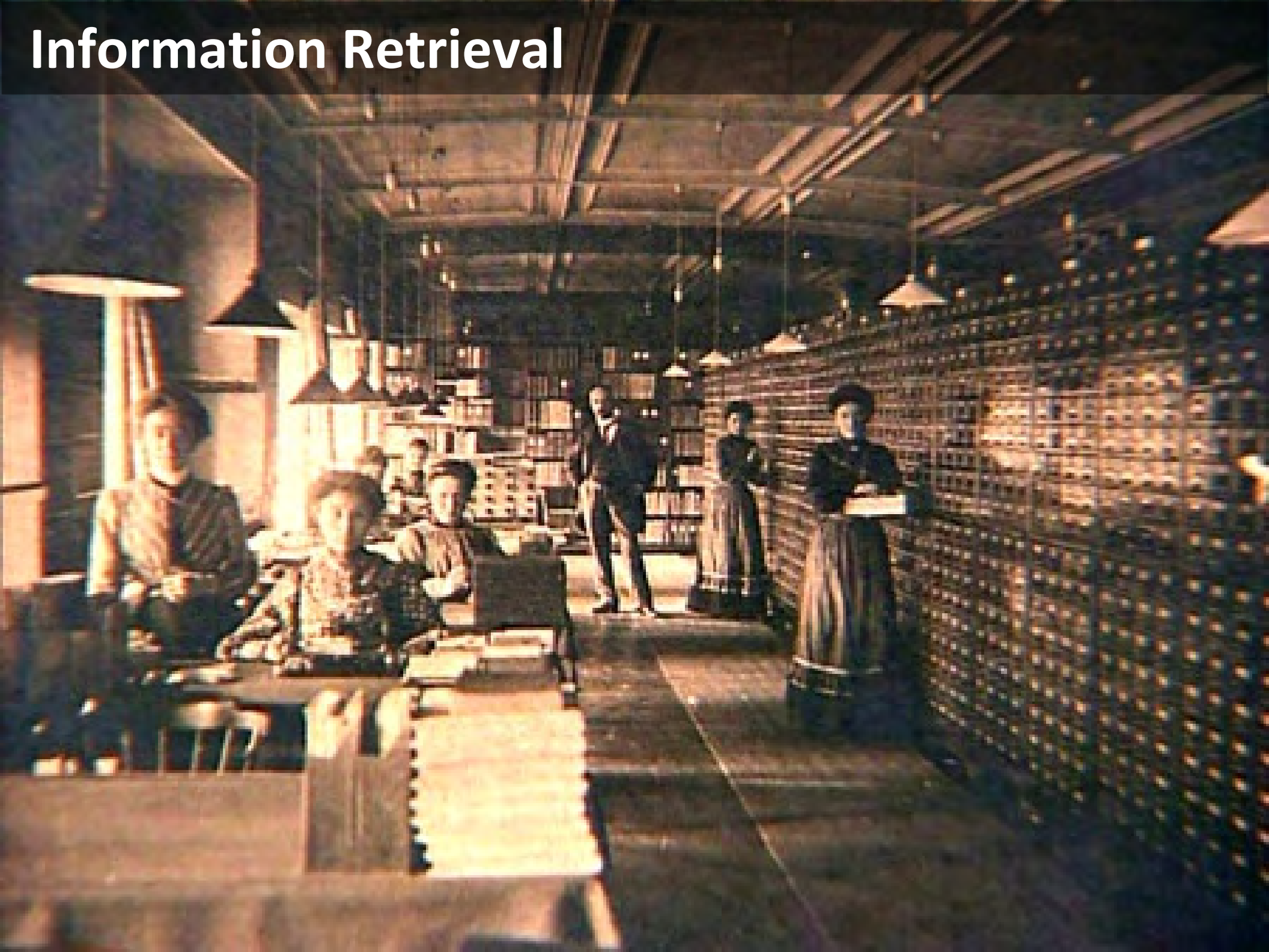
Writing to the Database, Note Multi-processing



