Strategic Planning for
IT Support of Grant-funded Research II

(http://www.esp.org/rjr/briite-04.pdf)

Robert J. Robbins

rrobbins@fhcrc.org

(206) 667 4778

Biomedical Research Institutions Information Technology Exchange

Strategic Planning for
IT Support of Grant-funded Research
II

Where are we now and where should we be going?

(http://www.esp.org/rjr/briite-04.pdf)

Robert J. Robbins

rrobbins@fhcrc.org

(206) 667 4778

Strategic Planning for
IT Support of Grant-funded Research
II

Where are we now and where should we be going?

How do we get there & what might get in our way?

(http://www.esp.org/rjr/briite-04.pdf)

Robert J. Robbins

rrobbins@fhcrc.org

(206) 667 4778

Biomedical Research Institutions Information Technology Exchange

Strategic Planning for
IT Support of Grant-funded Research
II

Where are we now and where should we be going?

How do we get there & what might get in our way?

(http://www.esp.org/rjr/briite-04.pdf)
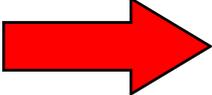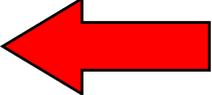
Robert J. Robbins

rrobbins@fhcrc.org

(206) 667 4778

Topics

- STRATEGIC PLANNING

Topics

- STRATEGIC PLANNING

- GOAL: Where should we be going?

Topics

- STRATEGIC PLANNING

- GOAL: Where should we be going?

- START: Where are we now?

Topics

- STRATEGIC PLANNING

- GOAL: Where should we be going?

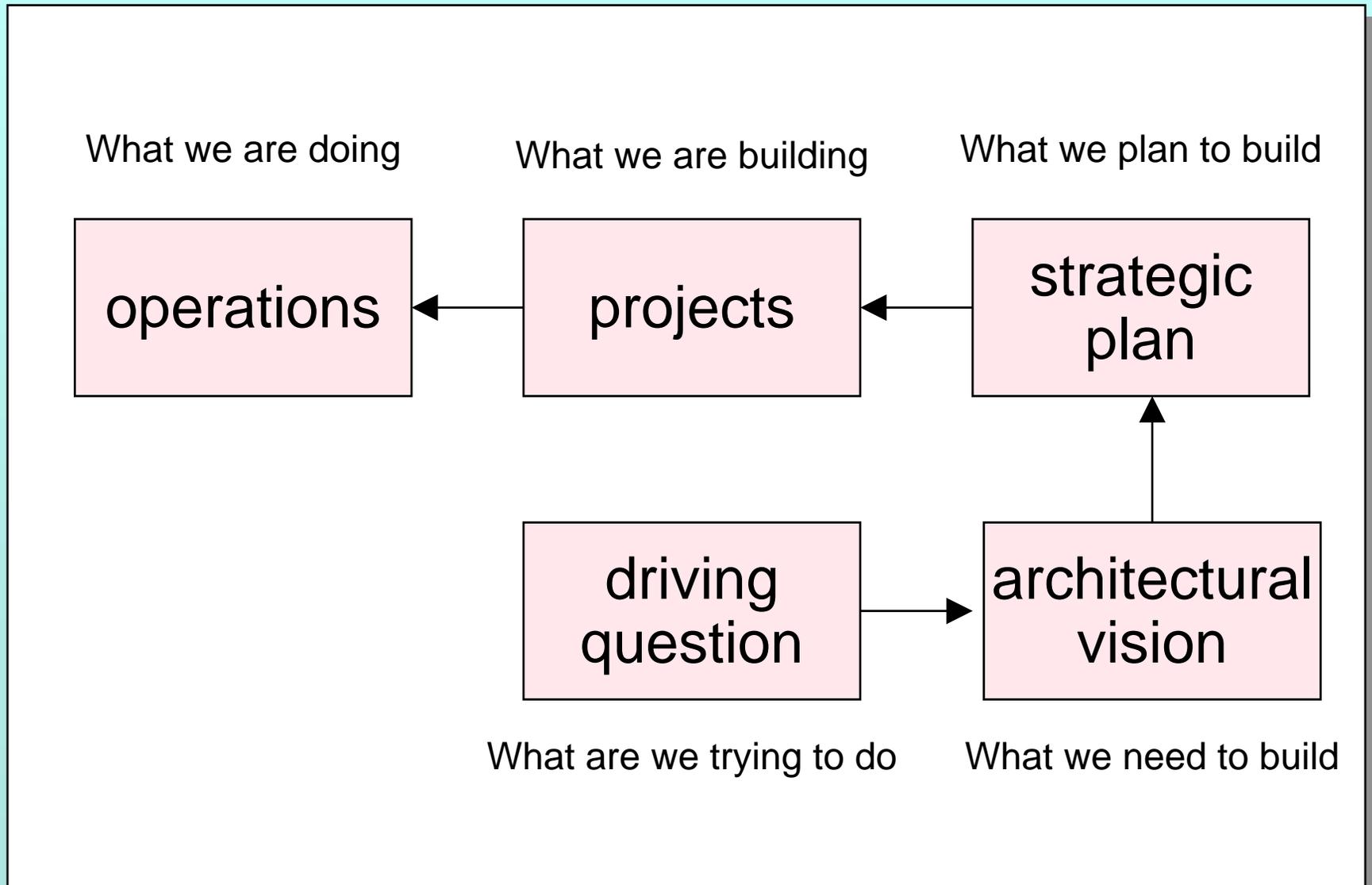- START: Where are we now?

- PATH: How do we get there from here?

Topics

- STRATEGIC PLANNING

- GOAL: Where should we be going?

- START: Where are we now?

- PATH: How do we get there from here?

- **SUMMARY: The role of BRIITE (& others)**

Strategic Planning

Fourth-Box Thinking

What we are doing   What we are building   What we plan to build

operations ← projects ← strategic plan

driving question → architectural vision

architectural vision ↑ strategic plan

What are we trying to do   What we need to build

Fourth-Box Thinking

Example of important driving question:

Q: What can be done to maximize the effectiveness of IT at the level of individual grants?

Fourth-Box Thinking

Example of important driving question:

Q: What can be done to maximize the effectiveness of IT at the level of individual grants?

A: That's a critical question. A case can be made for centralized support of distributed computing.

GOAL

Where should we be going?

Where are we going?

Goal:

Make it possible for nearly all biomedical researchers to use advanced information technology in support of their research at a cost barely more than the marginal costs necessitated by the specifics of their research.

Where are we going?

Goal:

Make it possible for nearly all biomedical researchers to use advanced information technology in support of their research at a cost barely more than the marginal costs necessitated by the specifics of their research…

That is, make it all work like a word processor…

Where are we going?

Goal:

NOTA BENE:

Use at marginal cost is really nothing more and nothing less than what is expected of any mature infrastructure.

That is, make it all work like a word processor…

Where are we going?

We are trying to develop generic systems to provide a robust IT infrastructure that allows use at marginal cost for all grant-funded biomedical researchers.

The solution must interoperate across multiple investigators, multiple research programs, multiple institutions, and multiple communities.

The interoperation should extend to reasonable interactions with biochemistry and on to chemistry and with biophysics and on to physics and with …

Where are we going?

We are trying to develop generic systems to provide a robust IT infrastructure that allows use at marginal cost for all grant-funded biomedical researchers.

The solution must interoperate across multiple investigators, multiple research programs, multiple institutions, and multiple communities.

The interoperation should extend to reasonable interactions with biochemistry and on to chemistry and with b... and with …

TALL ORDER…

START

Where are we now?

Topics

- We recognize that the problem exists.

- Commitment to address the problem exists at many levels.

- Recent technological advances make a solution seem possible.

- And, we are enthusiastic…

- But…

Biomedical Research Institutions Information Technology Exchange

Topics

- But, we must remember that

Topics

- But, we must remember that

  - wherever we are, we're late.

Biomedical Research Institutions Information Technology Exchange

Topics

- But, we must remember that

  - wherever we are, we're late.

  - whatever we've got, it isn't enough.

Topics

- But, we must remember that

  - wherever we are, we're late.

  - whatever we've got, it isn't enough.

  - whatever we do, it will be wrong.

Biomedical Research Institutions Information Technology Exchange

Topics

- But, we must remember that

  - wherever we are, we're late.

  - whatever we've got, it isn't enough.

  - whatever we do, it will be wrong.

  - standards are important, both to have and to avoid.

Biomedical Research Institutions Information Technology Exchange

Topics

- But, we must remember that

  - wherever we are, we're late.

  - whatever we've got, it isn't enough.

  - whatever we do, it will be wrong.

  - standards are important, both to have and to avoid.

  - we are at risk of "light's better" approaches.

Topics

- But, we must remember that

  - wherever we are, we're late.

  - whatever we've got, it isn't enough.

  - whatever we do, it will be wrong.

  - standards are important, both to have and to avoid.

  - we are at risk of "light's better" approaches.

  - there is no silver bullet (no one has all the answers).

START

Wherever we are,

we're late…

Topics

- To stay competitive, biomedical research institutions must deploy IT effectively.

- Moore's Law constantly transforms IT (and everything else).

- Moore's Law divides all IT project time into three periods:

