Information Technology Vision for Data Synthesis

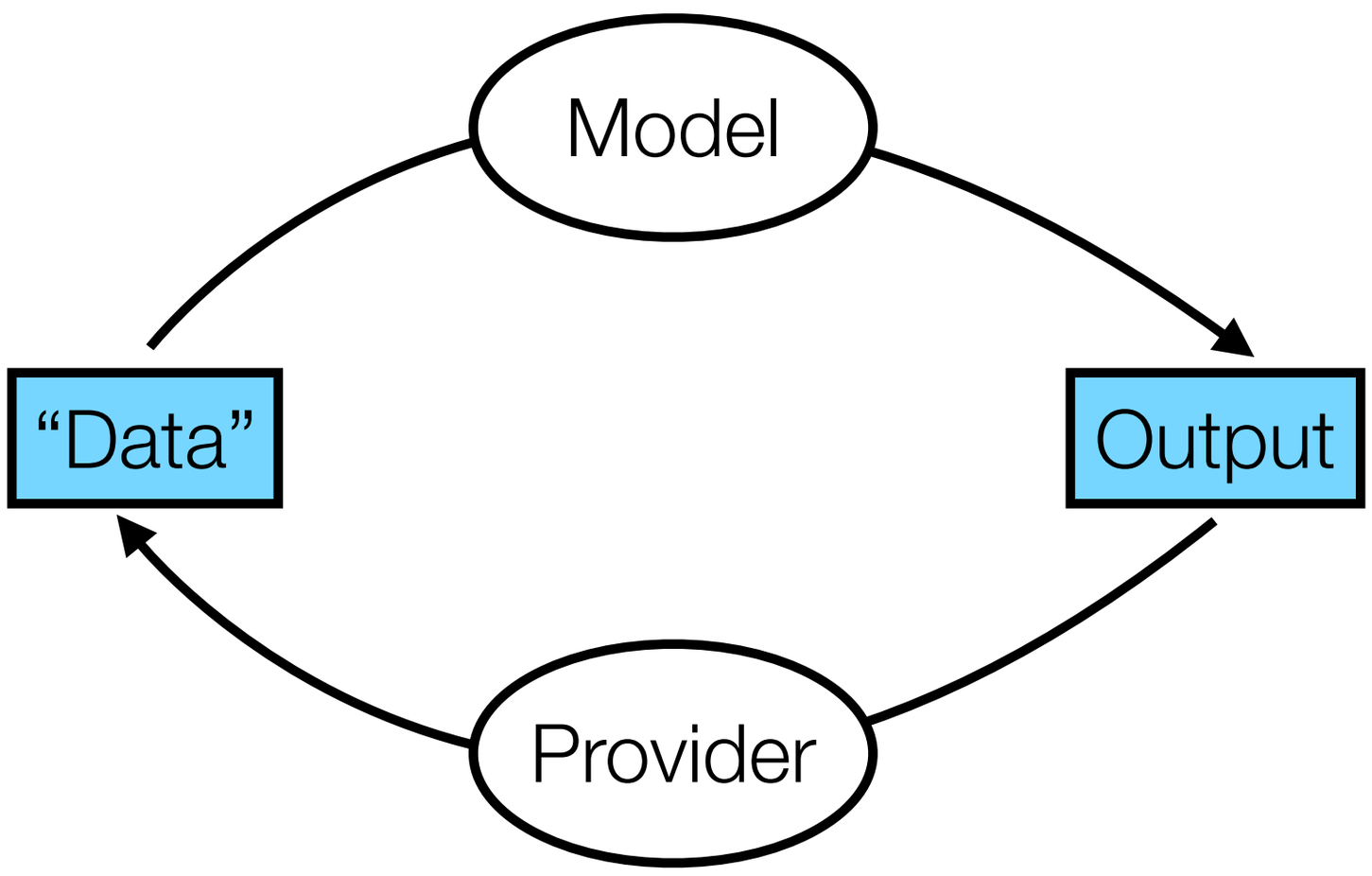
Mark A. Parsons

National Snow and Ice Data Center

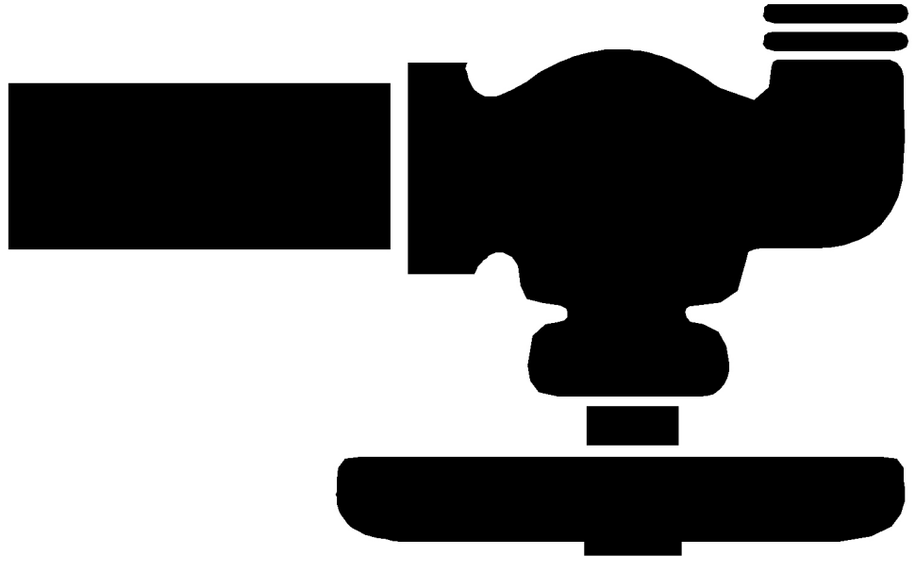
Data Management?

~~Information Technology~~ Vision for Data Synthesis

Mark A. Parsons


Model
“Data”
Output
Provider

1 1
0 0
1 0
1 1
1 0
1 1
0 1
1 0
0 0
0 1
0 1
1 1
1 0



What is a Utility?