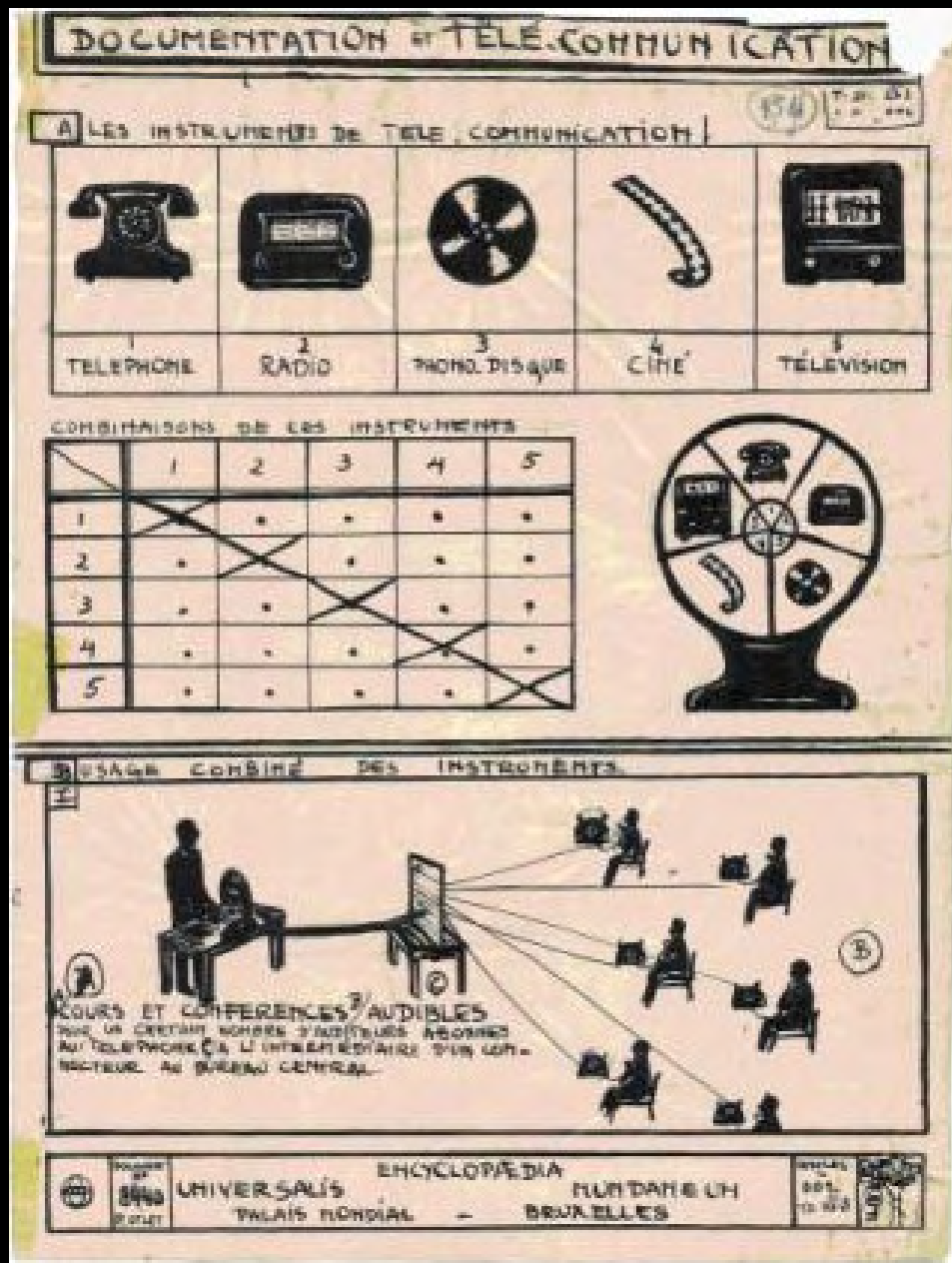
Large Scale Information Storage



Information Retrieval



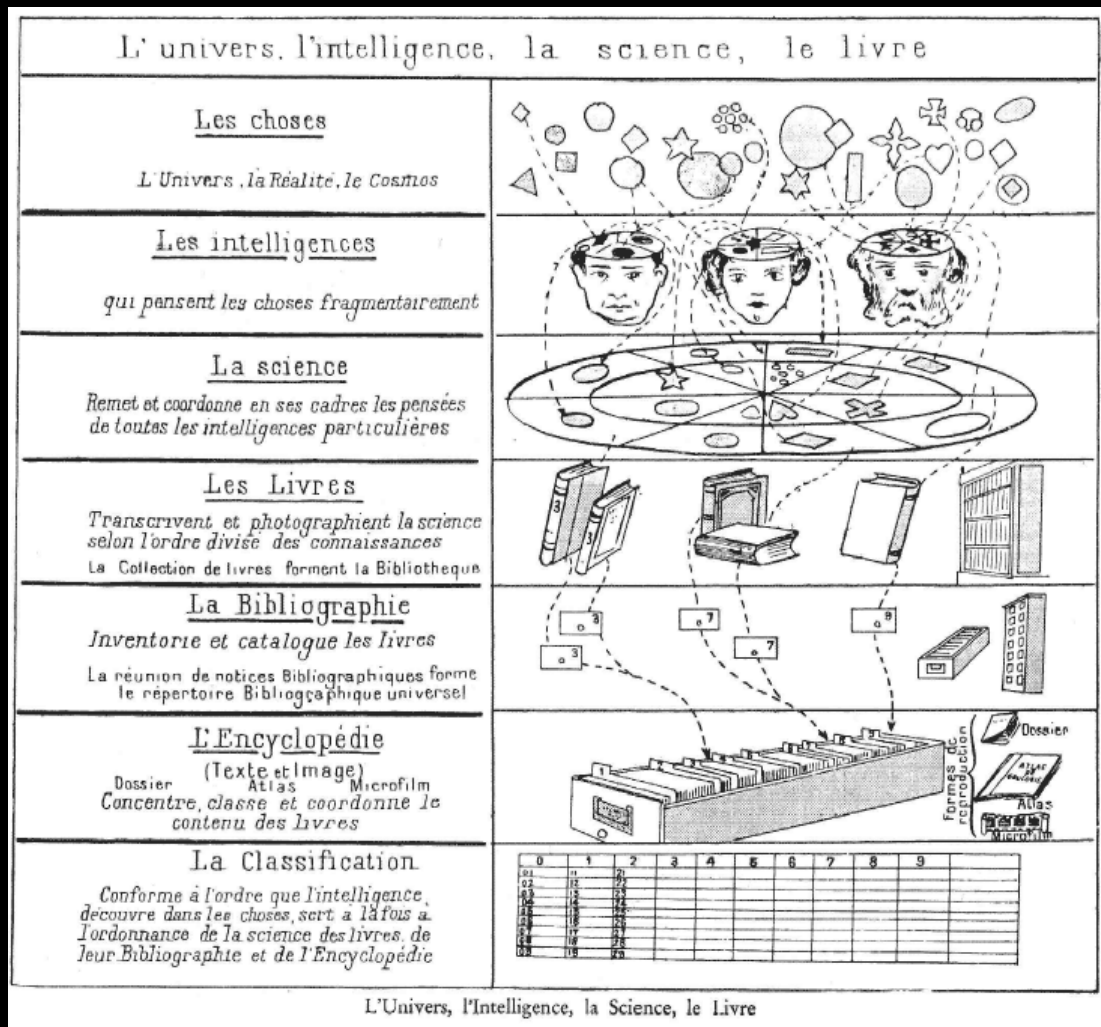
The Personal Computer Was Invented in 1934



Future vision:

- Technological developments will improve the ability to manage information
- Current technologies can be integrated to provide individual discovery, access and collaboration

The Mundaneum Worked, For a While



Two primary flaws of the Mundaneum:

- Static, top-down classification system
- Loading could not keep up with data production rates

Sounds familiar

Every technology is a tradeoff between something

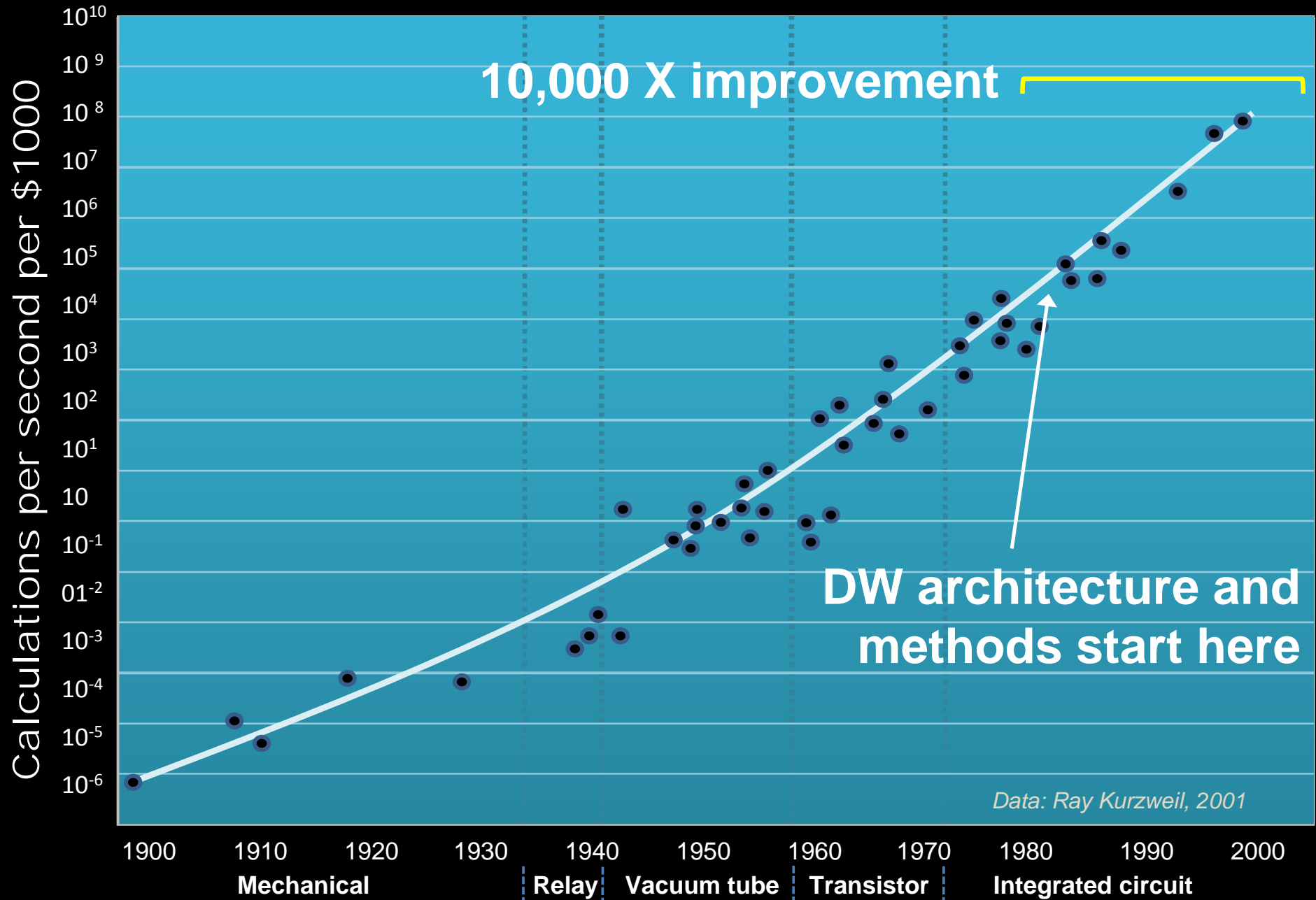
History is always the same:

- Top down vs. bottom up
- Hierarchy vs. network
- Disempowerment vs. autonomy
- Simplicity vs. flexibility



Top down authoritarian hierarchy usually wins because it's more pragmatic, but suffers from accumulated change leading to a revolution.

Technology Has Changed (a lot) But We Haven't





This is where technology exuberance gets you



Resources are scarce.

Therefore centralize: that solves all problems!



Creates bottlenecks

Causes scale problems

Enforces a single model

In some organizations and areas of business “data warehouse” is a bad word.

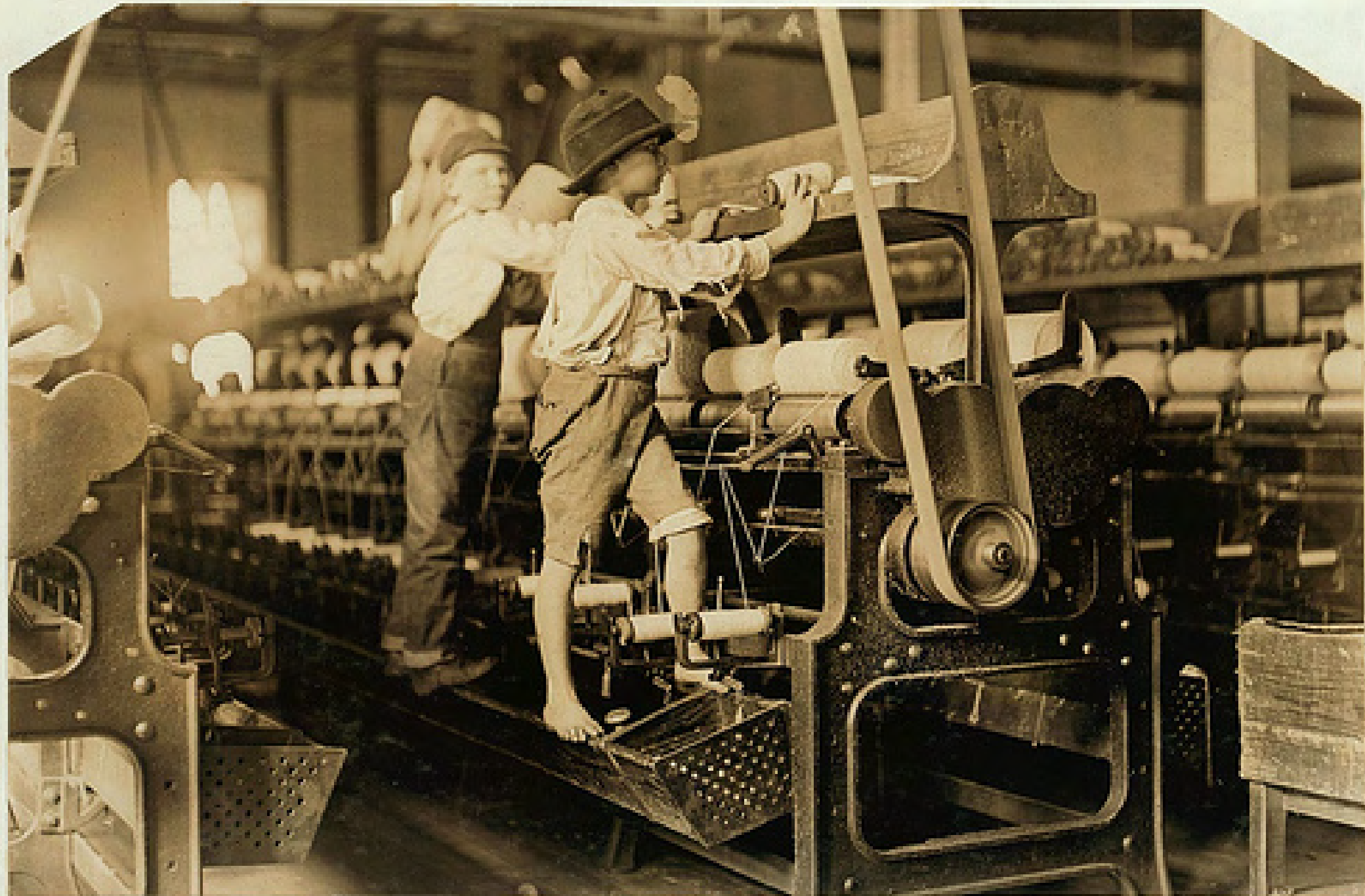
BI tools are a commodity



Commodity Shift is Enabling the Self-serve IT Bypass



...Right at a Time When IT is Challenged Meeting Needs



Users Are Uniform





Therefore standardize: it's simpler for everyone

How We Think of Users



Our design point is the passive consumer of information.

Proof: methodology

- IT role is requirements, design, build, deploy, administer
- User role is run reports

Self-serve BI is not like picking the right doughnut from a box.

How We Think of Users



How We Want Users to Think of Us



How We Think of Users



What Users Really Think



User Perspectives

They need to do things on their own schedule, not according to an IT plan.

The nature of most knowledge work is reacting to and analyzing new problems. For IT it's repeatability.

New problems and projects usually require new data.

Slow IT response = lost opportunity, frustration, increased shadow IT.