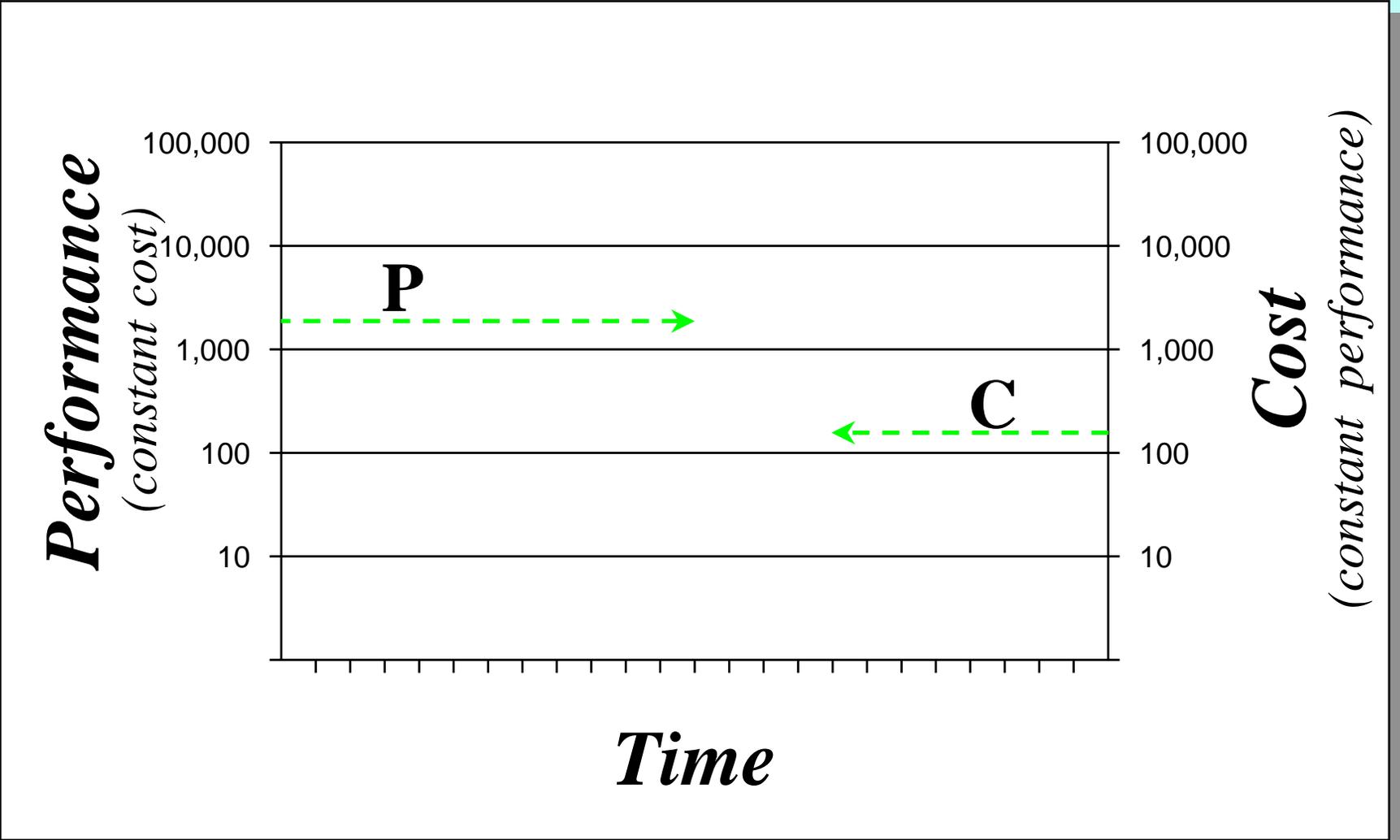 - Too hard

  - Too expensive

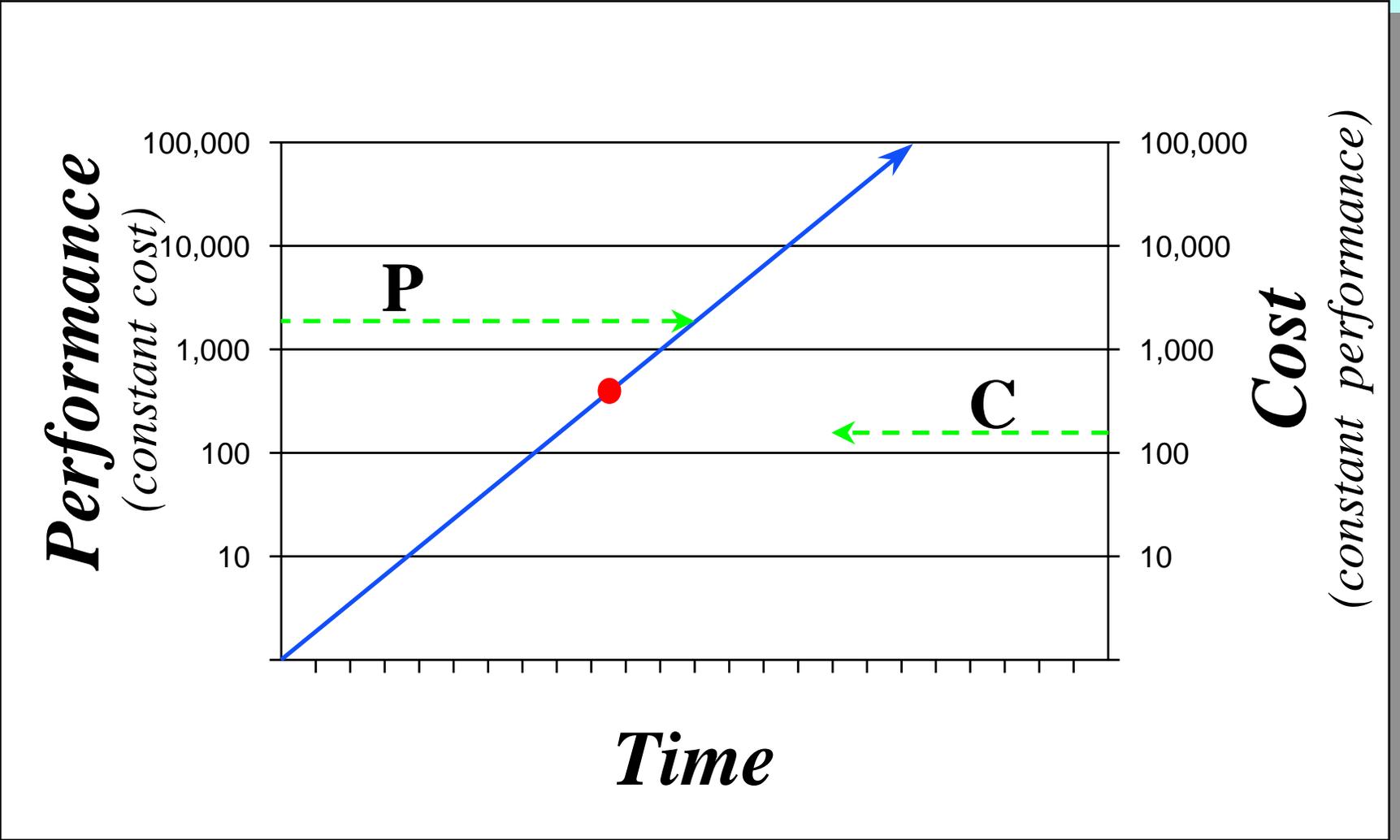 - Too late

Moore's Law: *The Effect*
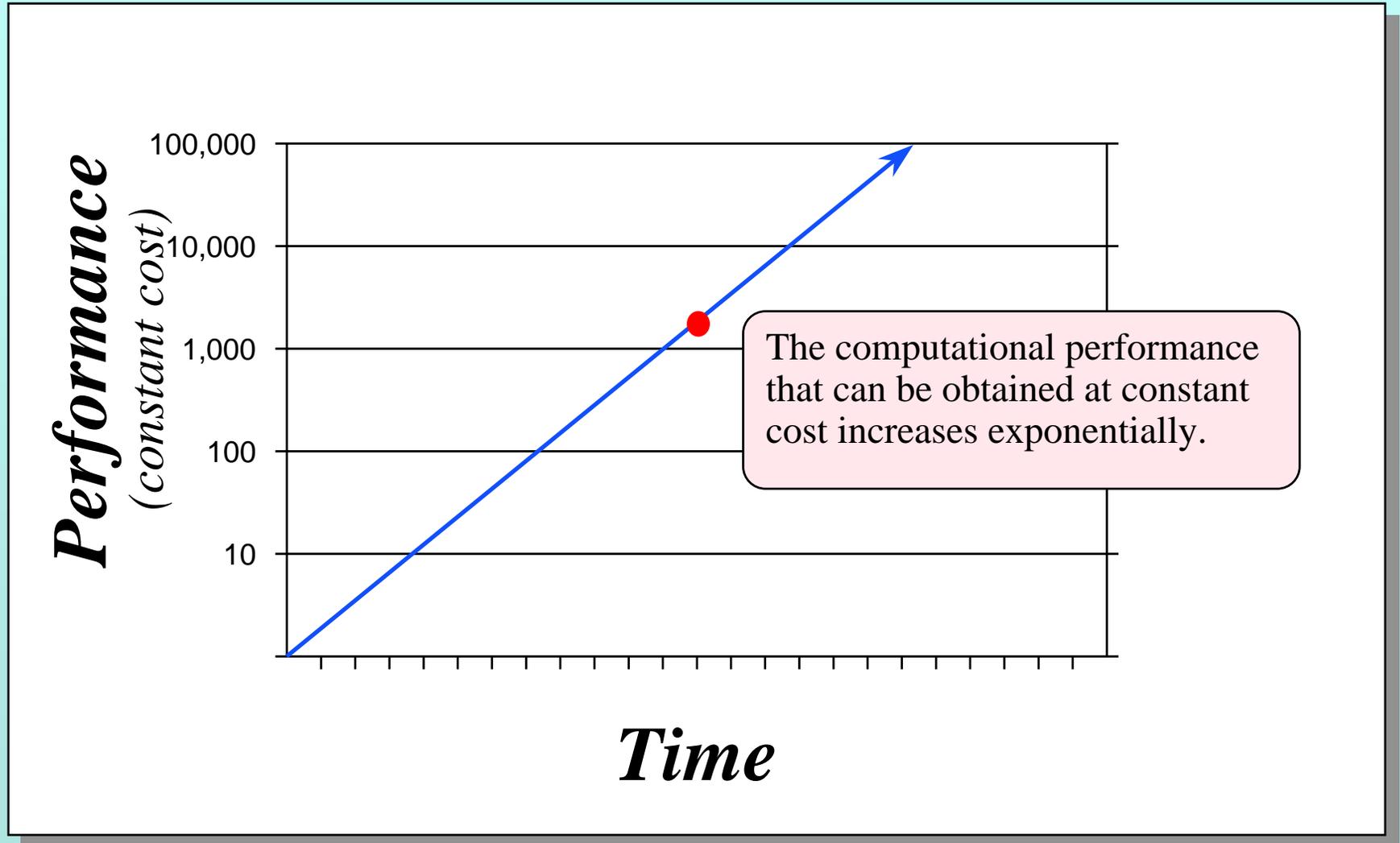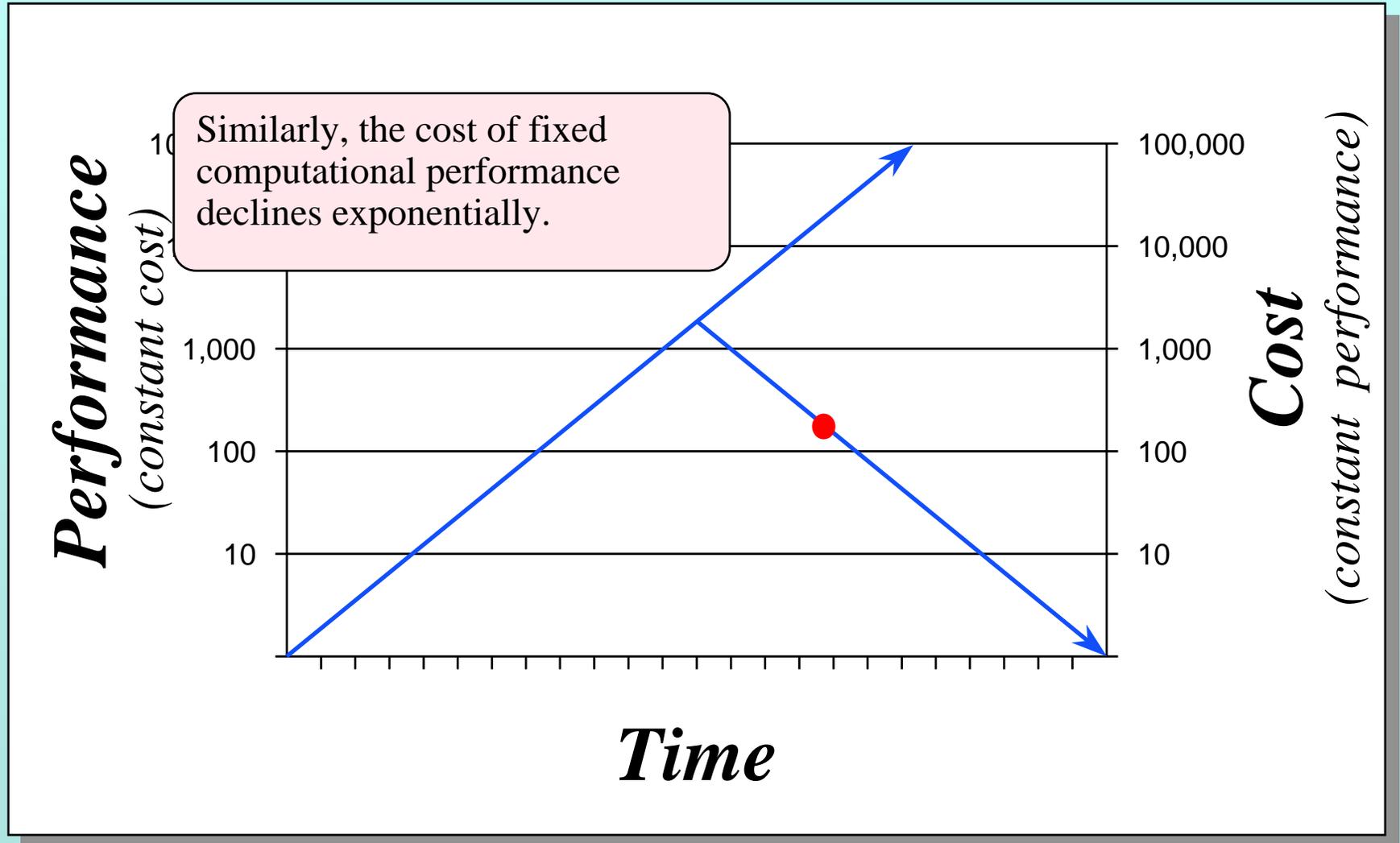
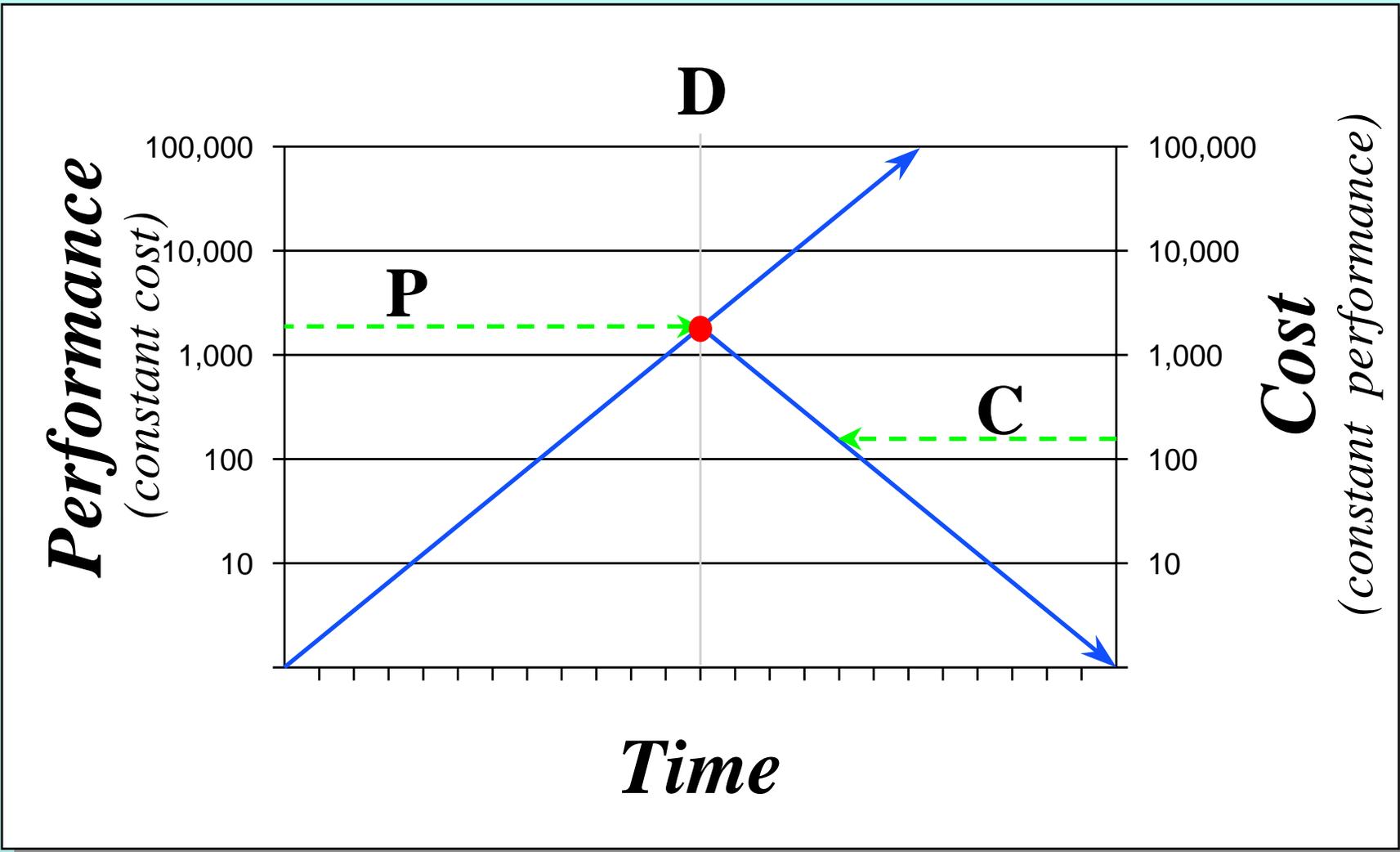The computational performance that can be obtained at constant cost increases exponentially.

Moore's Law: *The Effect*

Similarly, the cost of fixed computational performance declines exponentially.

Performance *(constant cost)*

Cost *(constant performance)*

Time

Moore's Law: *The Effect*

Moore's Law: *The Effect*

Performance *(constant cost)*

Cost *(constant performance)*
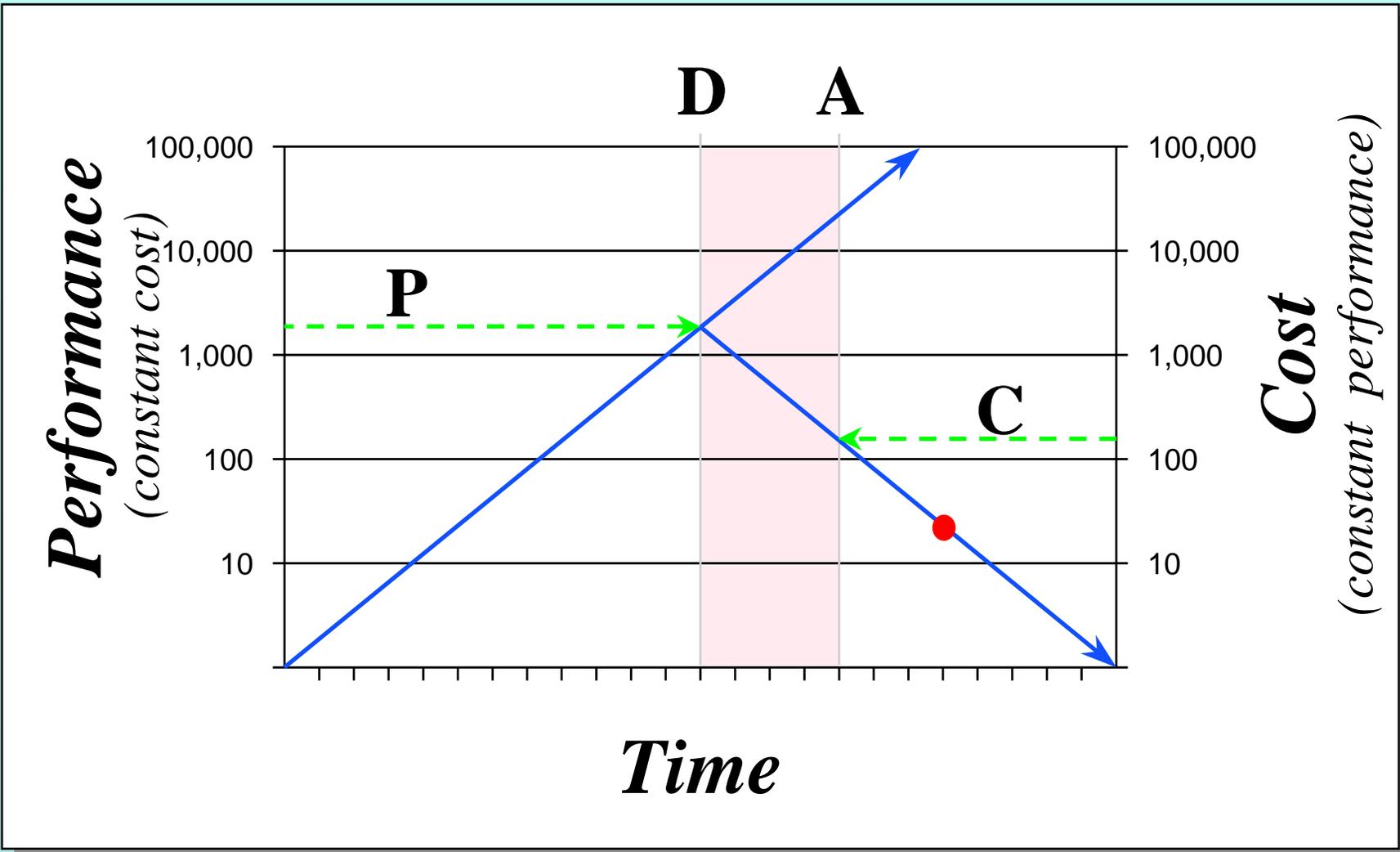
100,000
10,000
1,000
100
10