- Simple
- Predictable
- Reliable
- Extensible
- Accessible, i.e. usable
- Durable

National Snow and Ice Data Center

What is a Utility?

- Simple
- Predictable
- Reliable
- Extensible
- Accessible, i.e. usable
- Durable

It is infrastructure

National Snow and Ice Data Center

What does this mean?

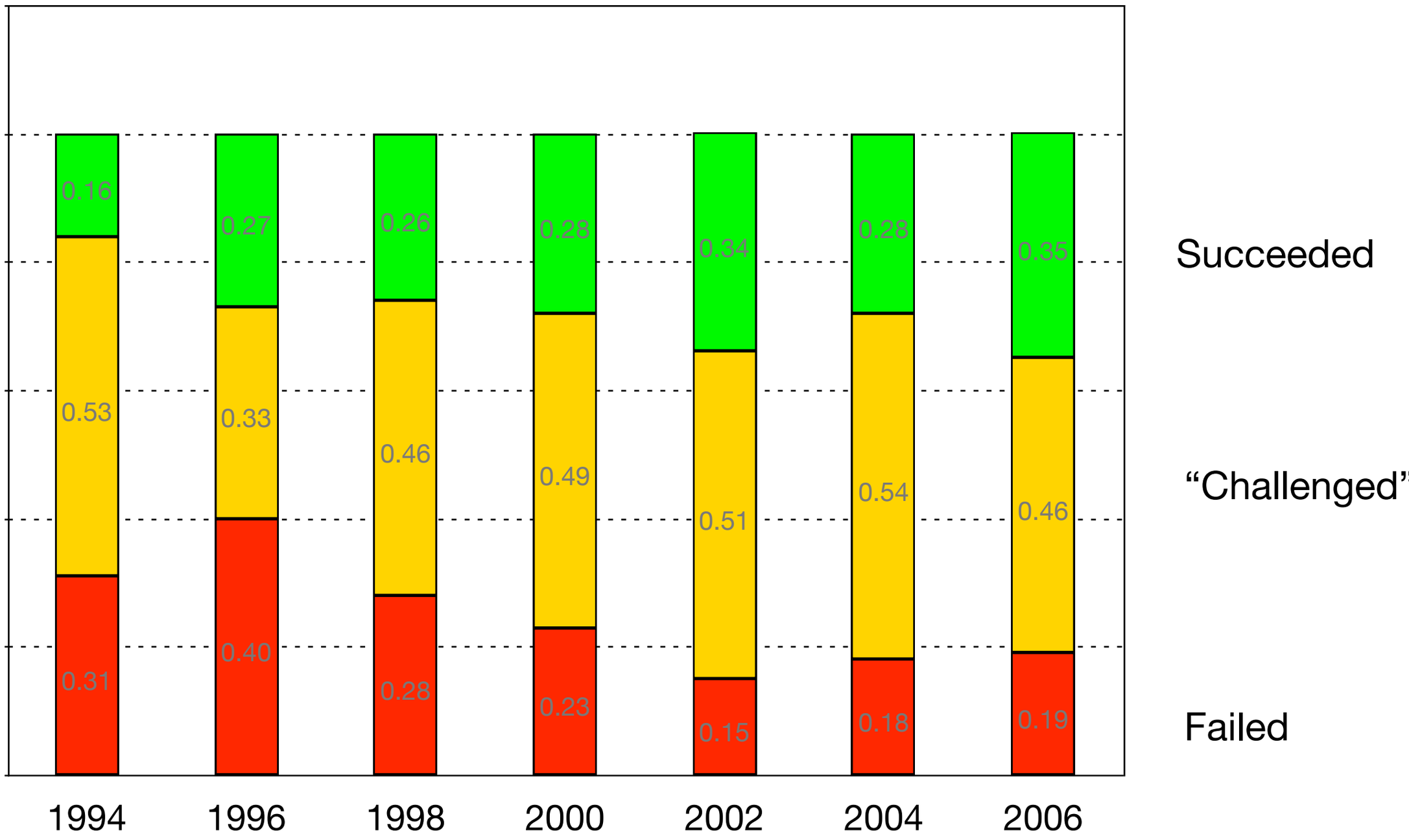
It can guide our thinking about:

- Interface design
- Interoperability
- Software design
- Cost models
- Data preservation
- Distributed vs. centralized data management
- Technological choices

NSIDC
National Snow and Ice Data Center

Systems and Innovation

> "We're entering a new world in which data may be more important than software."
>
> - Tim O'Reilly



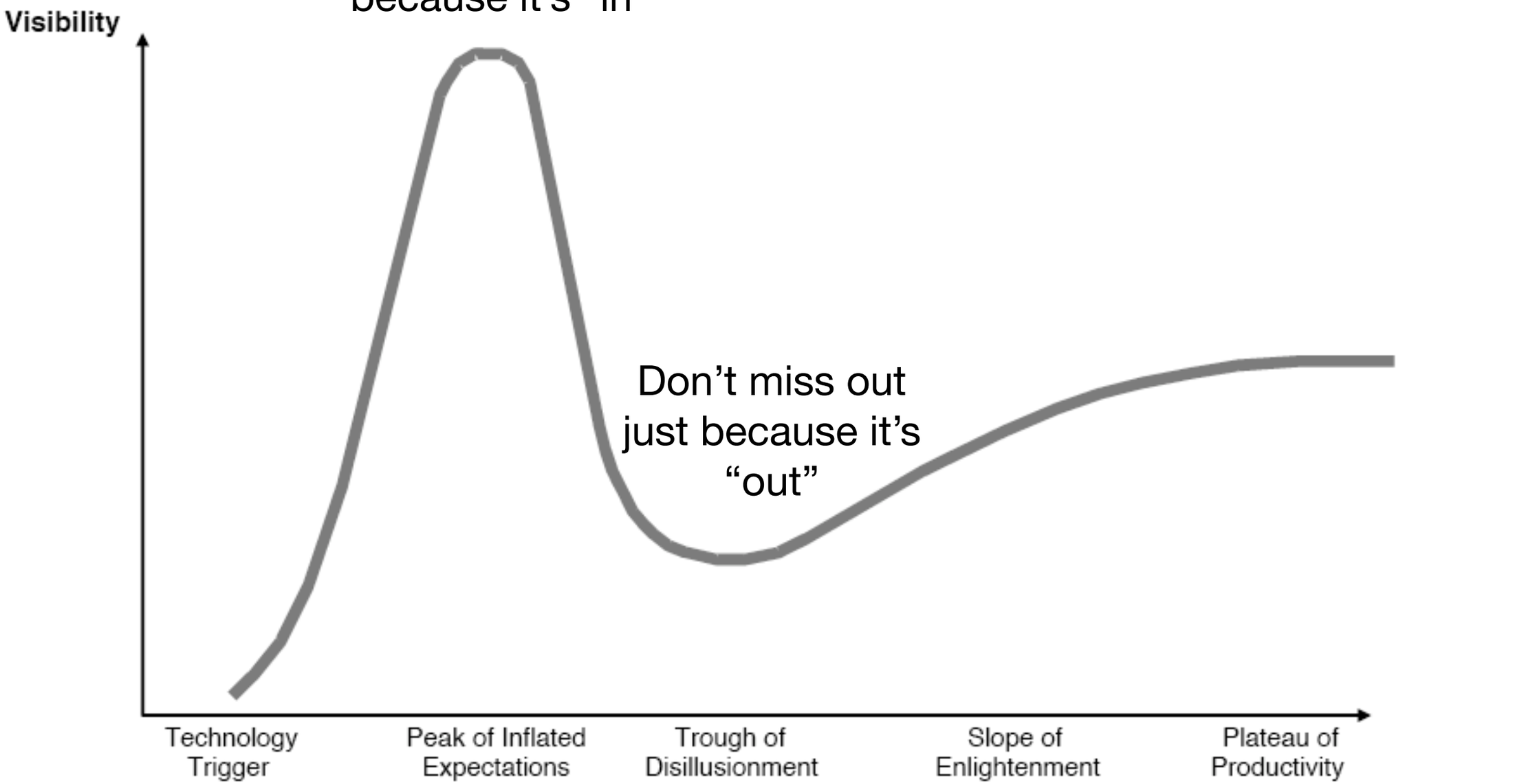
The Standish Group's "CHAOS report". An assessment of over 40,000 IT application projects