BI is not the end of the line



We think of BI as publishing, which it isn't

Excel is not the Problem, BI is the Problem

Because we see BI as the endpoint, we see Excel as a viewer. It isn't. This is why export-to-excel doesn't help.

For an *individual*, Excel is often the easiest path: add missing data, derive data, share results, usability.

People don't want to learn more tools unless they are directly related to their jobs.



Excel is also nearly universal in reach and familiarity.

Models are stable.

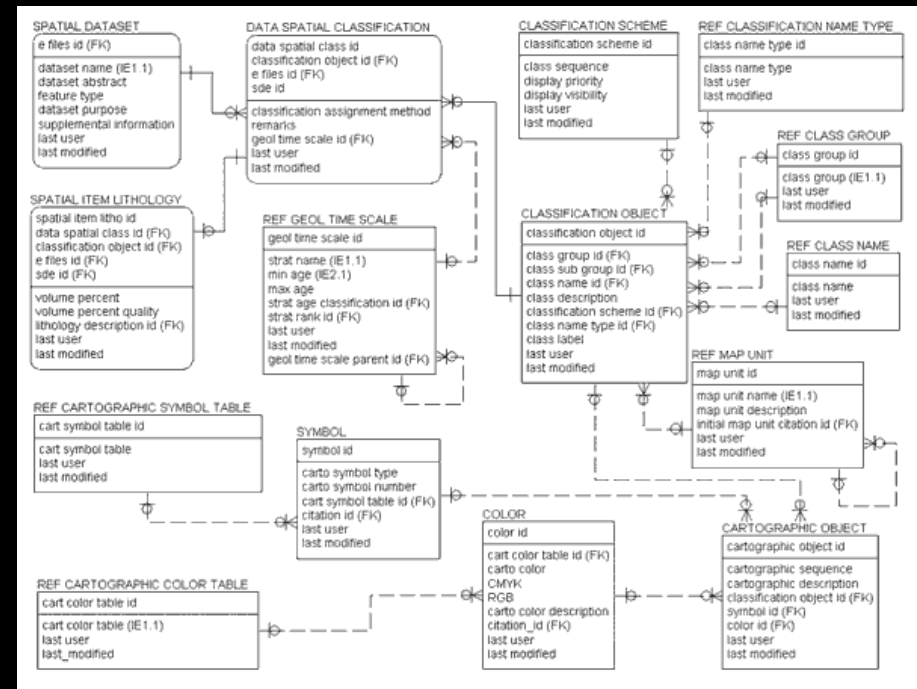
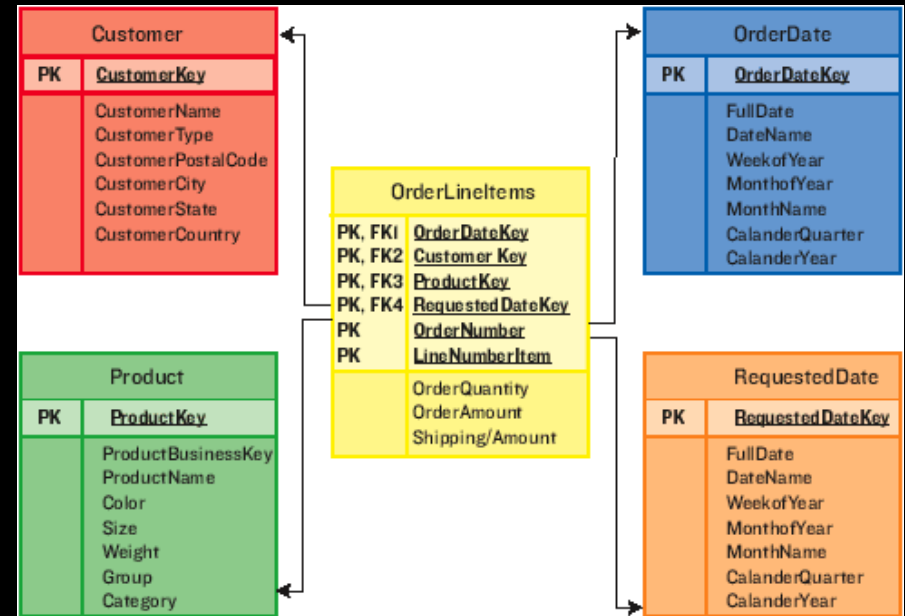


Concepts About Fixed Models Need to Change

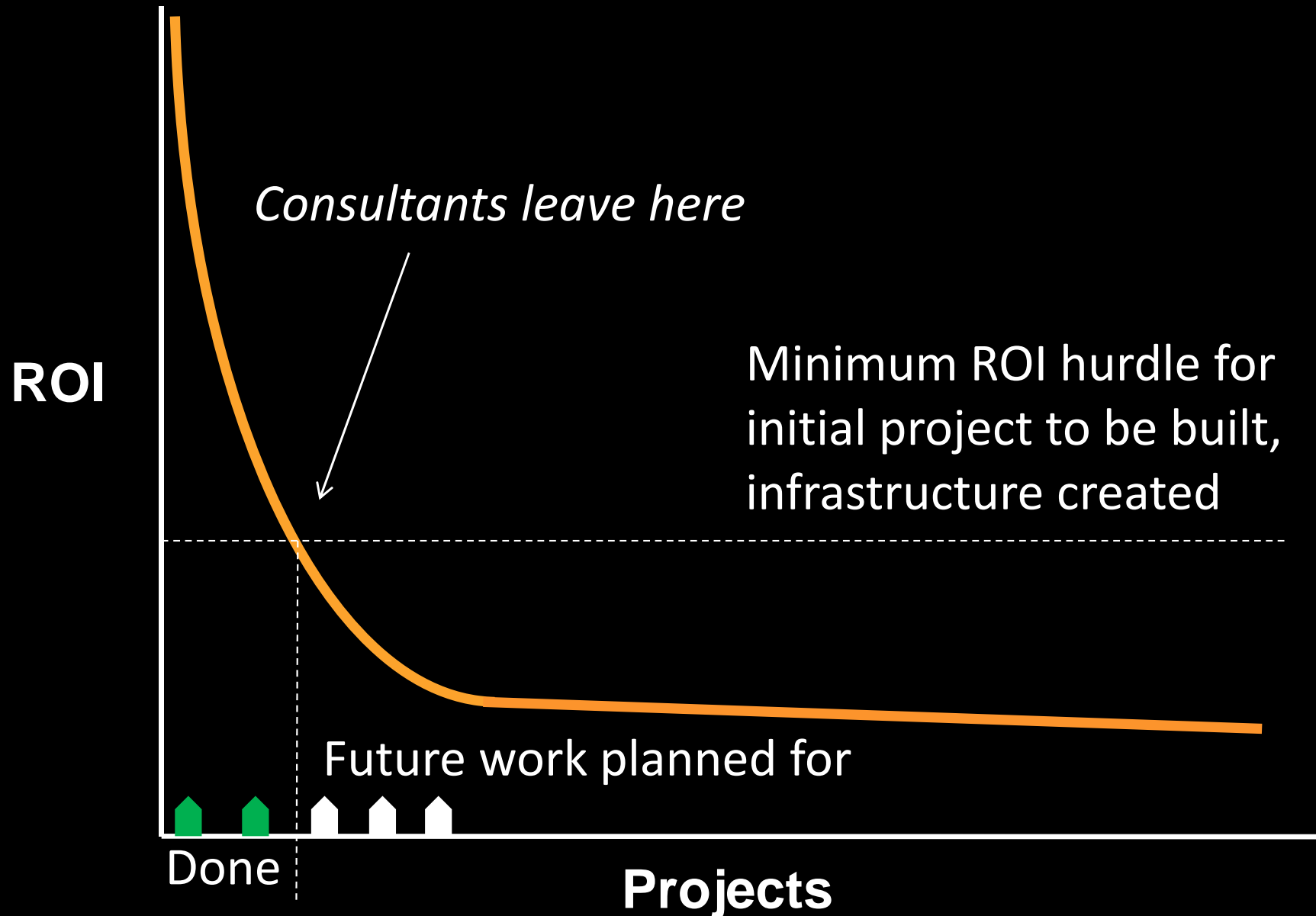
Current techniques for modeling and managing data are too rigid and incapable of describing all the possible relationships, like using a card catalog.