P

C

Time
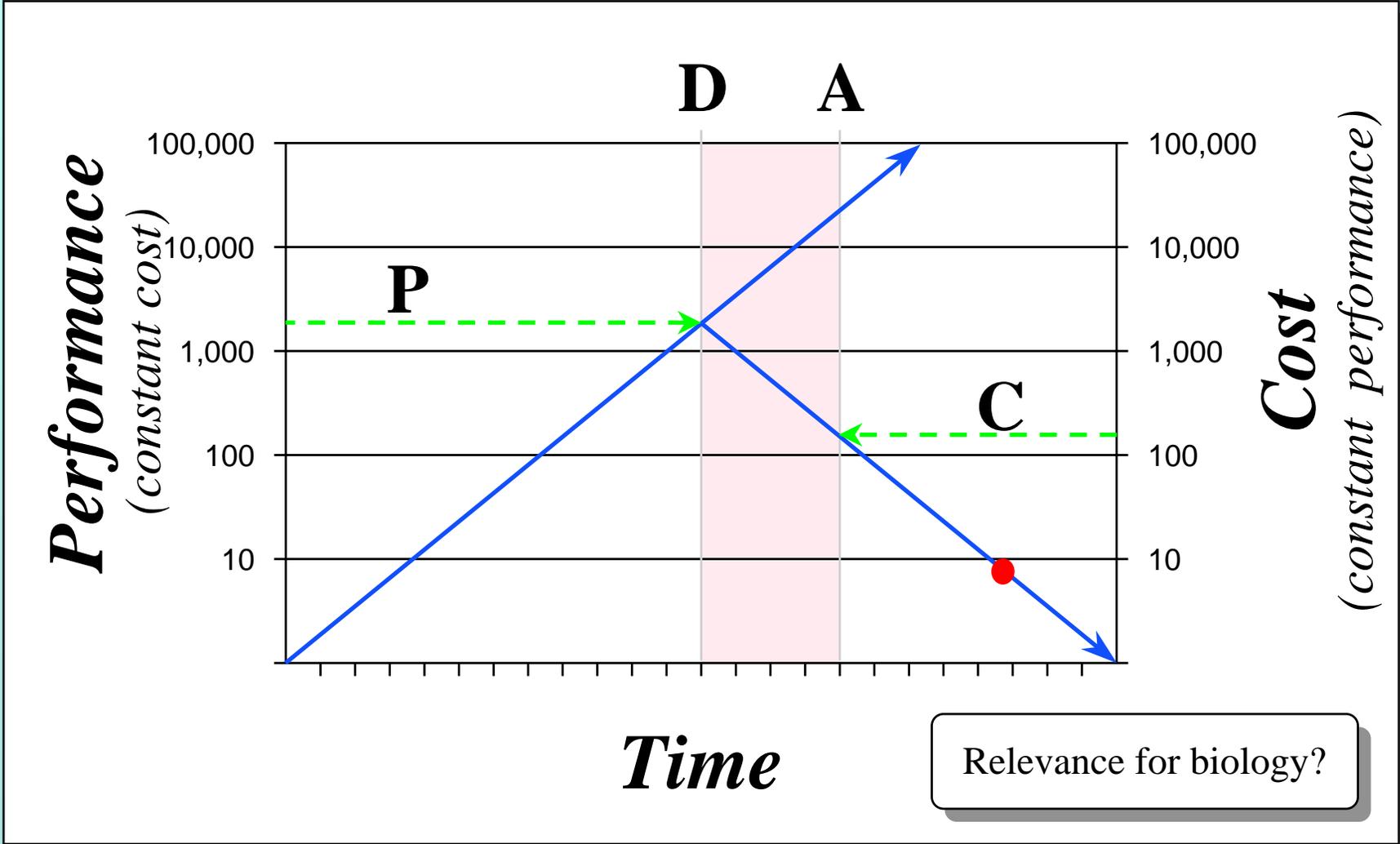
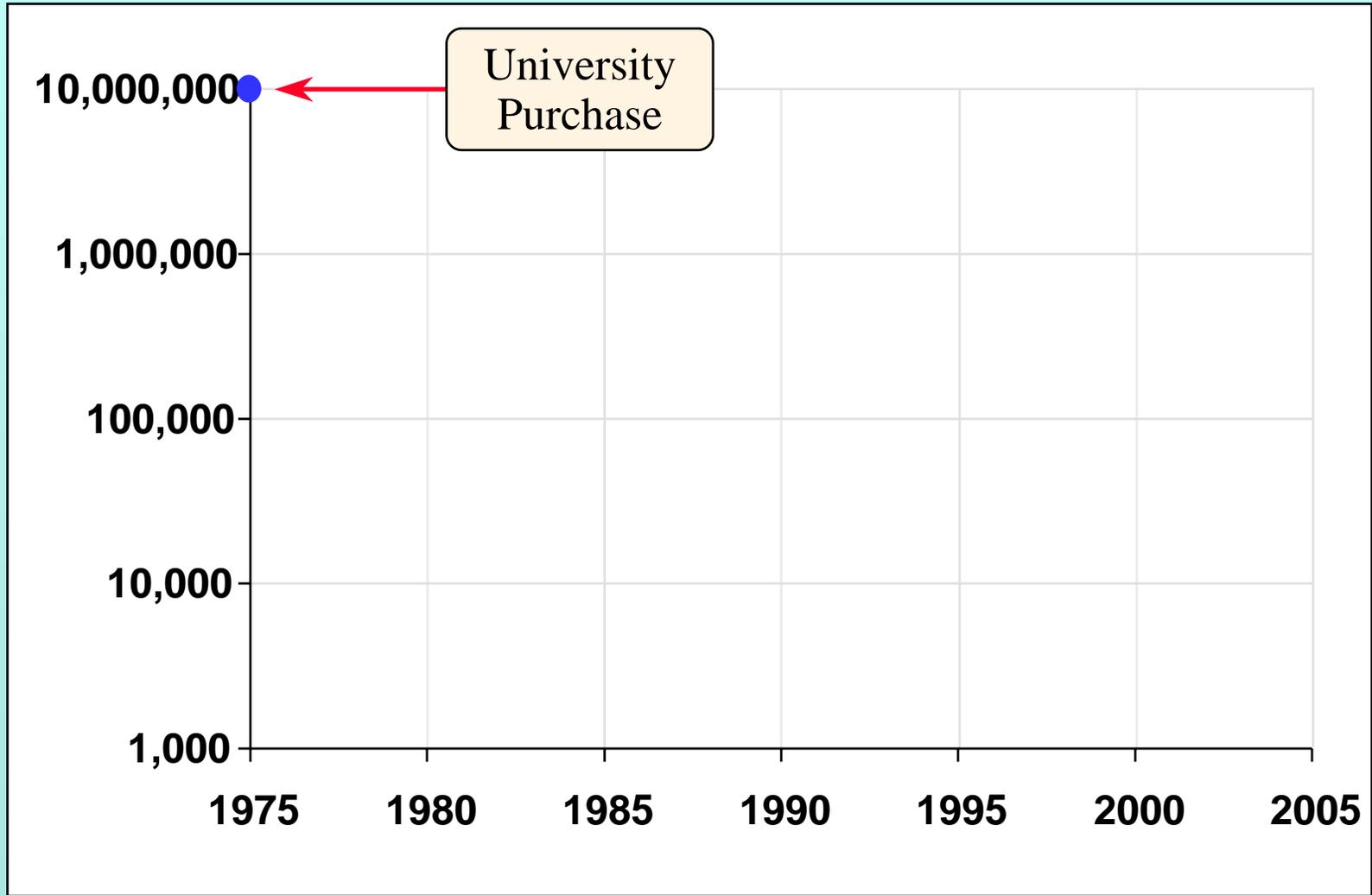# Moore's Law: *The Effect*

Moore's Law: *The Effect*
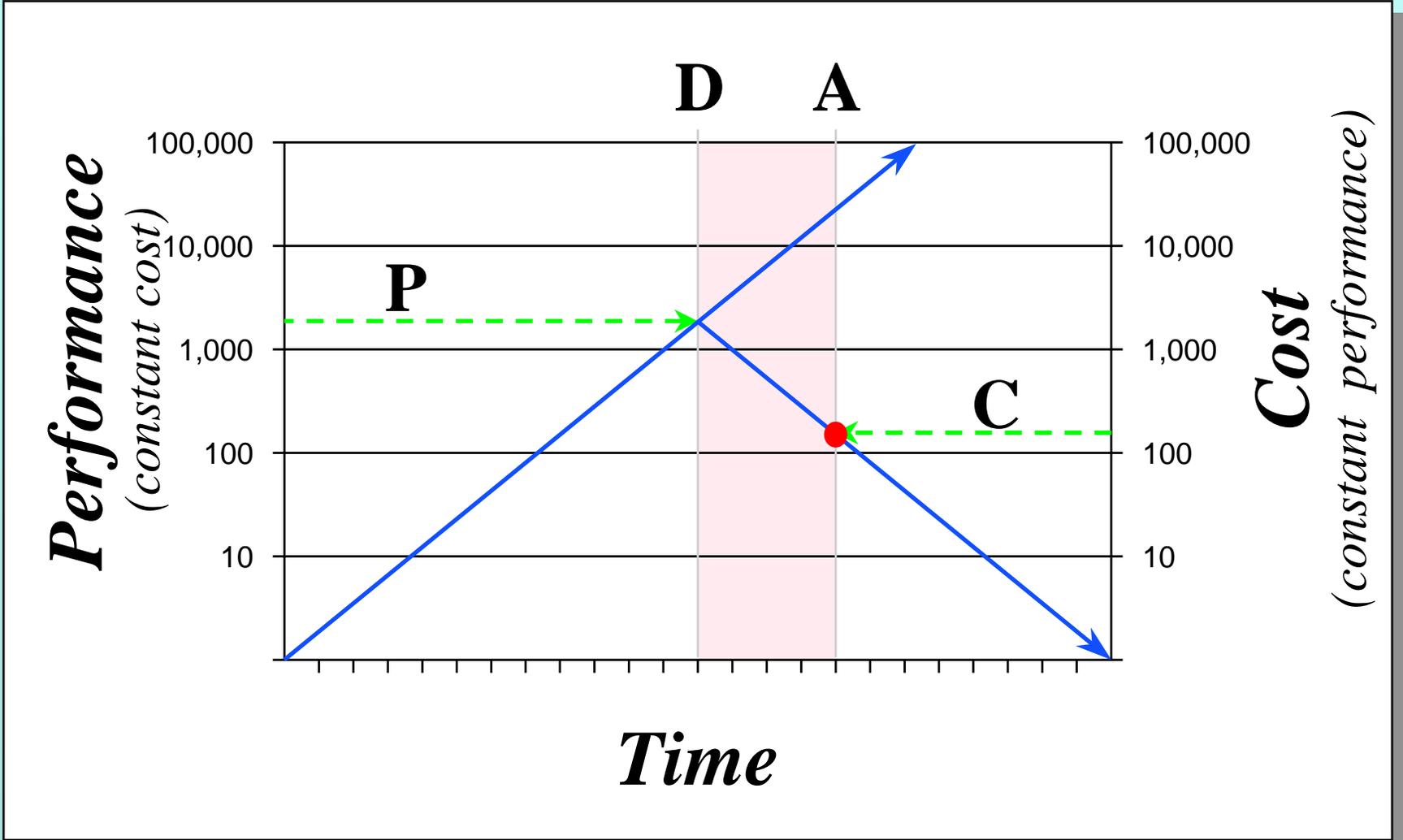
Moore's Law: *The Effect*

Moore's Law: *The Effect*

Cost (constant performance)

University Purchase

10,000,000

1,000,000

100,000

10,000

1,000

1975   1980   1985   1990   1995   2000   2005

Biomedical Research Institutions Information Technology Exchange   39

Cost (constant performance)
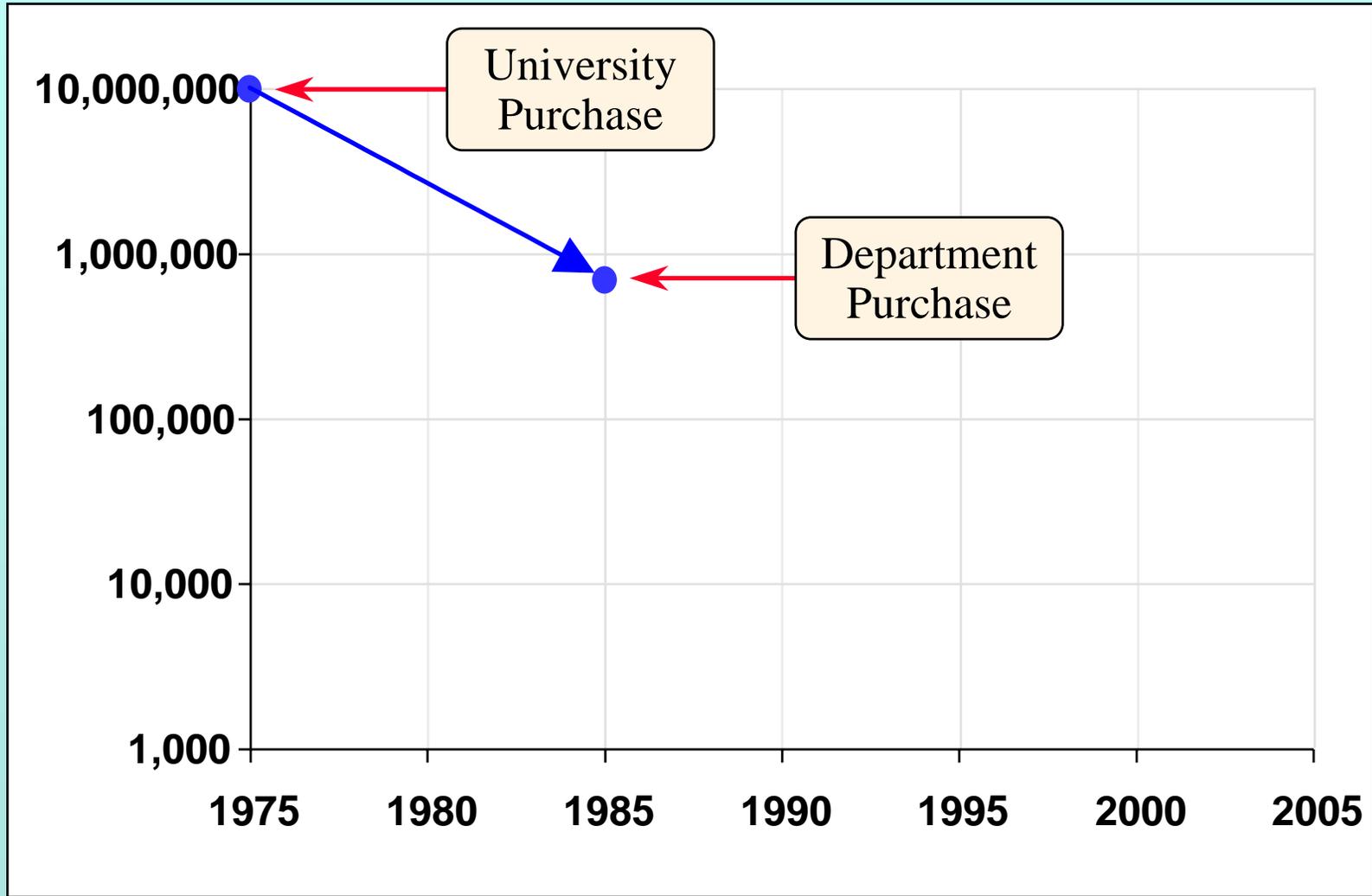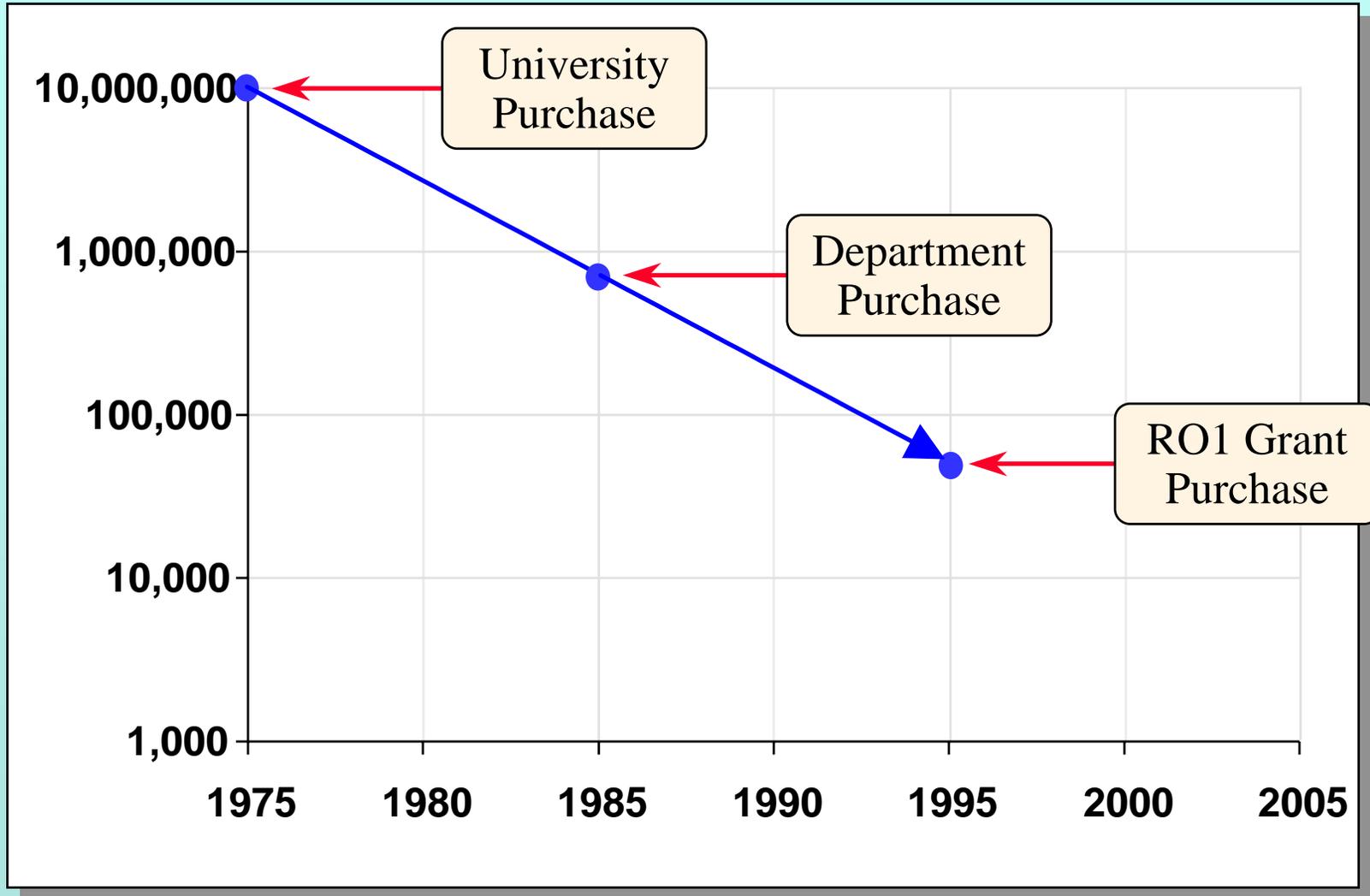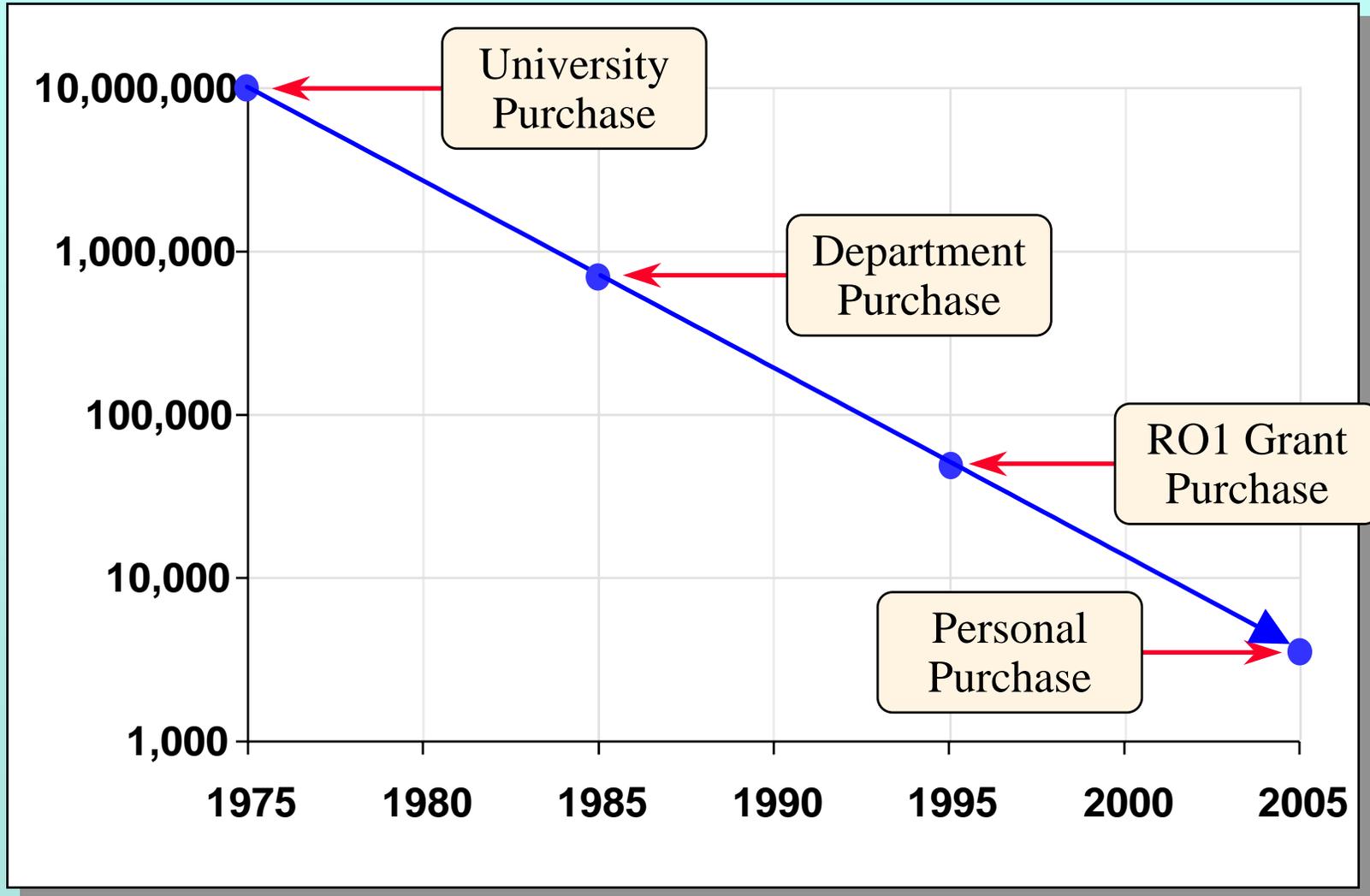
Cost (constant performance)