The Hype Cycle

Gartner

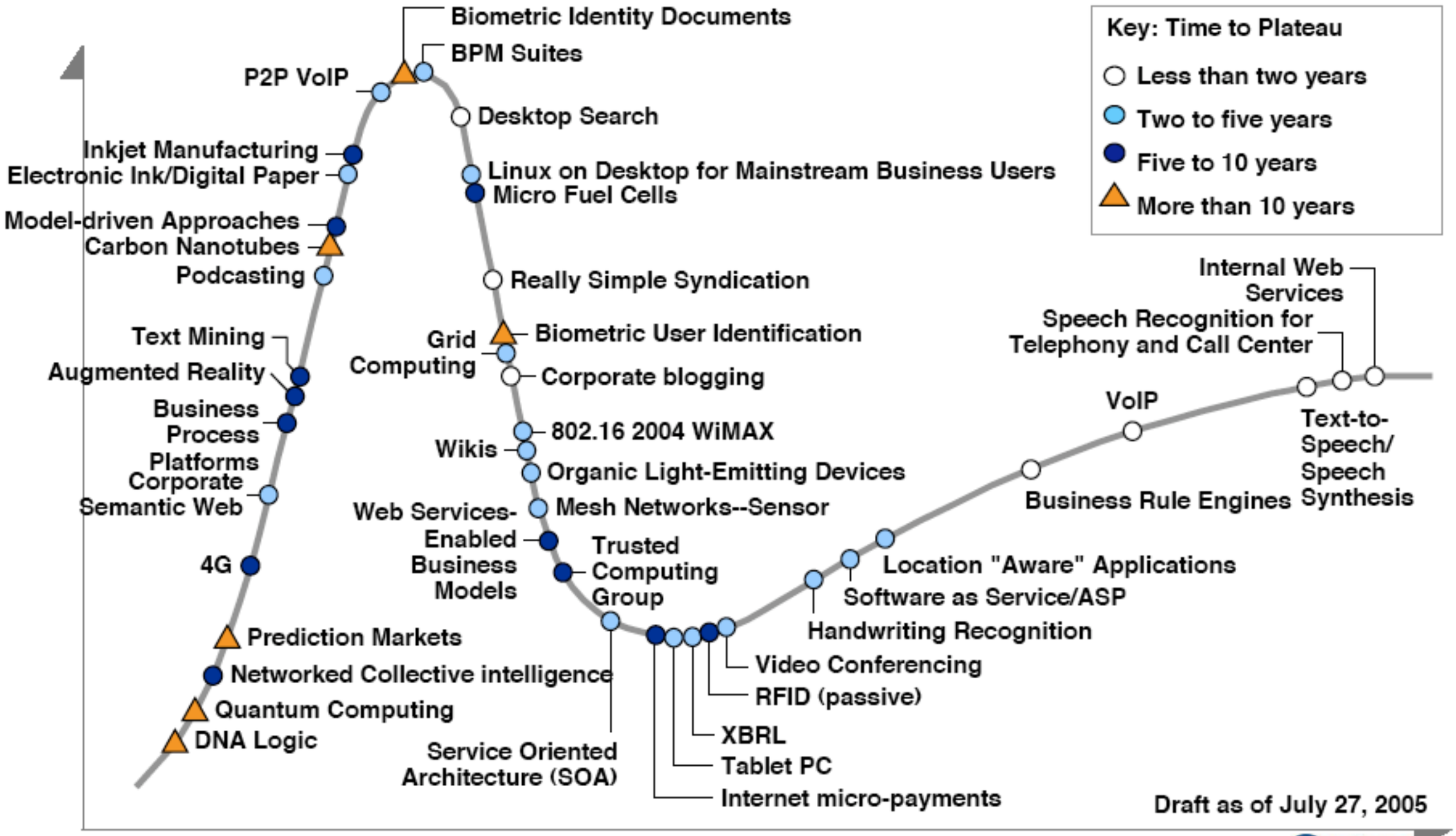
The Hype Cycle

Don’t join in just because it’s “in”