We assume that the data model, the integration, and the use can be isolated from one another.

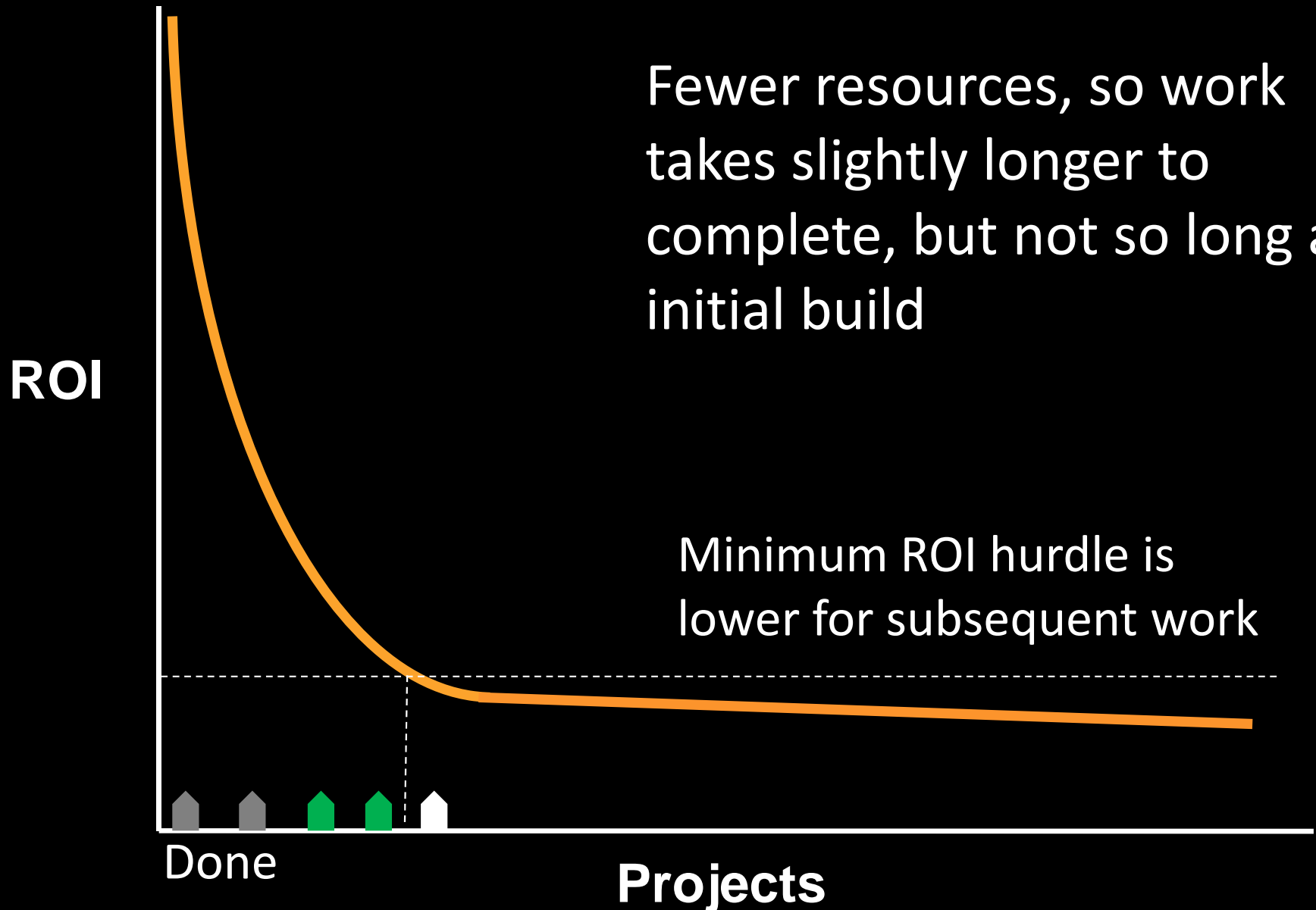
Forget about the model, worry about the keys.