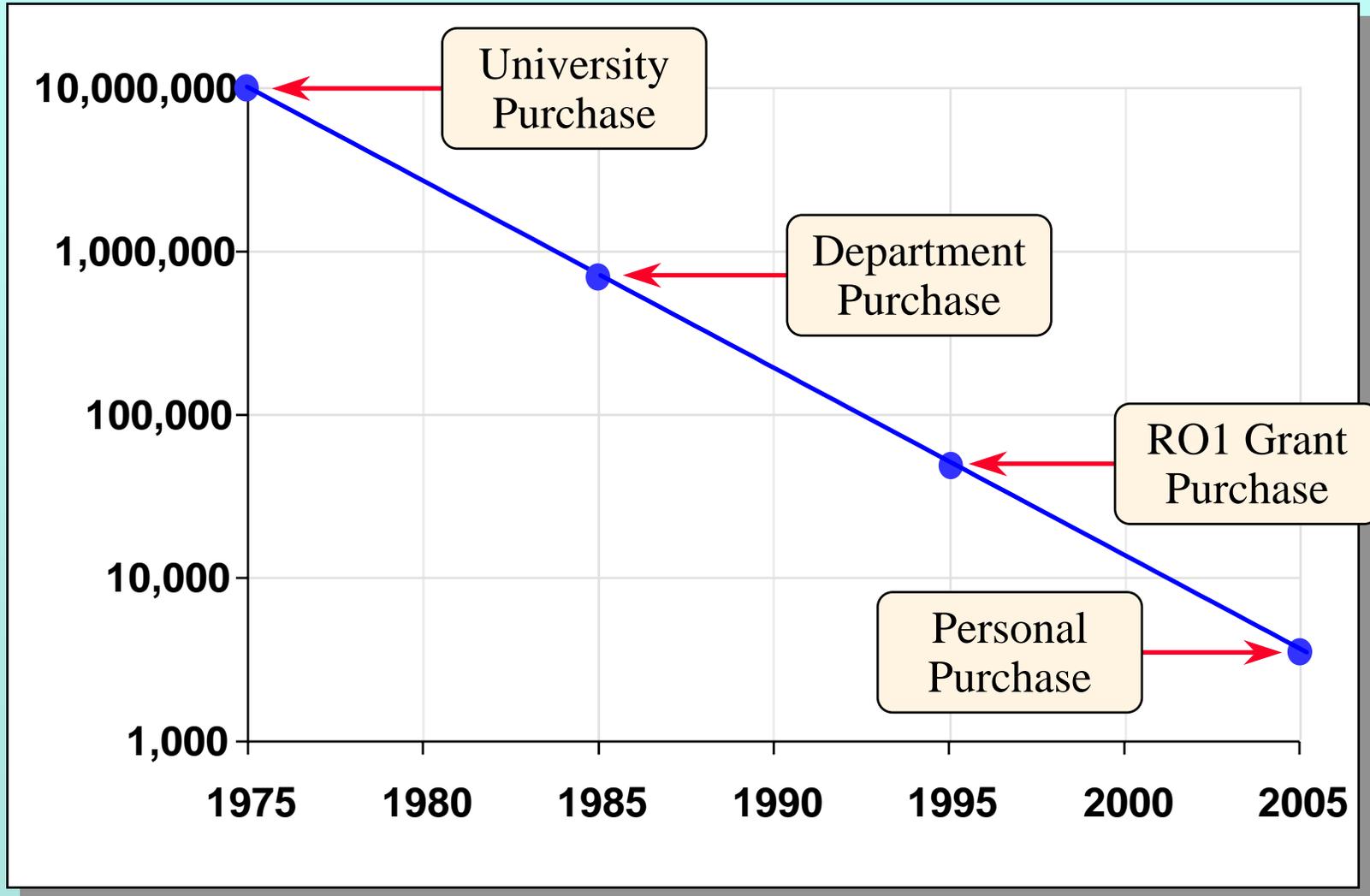# Cost (constant performance)

Cost (constant performance)

University Purchase

Department Purchase

RO1 Grant Purchase

Personal Purchase

10,000,000
1,000,000
100,000
10,000
1,000

1975   1980   1985   1990   1995   2000   2005

We're Late

- Late is good: it motivates.

- Late is bad: it encourages "light's better" thinking.

We're Late

- Late is good: it motivates.

- Late is bad: it encourages "light's better" thinking.

REMEMBER: A shortcut is often the quickest way to someplace you weren't going.

START

Whatever we've got,

it isn't enough…

Reality check: budgets

Reality Check

Resource Availability:

- Compared to the recent past, current government spending on biomedical information infrastructure is huge.

Reality Check

Resource Availability:

- Compared to the recent past, current government spending on biomedical information infrastructure is huge.

- Compared to what's needed, current government spending on bio-medical information infrastructure is tiny.

Rhetorical Question

Which is likely to be more complex:

- identifying, documenting, and tracking the whereabouts of **all parcels** in transit in the UPS system at one time

Rhetorical Question

Which is likely to be more complex:

- identifying, documenting, and tracking the whereabouts of **all parcels** in transit in the UPS system at one time

- identifying, documenting, and tracking all data, all materials, and all equipment relevant to all aspects of all publicly funded biomedical research, in all fields and on all topics.

Business Factoids

Five years ago, United Parcel Service:

- used redundant multi-terabyte databases to track all packages in transit

- had 4,000 full-time employees dedicated to IT

- spent one billion dollars per year on IT

- had an income of 1.1 billion dollars, against revenues of 22.4 billion dollars

Business Comparisons

Company	Revenues	IT Budget	Pct
Chase-Manhattan	16,431,000,000	1,800,000,000	10.95 %
AMR Corporation	17,753,000,000	1,368,000,000	7.71 %
Nation's Bank	17,509,000,000	1,130,000,000	6.45 %
Sprint	14,235,000,000	873,000,000	6.13 %
IBM	75,947,000,000	4,400,000,000	5.79 %
MCI	18,500,000,000	1,000,000,000	5.41 %
Microsoft	11,360,000,000	510,000,000	4.49 %
United Parcel	22,400,000,000	1,000,000,000	4.46 %
Bristol-Myers Squibb	15,065,000,000	440,000,000	2.92 %
Pfizer	11,306,000,000	300,000,000	2.65 %
Pacific Gas & Electric	10,000,000,000	250,000,000	2.50 %
Wal-Mart	104,859,000,000	550,000,000	0.52 %
K-Mart	31,437,000,000	130,000,000	0.41 %

Federal Funding of Biomedical-IT

Appropriate funding level:

- approx. 5-15% of research funding
- *i.e.,* **billions** of dollars per year

Biomedical Research Institutions Information Technology Exchange

Federal Funding of Biomedical-IT

Appropriate funding level:

- approx. 5-15% of research funding

- *i.e.,* **billions** of dollars per year

Seem high?

What percent of institutional operating budgets goes to other mature infrastructure?

Federal Funding of Biomedical-IT

A

Warning:

Until more resources become available, finding true SOLUTIONS to biomedical-IT problems will be impossible.

Seem high?

What percent of institutional operating budgets goes to other mature infrastructure?

Reality check:
Inadequate technology
&
"light's better" solutions

Caution from the Past

Scientific Database Management

Final Report

edited by

James C. French, Anita K. Jones, and John L. Pfalz

Report of the Invitational NSF Workshop on

Scientific Database Management

12–13 March 1990

Charlottesville, Virginia

Anita K. Jones, Chairperson

Technical Report 90-21

August 1990

Caution from the Past

U Va Tech Reports:

- ## CS-90-21

  J.C. French, A.K. Jones and J.L. Pfaltz, Scientific Database Management (Final Report), August 1990.

  ftp://ftp.cs.virginia.edu/pub/techreports/CS-90-21.ps.Z

- ## CS-90-22

  J.C. French, A.K. Jones and J.L. Pfaltz, Scientific Database Management (Panel Reports and Supporting Material), August 1990

  ftp://ftp.cs.virginia.edu/pub/techreports/CS-90-22.ps.Z

Caution from the Past

Two major conclusions:

- The single unifying cry of the workshop is that existing data models are inadequate for science data needs. (p. 6)

Caution from the Past

Two major conclusions:

- The single unifying cry of the workshop is that existing data models are inadequate for science data needs. (p. 6)

- The data source dimension (e.g., single or multi-source), which is not generally mentioned in the database literature, may present the most fundamental challenge. (p. 3)

Data Model Problems

Topics

- Data-model problems

  No support for rapid schema change.

  No support for inter-database referential integrity.

  No support for tri-state logic.

  No support for object-identity ambiguity.

  And there's more…

Biomedical Research Institutions Information Technology Exchange

Topics

- Data-model problems

   If existing data models are inadequate, then much of the power (and extensibility) of DBMSs are lost, because systems built on them are fundamentally kludges.

   An information system built on a bad data model is like a house built on sand…

Data Source Problems

Topics

- Data-source problems

  Biology is a small-instrument, multi-source science.

  Integrating multi-source data is hard.
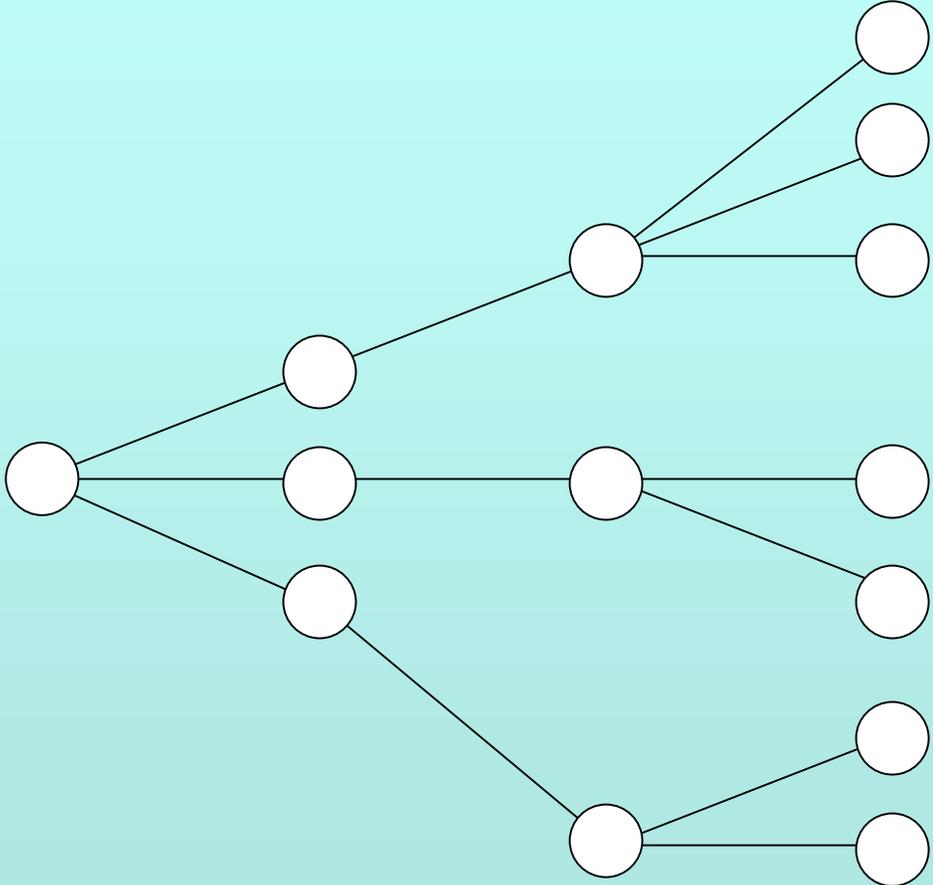
  Consistency flows in the wrong direction.