The Origin of BI Backlog: Initial Build

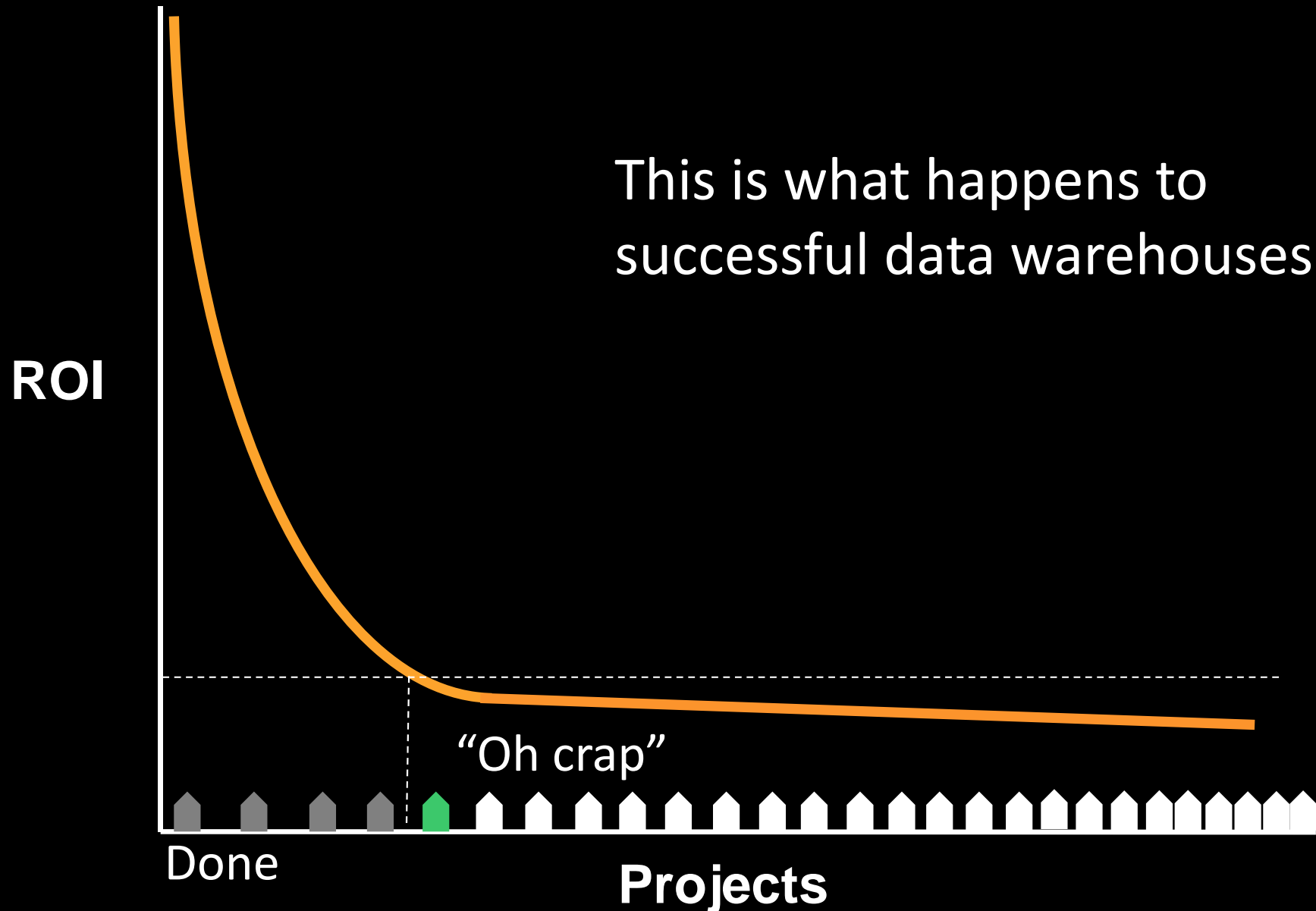


The Origin of BI Backlog: Next Phase

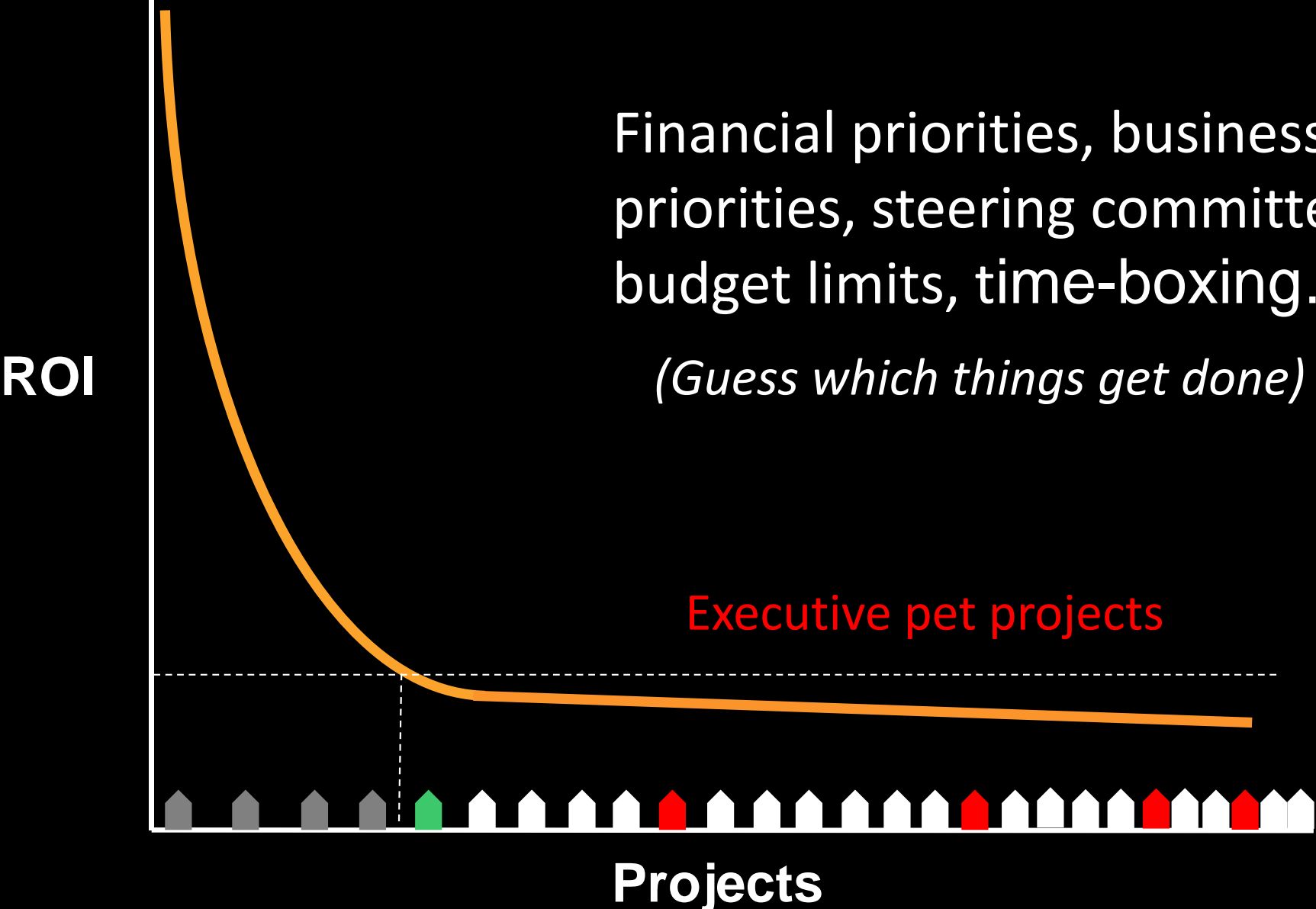


The Long Tail of BI

This is what happens to successful data warehouses

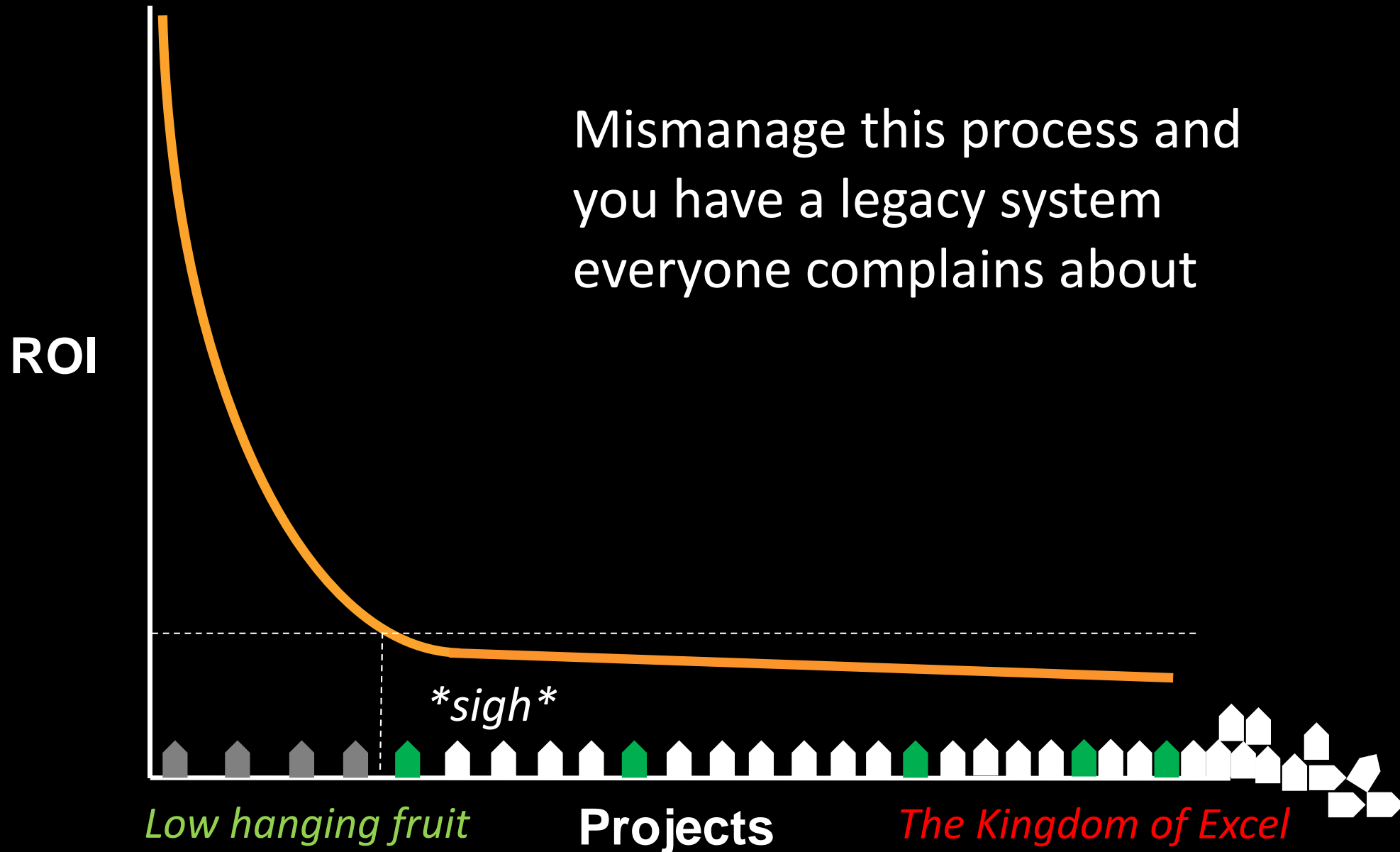


Prioritizing the Long Tail of BI

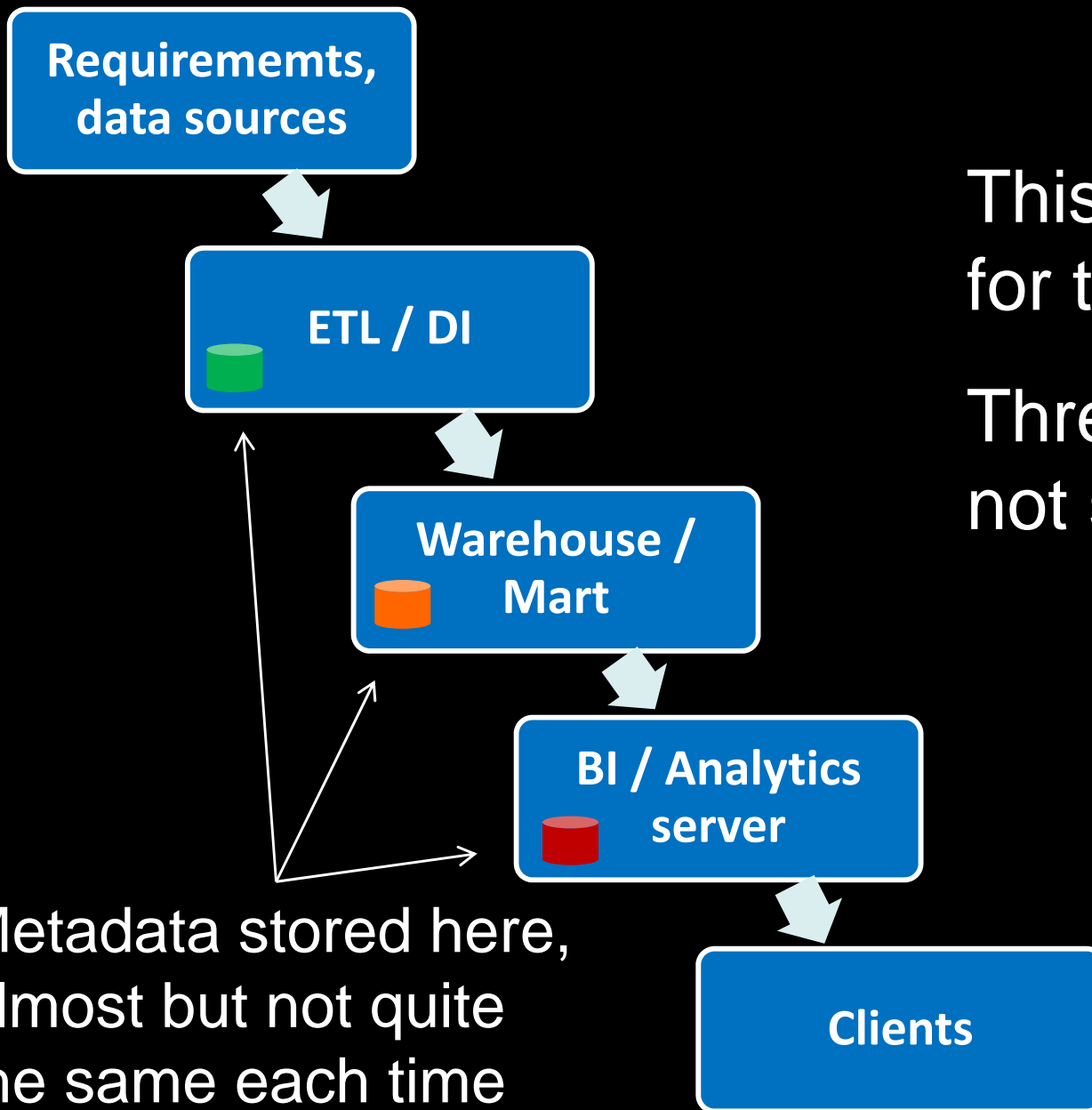


The Long Tail of BI: Why We Have Spreadmarts

Mismanage this process and you have a legacy system everyone complains about



Development Process Designed to Minimize Later Change



This process is fine for the initial build.