  GenBank is a false model.

Source I

Basics

Single-instrument Science

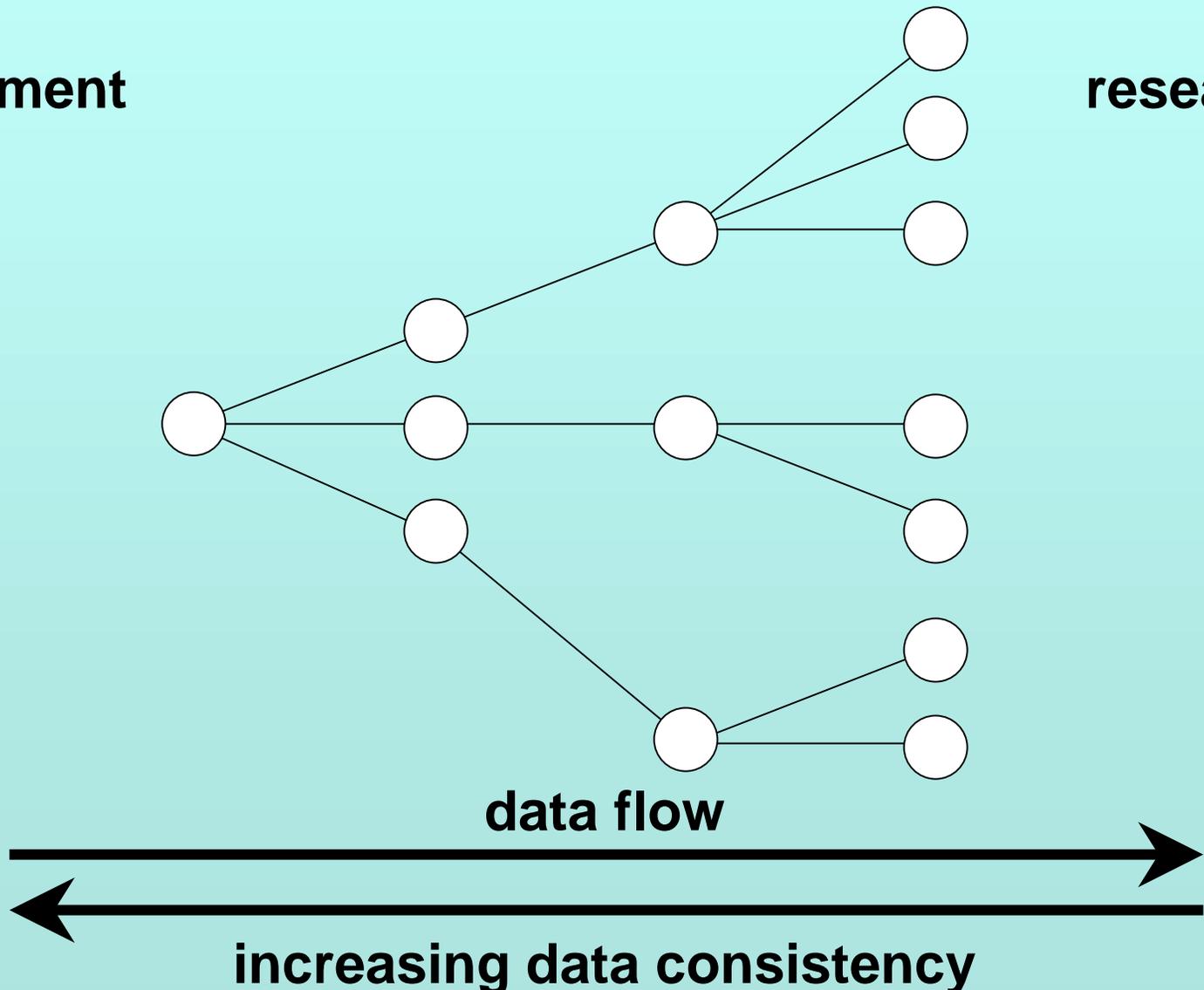**instrument**   **researchers**

data flow

Single-instrument Science

instrument

researchers

data flow

increasing data consistency

Single-instrument Science

instrument   researchers

RIGHT WAY:

With single-source science, data is MOST consistent nearest the source, making integration unnecessary (but making the need for path documentation high).

data flow

increasing data consistency

Multi-instrument Science

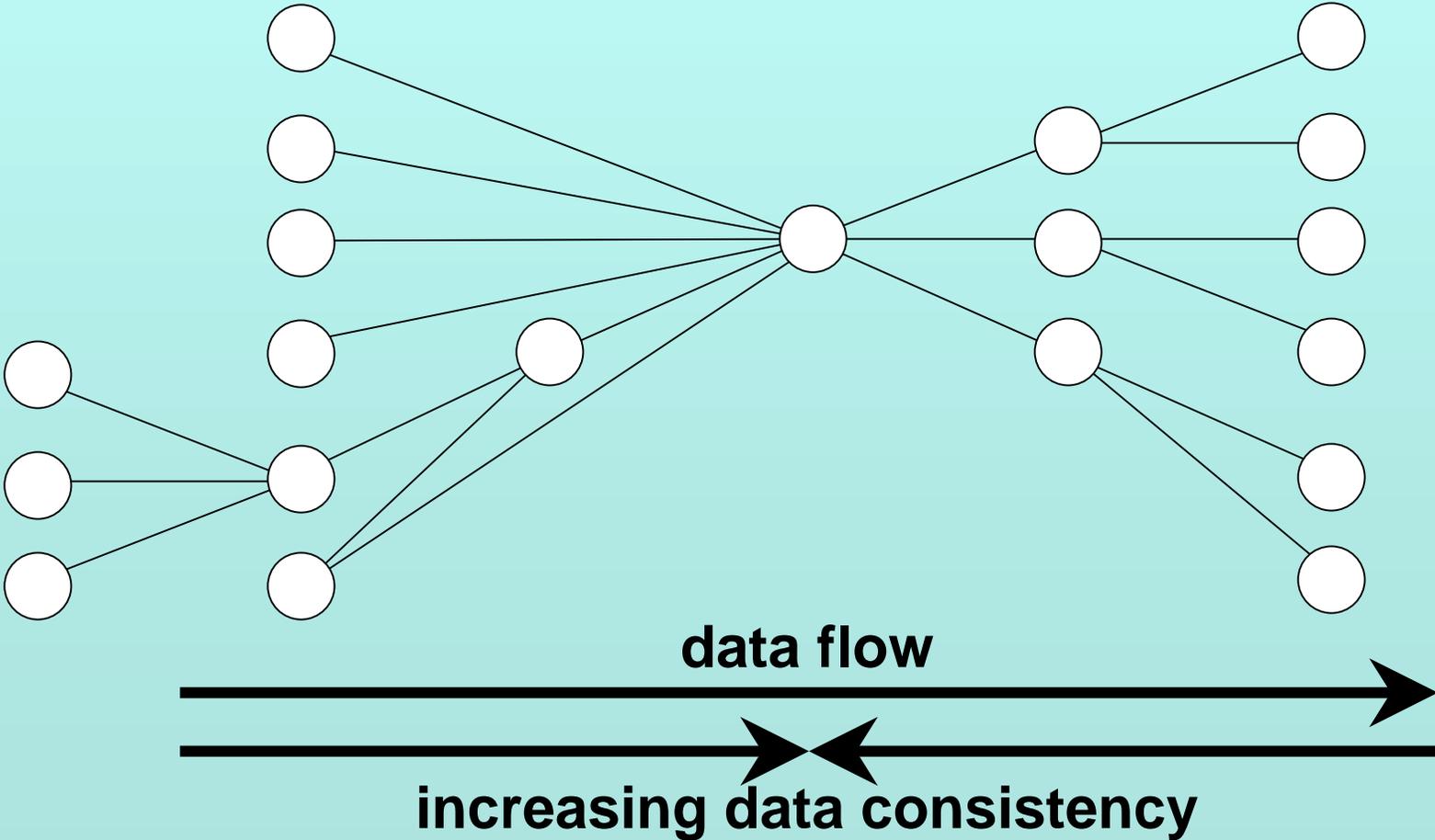**researchers**　　　　　**data resource(s)**　　　　　**researchers**

data flow

　　　Biomedical Research Institutions Information Technology Exchange
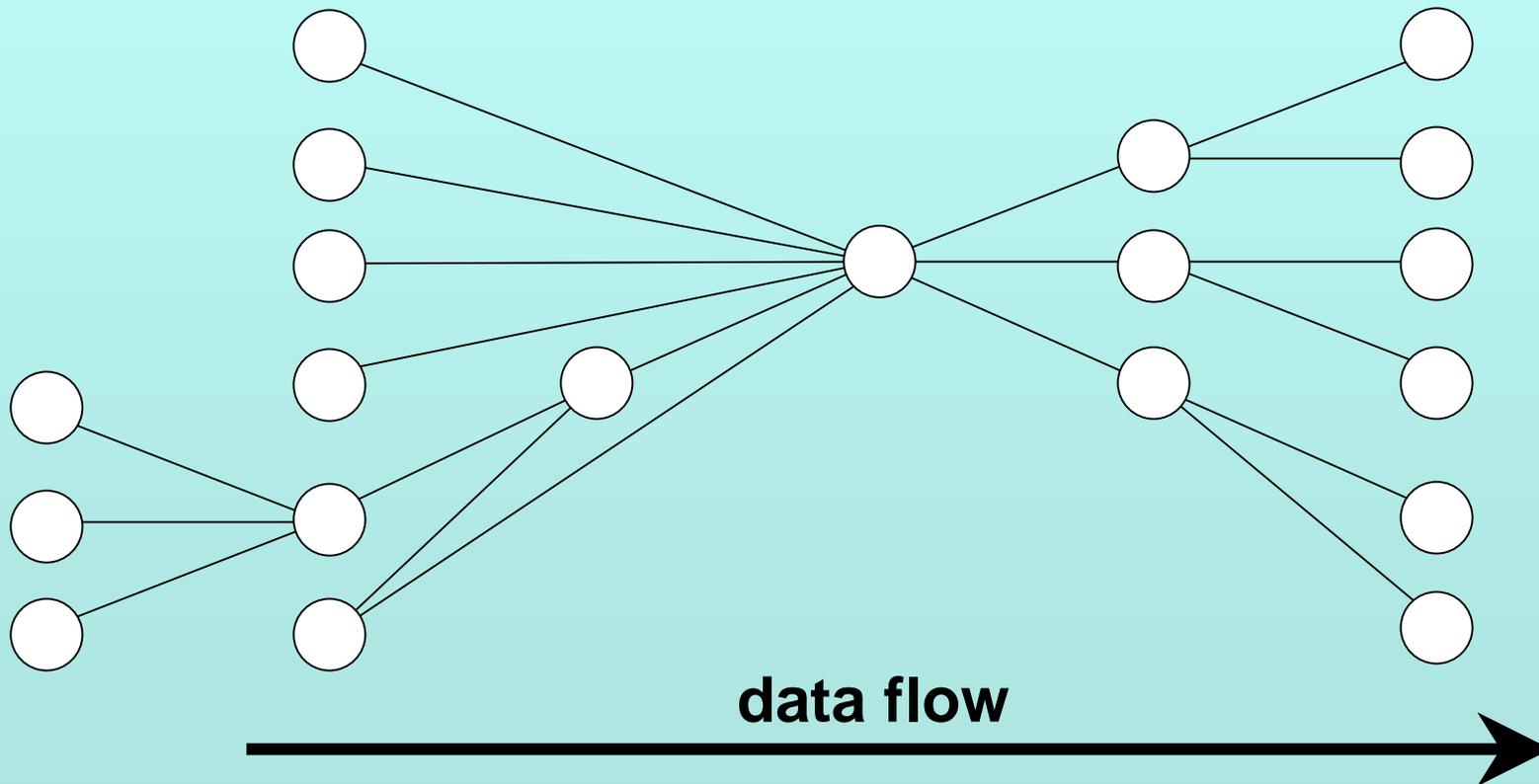
Multi-instrument Science

Multi-instrument Science

researchers　　　　　data resource(s)　　　　　researchers

STOP – WRONG WAY:

With multi-source science, data is LEAST consistent nearest the source, making true integration difficult.
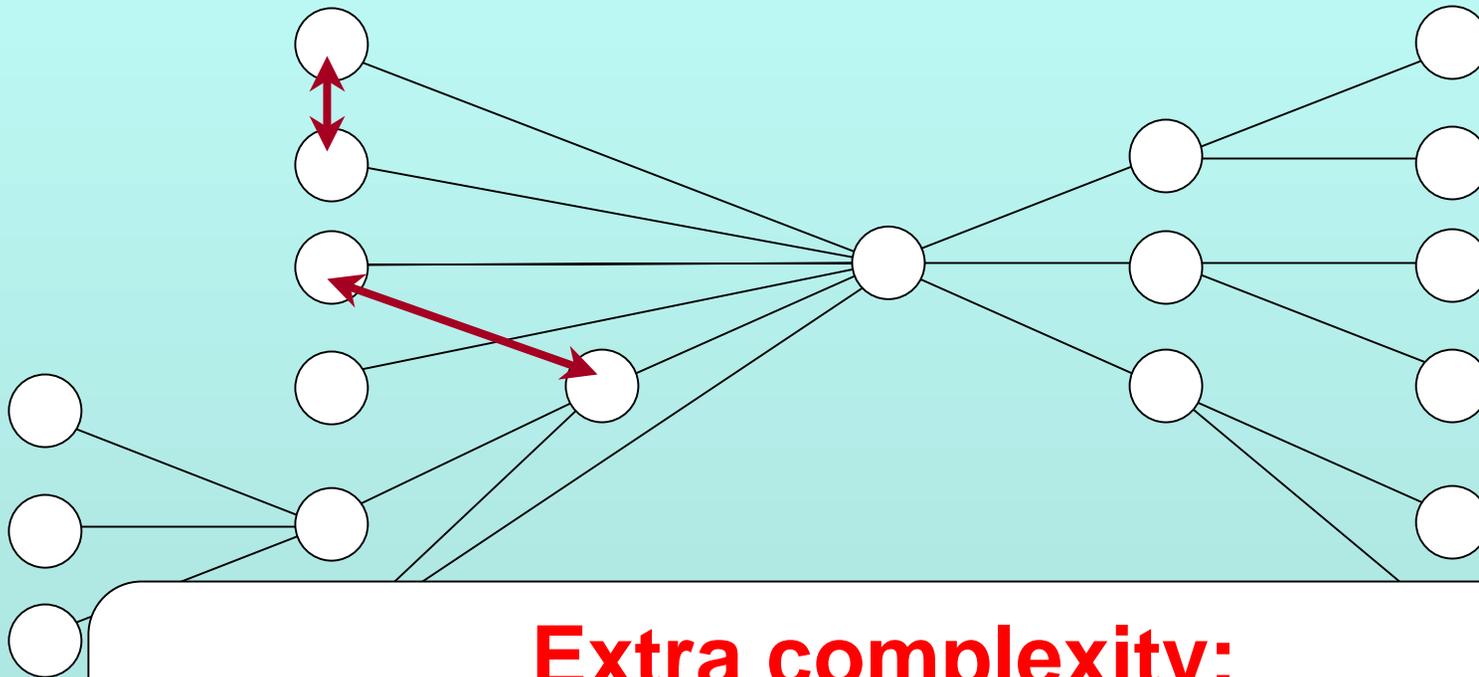
data flow

increasing data consistency

Multi-instrument Science

researchers   **data resource(s)**   **researchers**

Extra complexity:

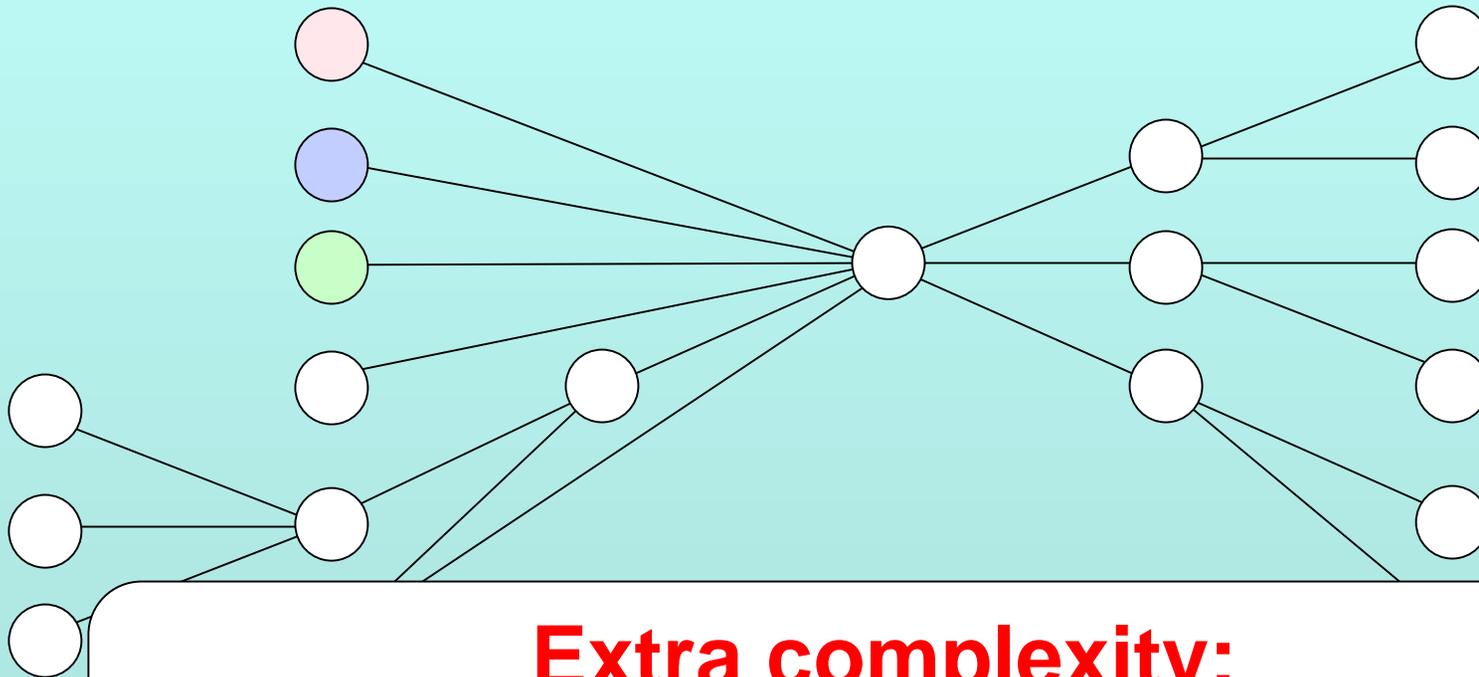Undocumented, uncoordinated local data exchange among producers

Multi-instrument Science

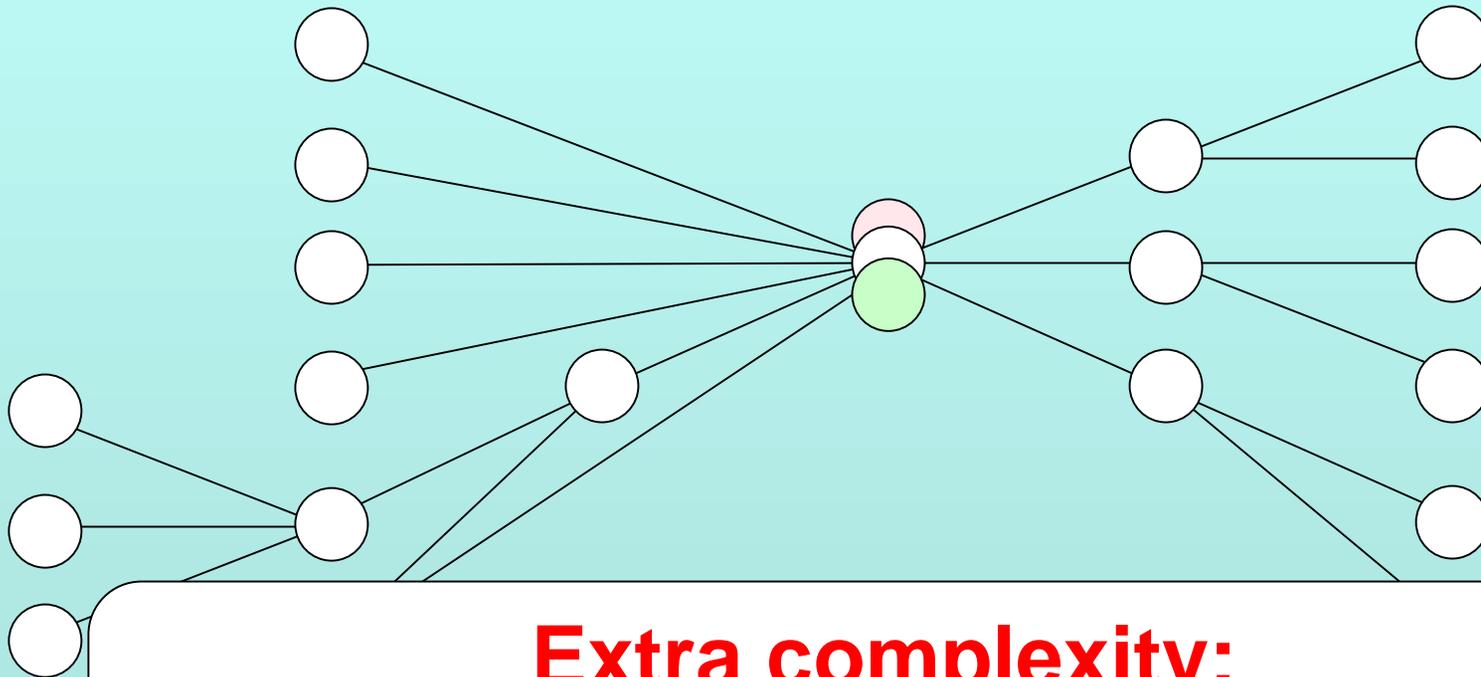researchers   data resource(s)   researchers

Extra complexity:

Data collected locally to meet local needs are not globally consistent - or even equivalent.

Biomedical Research Institutions Information Technology Exchange

75

Multi-instrument Science

researchers   data resource(s)   researchers

Extra complexity:

Multiple centralized resources may exist, meaning there is no authoritative source.

Source II
Scope

Data-source Scope Issues

Problems occur at many levels:

- Integrating sequence data into GenBank

- Connecting GenBank with other genomic resources

- Connecting genomic data with other biological data

- Connecting all biological data with medical data

- Connecting all biomedical data with…

Reality check:
"light's better" solutions

"Light's better" Solutions

- GenBank provides a model for public biomedical databases

- LSIDs can serve as database keys

- True DB solutions are too hard; IR-like solutions are adequate.

- Perfection is the enemy of the good; what we need is a commitment to getting things done using an 80:20 approach.
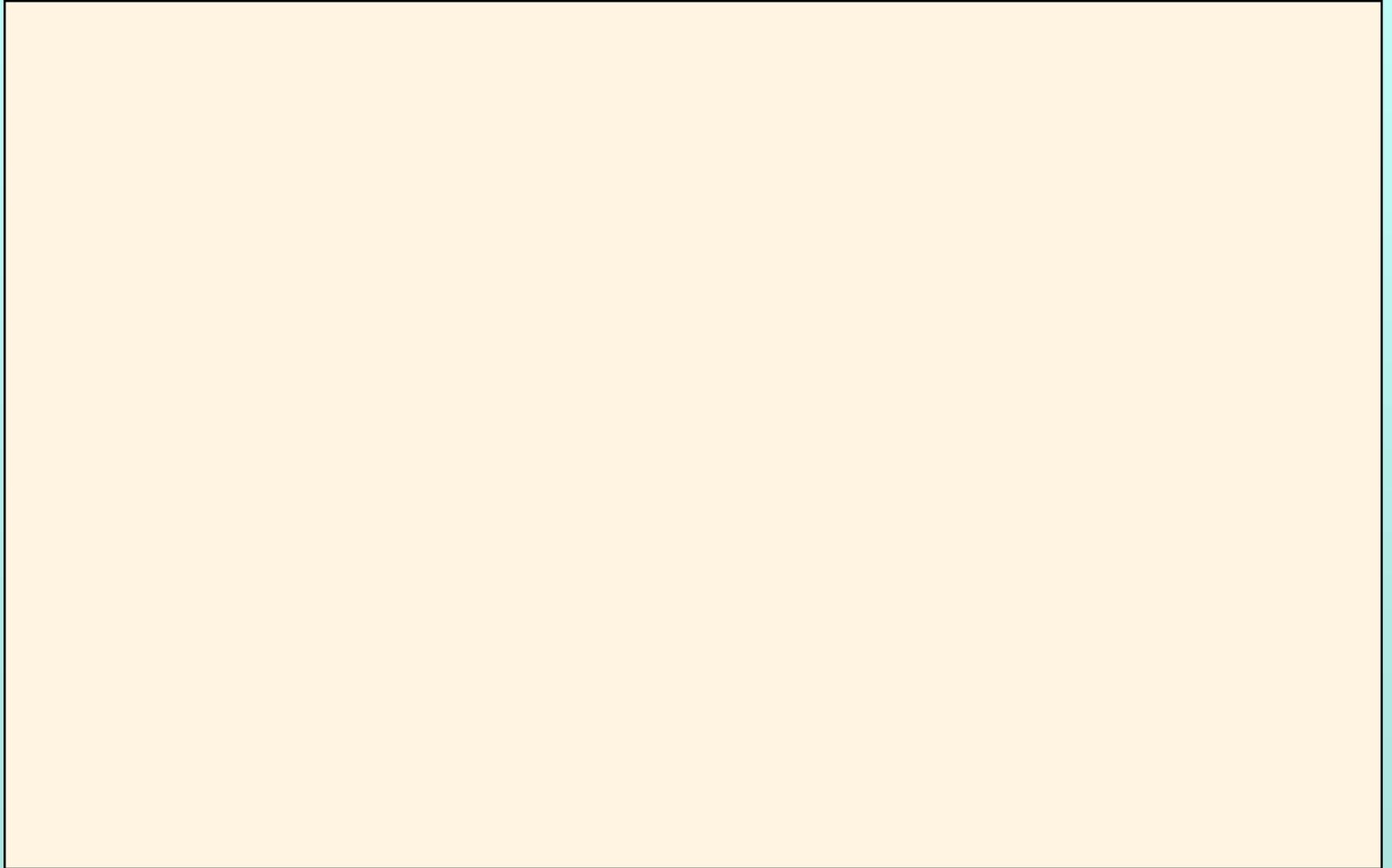
GenBank as a False Model

- Classic Kuhnian paradigm science

- Simple, unambiguous data type (string)

- Symbiotic relationship with publishers

- Sequences are things, not actions

Syntactics vs. Semantics, I

Syntactics vs. Semantics, I

A: If you call a dog's tail a leg, how many legs does a dog have?

B: Uh, five?

Syntactics vs. Semantics, I

A: If you call a dog's tail a leg, how many legs does a dog have?

B: Uh, five?

A: Nope, four! Calling a tail a leg don't make it one.

Syntactics vs. Semantics, II

A: If you use LSIDs as primary keys for biomedical databases, does that solve the problem of inter-database referential integrity.

B: Uh, maybe?

Syntactics vs. Semantics, II

A: If you use LSIDs as primary keys for biomedical databases, does that solve the problem of inter-database referential integrity.

B: Uh, maybe?

A: Nope! Calling an LSID a key don't make it one.

Biomedical Research Institutions Information Technology Exchange

Reality check:
No silver bullet

Data-source Solutions

Institutional Solutions:

- Getting from RO1 science to international standards is too big a step

- We need solutions at the levels of research institution and research community.

- Biomedical research organizations need to provide coherent support for biomedical IT, just as they do for biomedical bench research.

- Integrating institutional solutions is feasible; integrating individual lab solutions is not.

PATH

How do we get there from here?

PATH Issues

- Multiple solutions require multiple paths.

PATH Issues

- Multiple solutions require multiple paths.

And these paths should be followed by different actors, concerned with different scopes, acting at different scales…

PATH Issues

- Multiple solutions require multiple paths.

- Multiple solutions require an over-arching architectural vision, informed by an appropriate level of conceptual abstraction.

PATH Issues

- Multiple solutions require multiple paths.

- Multiple solutions require an over-arching

Combining an architectural vision with appropriate abstraction will be critical. The biggest challenge is that a truly open federated infrastructure will have to be context-free, and removing context removes most of the familiar guidelines for good design.

PATH Issues

- Multiple solutions require multiple paths.

- Multiple solutions require an over-arching architectural vision, informed by an appropriate level of conceptual abstraction.

- Standards must be chosen wisely: enabling standards must be embraced, constraining standards rejected, and premature standards avoided.

PATH Issues

- Multiple solutions require multiple paths.

- Multiple solutions require an over-arching architectural vision, informed by an appropriate level of conceptual abstraction.

- Standards must be chosen wisely: enabling standards must be embraced, constraining standards rejected, and premature standards avoided.

- Unexpected change (in everything) must be assumed and accommodated.

PATH Issues

- We must track industry trends and plot a course that matches our efforts with the prevailing industry winds.
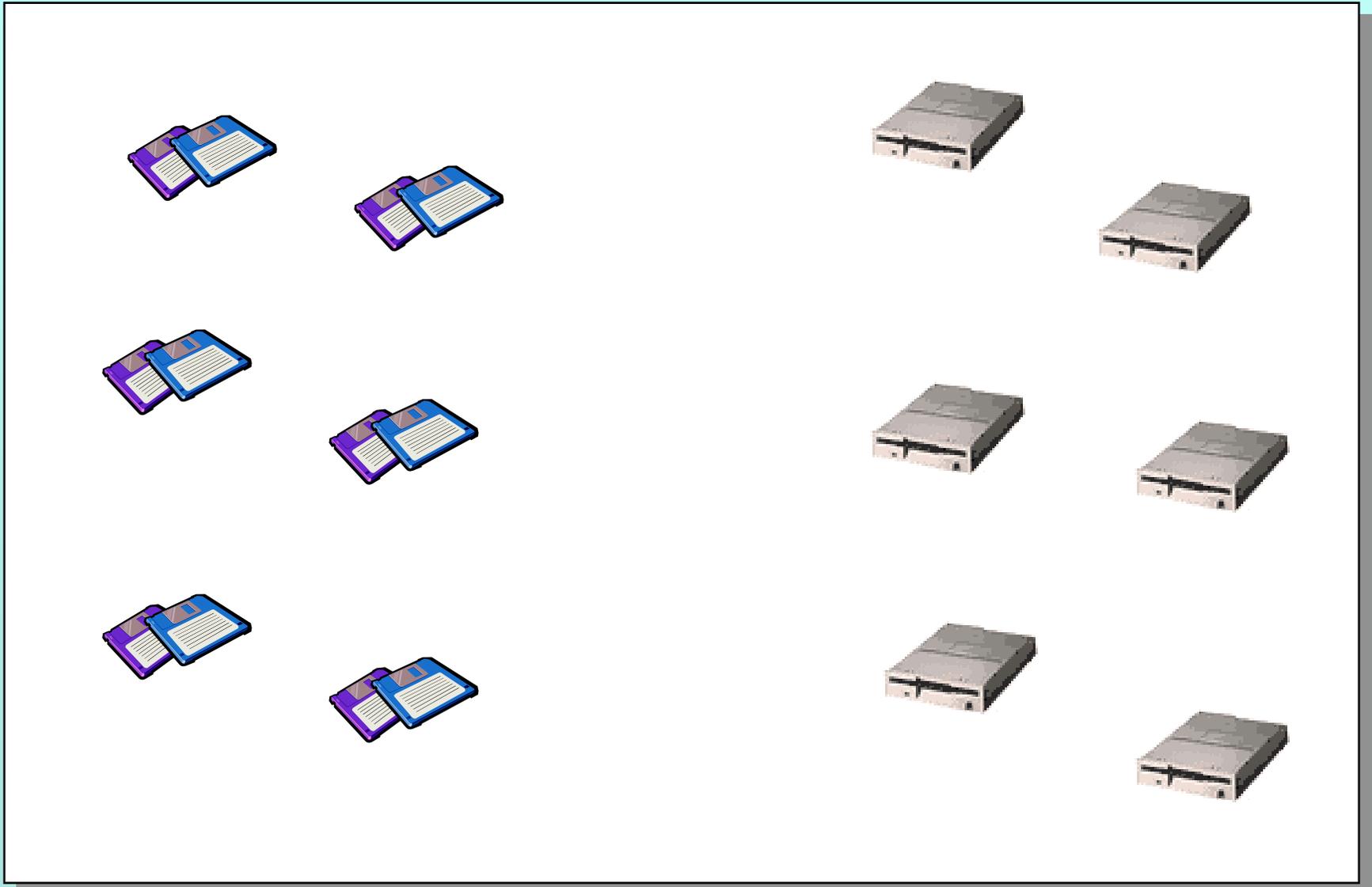
PATH Issues

- We must track industry trends and plot a course that matches our efforts with the prevailing industry winds.

The biomedical research community cannot control the direction of IT research and development, but we can take advantage of if (and possibly influence it somewhat).
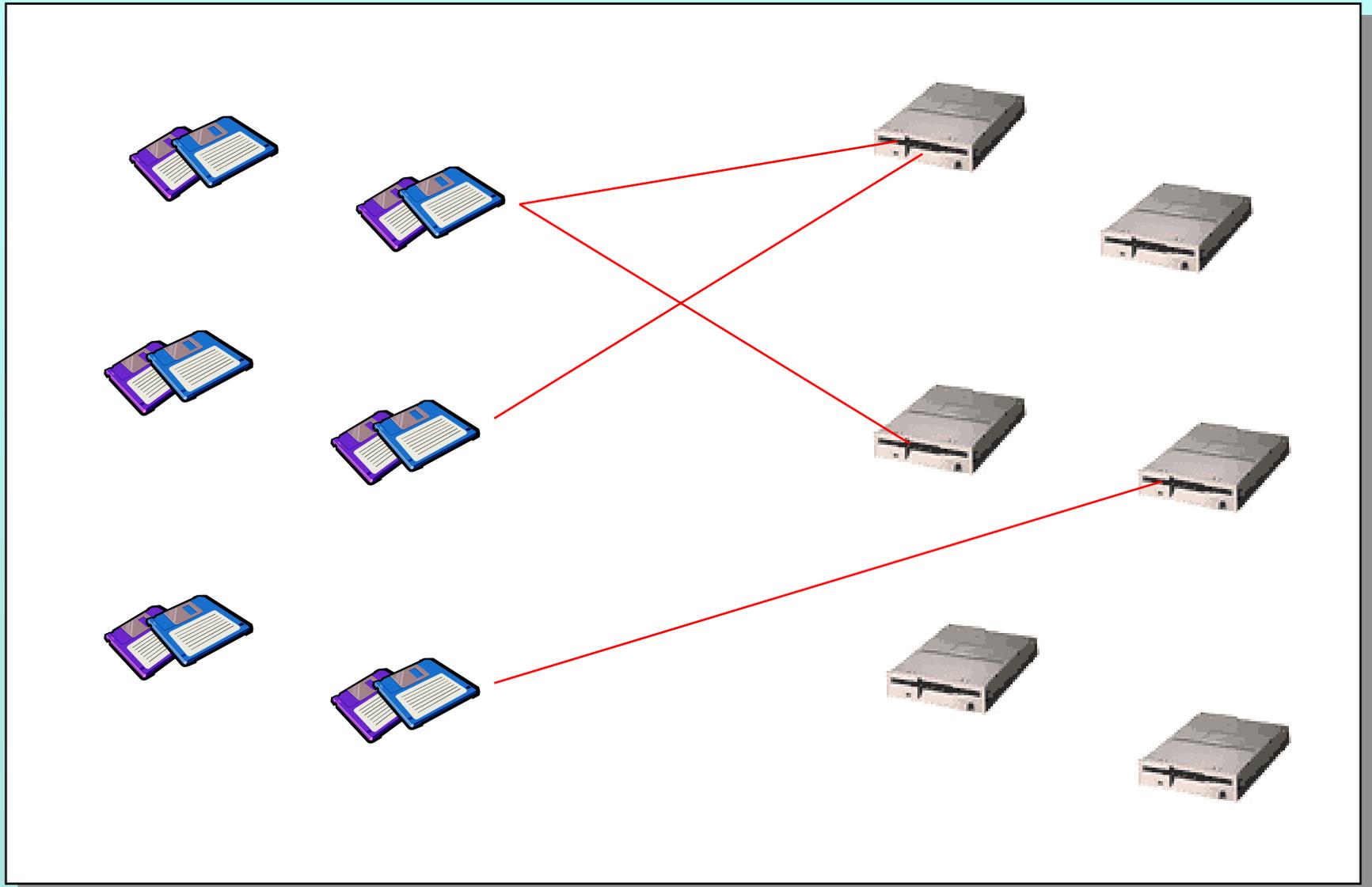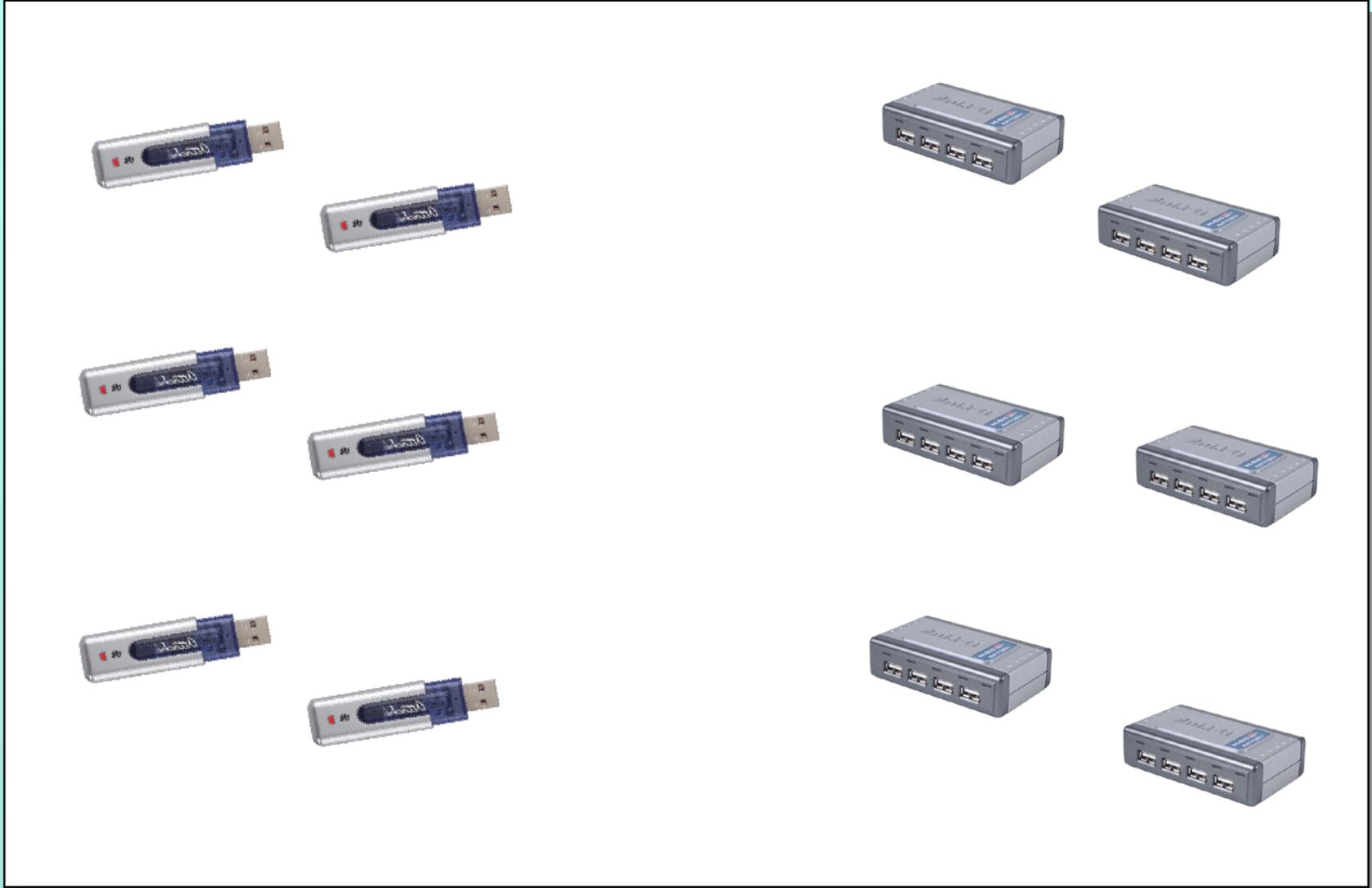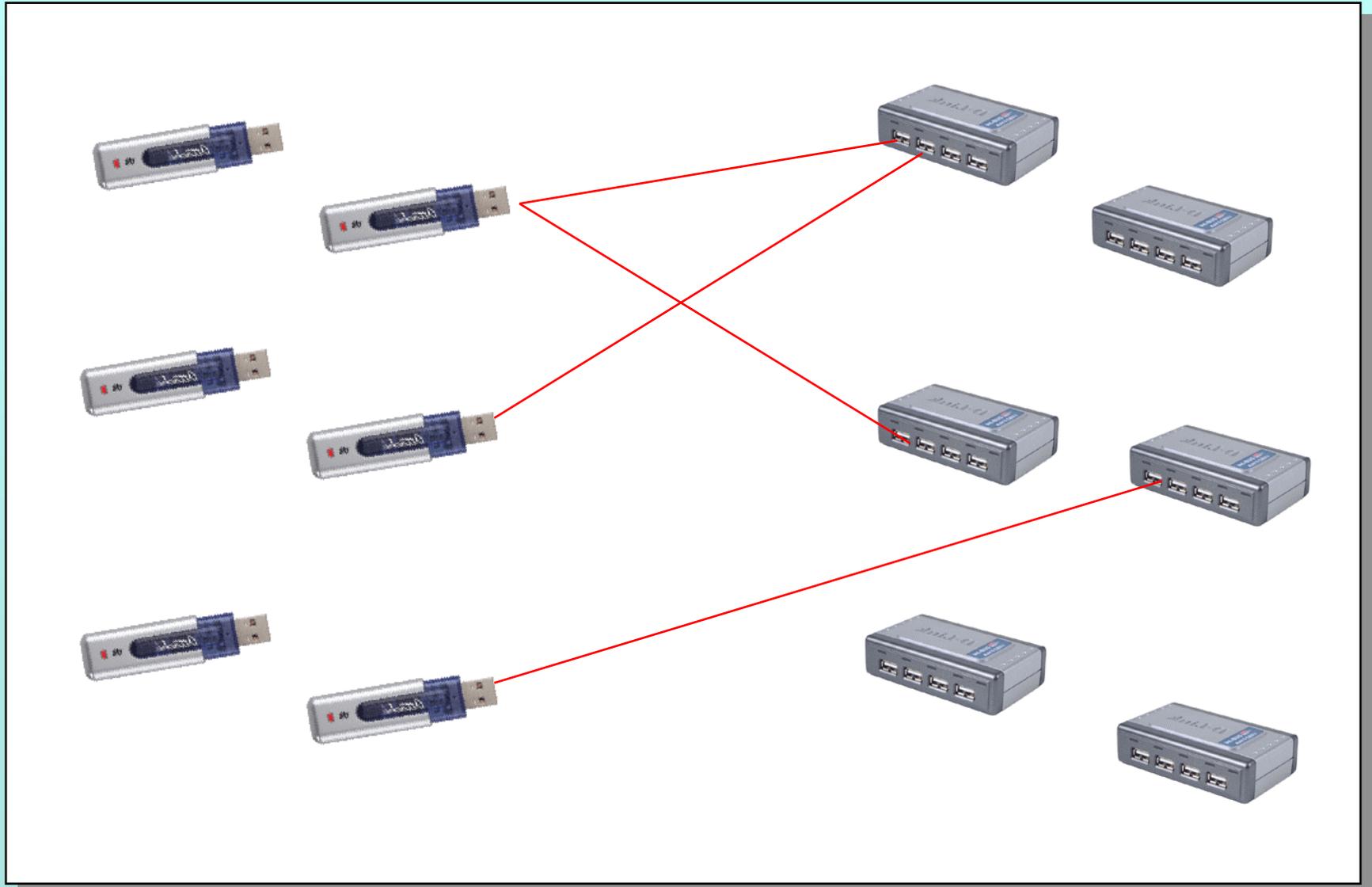
Standards

Bad Data-exchange Standard
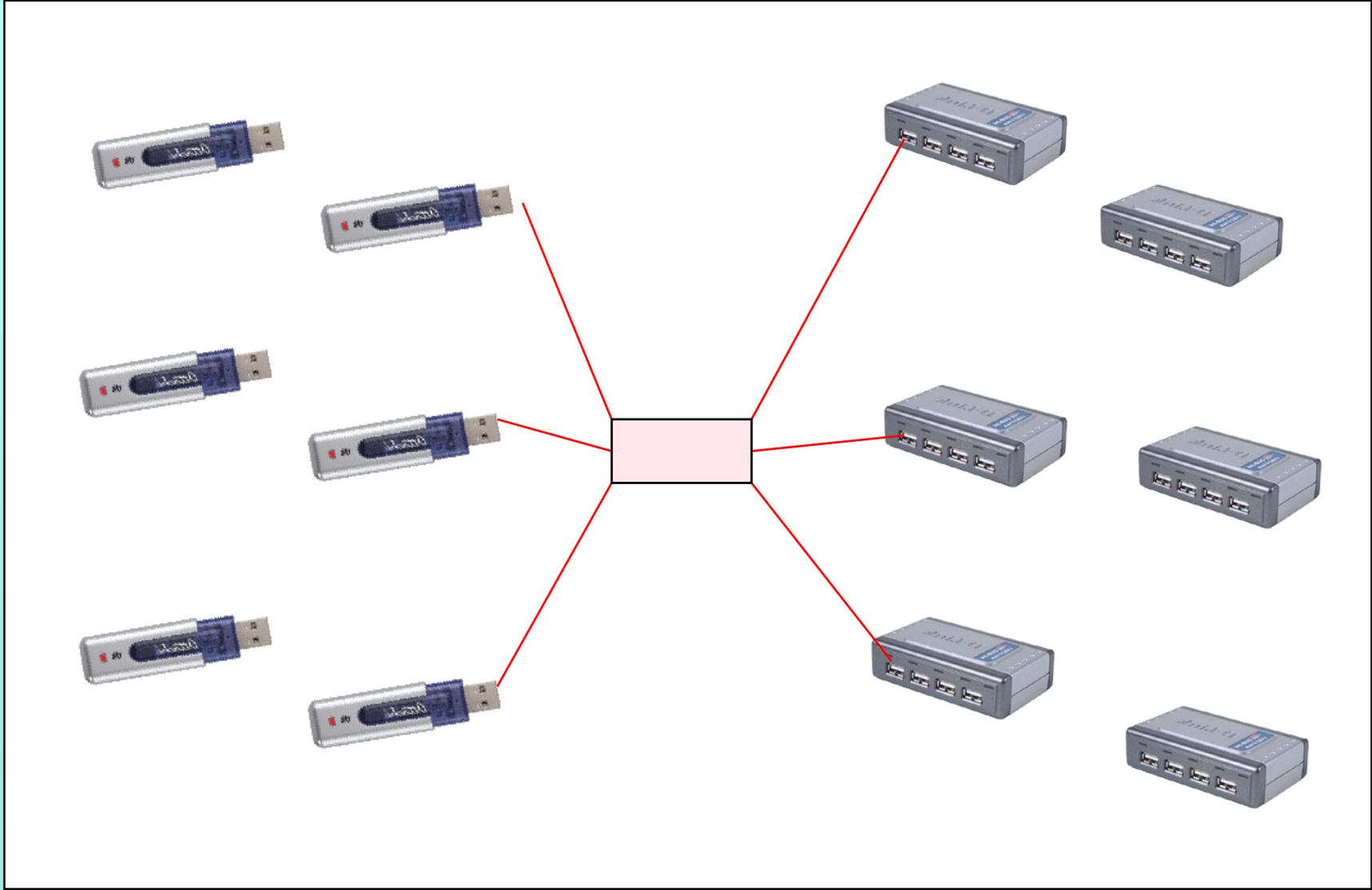
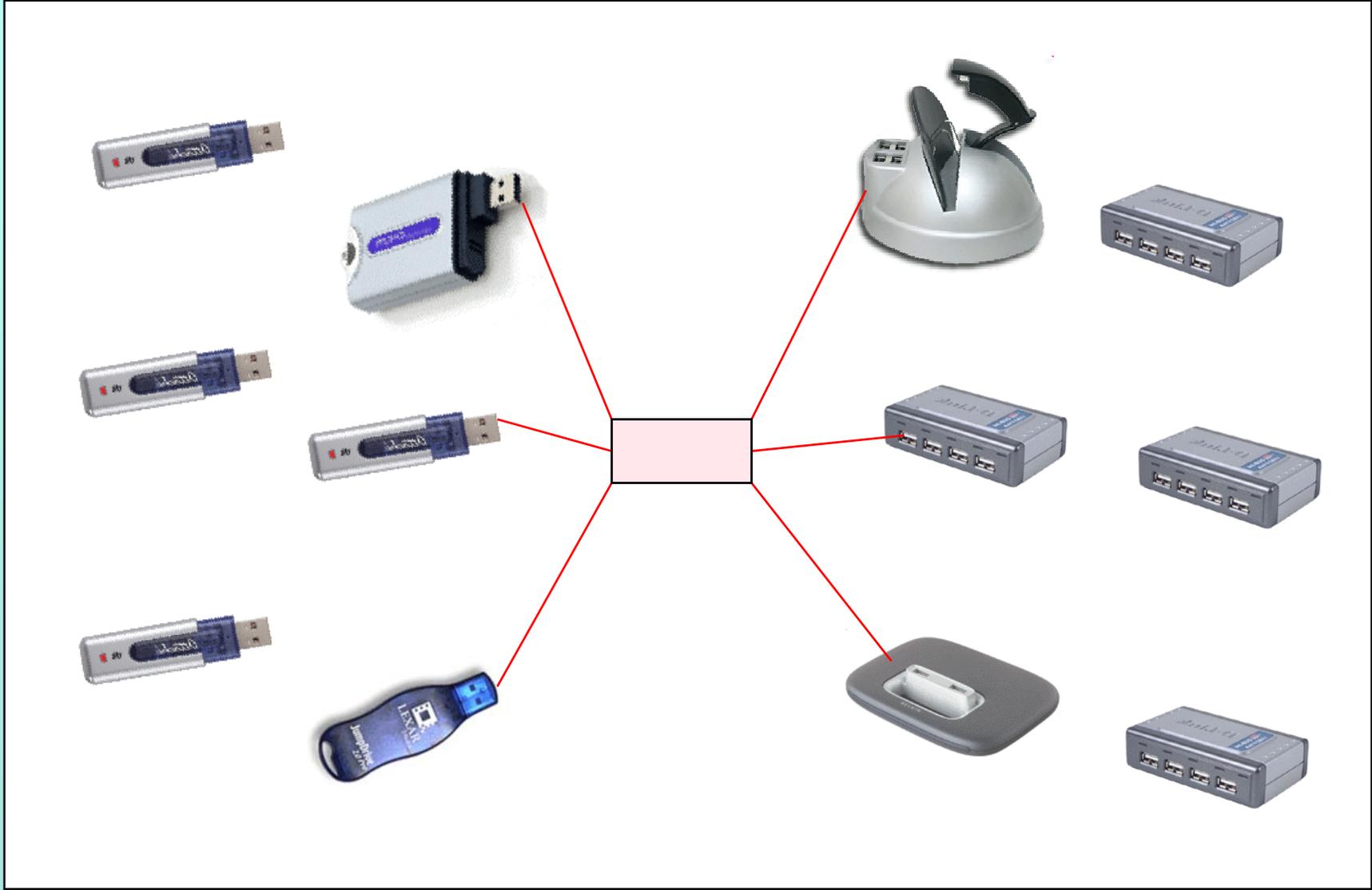# Bad Data-exchange Standard