Three months later, not so much.

 *The common MD repo was supposed to fix this (#fail)*

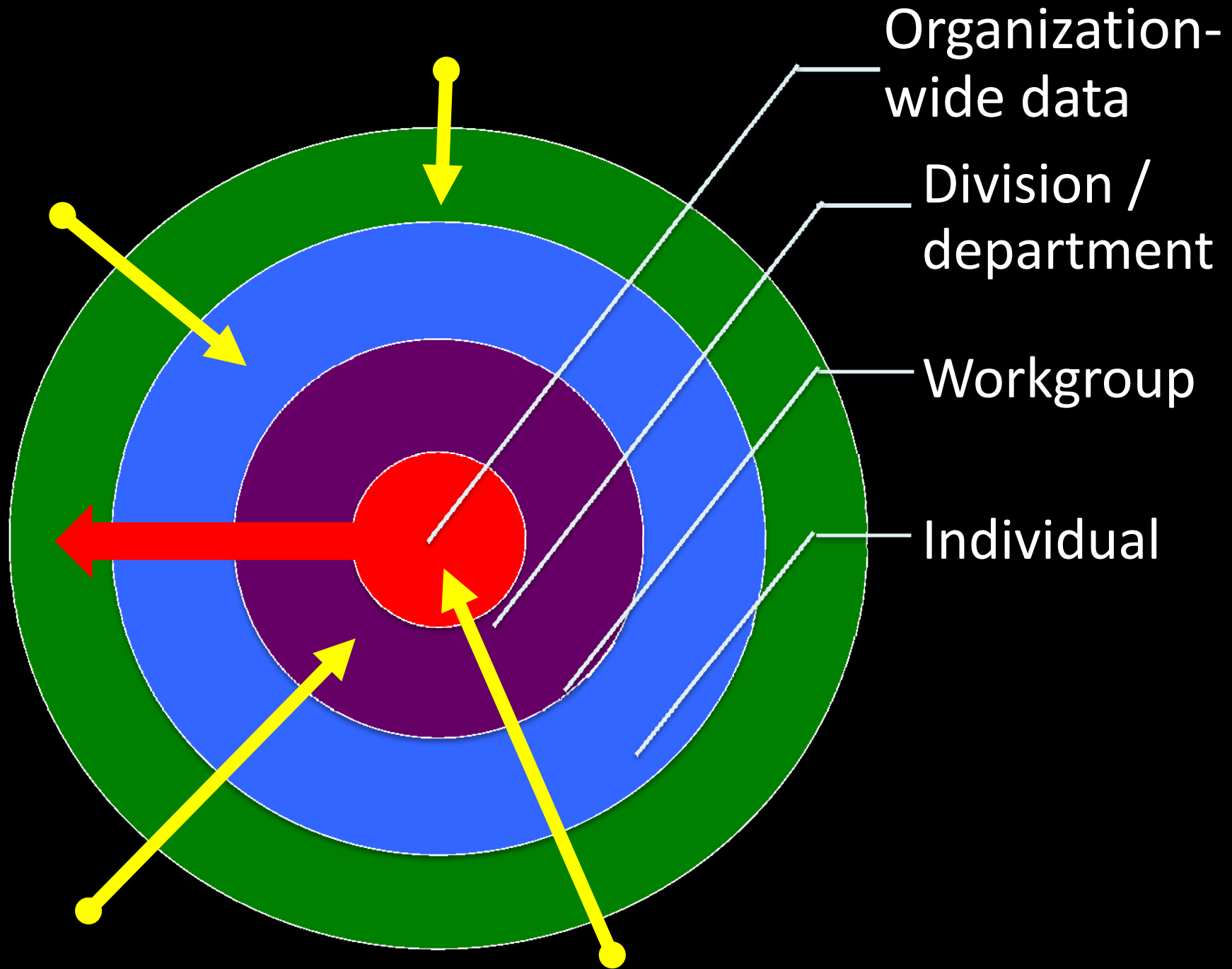


Control

BI is still using the old assumptions of hierarchy and centralized control rooted in organizing around the economics of scarcity.

People don't want buckets, they want fixtures.

We Need to Establish Zones of Control



Methods and architecture have to change along with tools and platforms



Don't Pave the Cowpath





CC Image Attributions

Thanks to the people who supplied the creative commons licensed images used in this presentation:

cuneiform_undo.jpg - <http://www.flickr.com/photos/charlestilford/2552654321/>

cuneiform_proto_3000bc.jpg - <http://www.flickr.com/photos/takomabibelot/3124619443/>

scroll_kerouac.jpg - <http://www.flickr.com/photos/ari/93966538/>

firemen not noticing fire.jpg - <http://flickr.com/photos/oldonliner/1485881035/>

outdated gumshoe.jpg - <http://flickr.com/photos/olivander/372385317/>

manuscript_page.jpg - <http://www.flickr.com/photos/calliope/306564541/>

manuscript_illum.jpg - http://www.flickr.com/photos/diorama_sky/2975796332/

well town hall.jpg - <http://flickr.com/photos/tuinkabouter/1135560976/>

pyramid_camel_rider.jpg - <http://www.flickr.com/photos/khalid-almasoud/1528054134/>

uniform_umbrellas.jpg - <http://www.flickr.com/photos/mortimer/221051561/>

open air market - <http://flickr.com/photos/baboon/309793875/>

train_to_sea.jpg - <http://www.flickr.com/photos/innoxius/457069767/>

wheat_field.jpg - <http://www.flickr.com/photos/ecstaticist/1120119742/>

Open air market - <http://flickr.com/photos/baboon/309793875/>

changing of the guard.jpg - <http://flickr.com/photos/mambo1935/160739264/>

Gare do Oriente Lisbon airport bridge.jpg - <http://flickr.com/photos/higaara/228673603/>

winding_road.jpg - http://www.flickr.com/photos/batt_57/4000701633/

Tokyo forum - <http://flickr.com/photos/fukagawa/2004106475/>

riot police line small - <http://flickr.com/photos/73594239@N00/25719098/>

About the Presenter



Mark Madsen is president of Third Nature, a technology research and consulting firm focused on business intelligence, analytics and information management. Mark is an award-winning author, architect and former CTO whose work has been featured in numerous industry publications. During his career Mark received awards from the American Productivity & Quality Center, TDWI, Computerworld and the Smithsonian Institute. He is an international speaker, contributing editor at Intelligent Enterprise, and manages the open source channel at the Business Intelligence Network. For more information or to contact Mark, visit <http://ThirdNature.net>.

About Third Nature



Third Nature is a research and consulting firm focused on new and emerging technology and practices in business intelligence, data integration and information management. If your question is related to BI, open source, web 2.0 or data integration then you're at the right place.

Our goal is to help companies take advantage of information-driven management practices and applications. We offer education, consulting and research services to support business and IT organizations as well as technology vendors.

We fill the gap between what the industry analyst firms cover and what IT needs. We specialize in product and technology analysis, so we look at emerging technologies and markets, evaluating the products rather than vendor market positions.