Good Data-exchange Standard

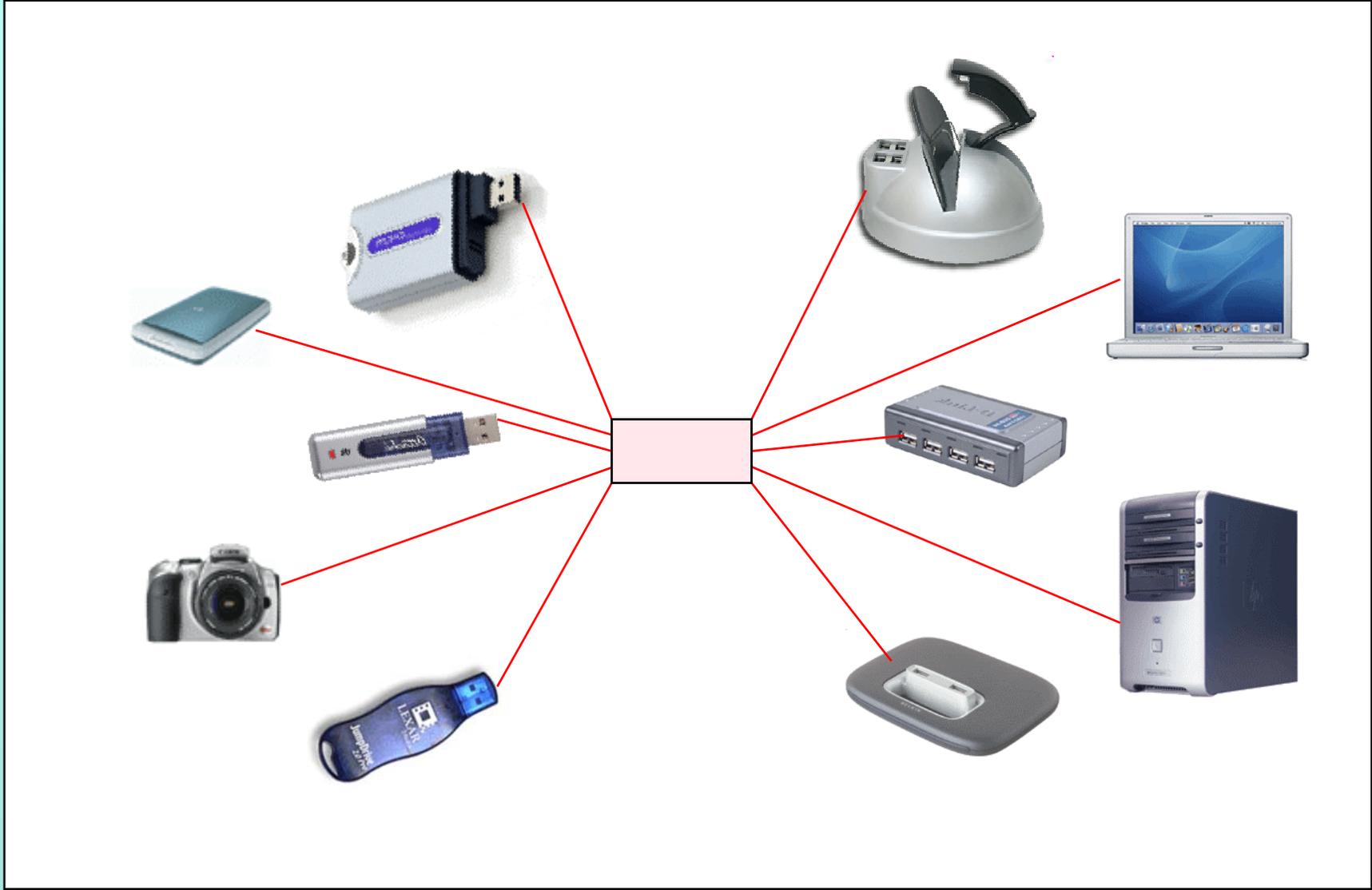# Good Data-exchange Standard

Good Data-exchange Standard

Good Data-exchange Standard

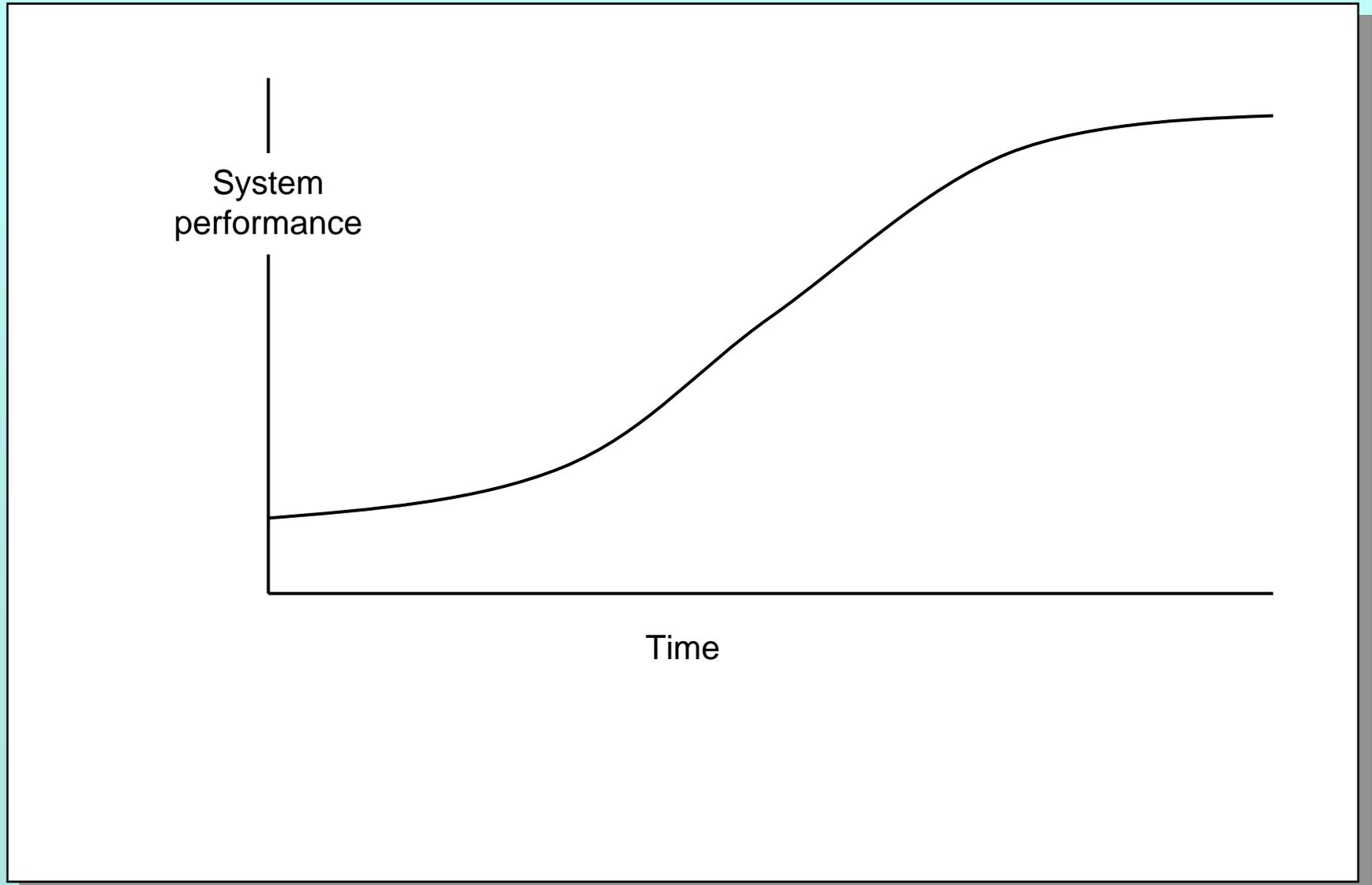# Good Data-exchange Standard

Industry Trends

Industry Trends

System
performance

Time

Industry Trends

System
performance

Level of
performance
needed by
most users

Time

Industry Trends

System performance

Level of performance needed by most users

Transition point where technology satisfies basic needs

Time

Biomedical Research Institutions Information Technology Exchange

Industry Trends

System performance

Level of performance needed by most users

Unmet need

Transition point where technology satisfies basic needs

Time

Industry Trends

Transition point where technology satisfies basic needs

System performance

Level of performance needed by most users

Unmet need

Technology dominates

Time

Industry Trends

System performance

Level of performance needed by most users

Transition point where technology satisfies basic needs

Unmet need

Technology dominates

High Technology

Users want more technology, better performance

Industry Trends

System
performance

Transition point
where technology
satisfies basic needs

Level of
performance
needed by
most users

Unmet need

Excess technology,
most users not
interested in this region.

Technology
dominates

High Technology

Users want
more technology,
better performance

Industry Trends

System performance

Level of performance needed by most users

Transition point where technology satisfies basic needs

Excess technology, most users not interested in this region.

Unmet need

Technology dominates

Technology is "good enough" and therefore irrelevant. user experience dominates

High Technology

Users want more technology, better performance

Industry Trends

Transition point where technology satisfies basic needs

System performance

Level of performance needed by most users

Excess technology, most users not interested in this region.

Unmet need

Technology dominates

Technology is "good enough" and therefore irrelevant. user experience dominates

High Technology

Users want more technology, better performance

Consumer Commodity

Users want convenience, reliability, low cost

Industry Trends

Transition point
where technology

Evolution into the "commodity" space results in a demand for "appliance-like" solutions.

Note that appliances have "use at marginal cost" characteristics.

Lev
perfo
need
most

on.

Users want
more technology,
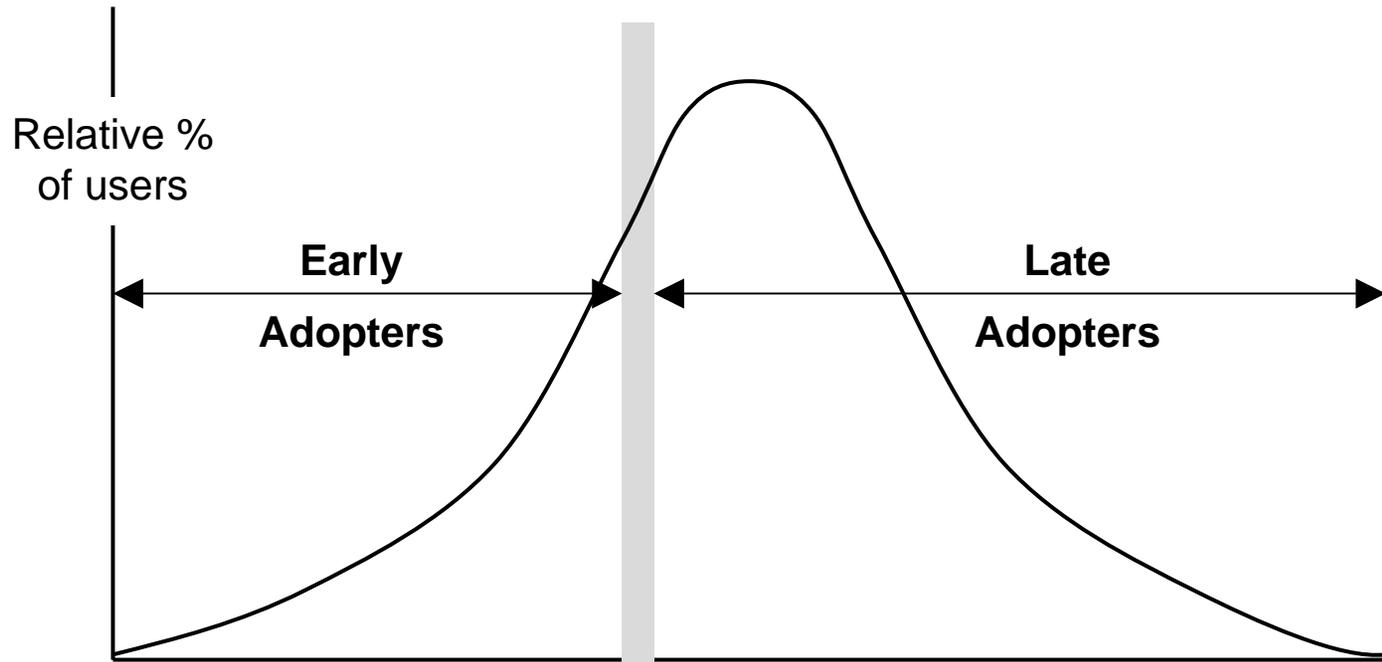better performance

Users want
convenience,
reliability, low cost

Industry Trends

Early adopters drive the technical capabilities of the system, forcing the bar of acceptable performance upward. However, at some point the bar stabilizes and late adopters come to dominate the market for (and hence the design of) technology products.

SUMMARY

The role of BRIITE (& others)

Location of solution components

LABORATORY: QA/QC; basic data management and analysis

INSTITUTION: Shared resources; basic storage & management; statistics and analysis support; digital publishing support;

RES. COMMUNITY: Information appliances; public data collections; analytical support

FUNDING AGENCY: Core grant support; caBIG; BISTI

GLOBAL: Identity management; authentication, authorization, auditing

Slides:

http://www.esp.org/rjr/briite-04.pdf

END