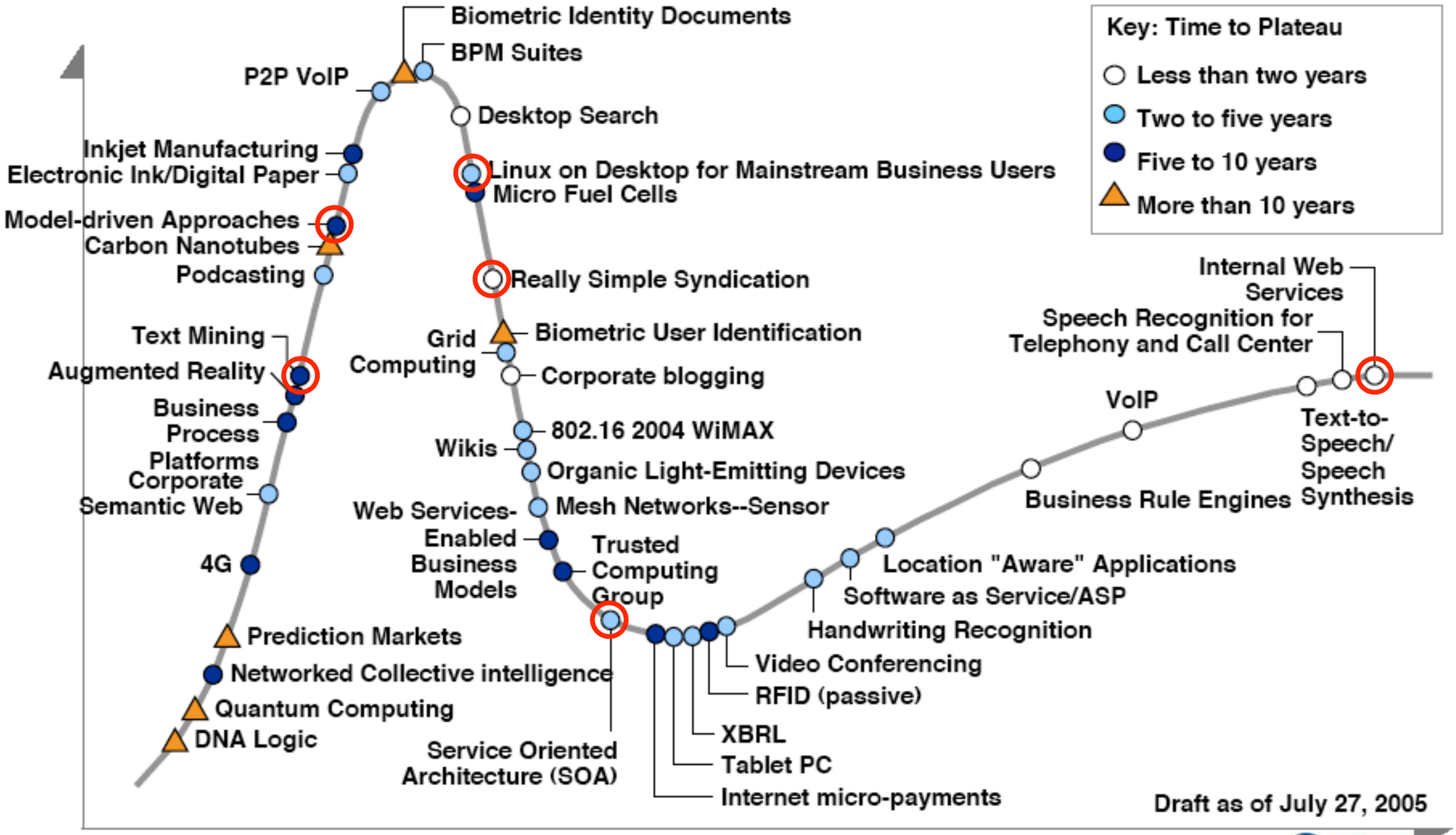
Gartner

Emerging Technologies Hype Cycle 2005



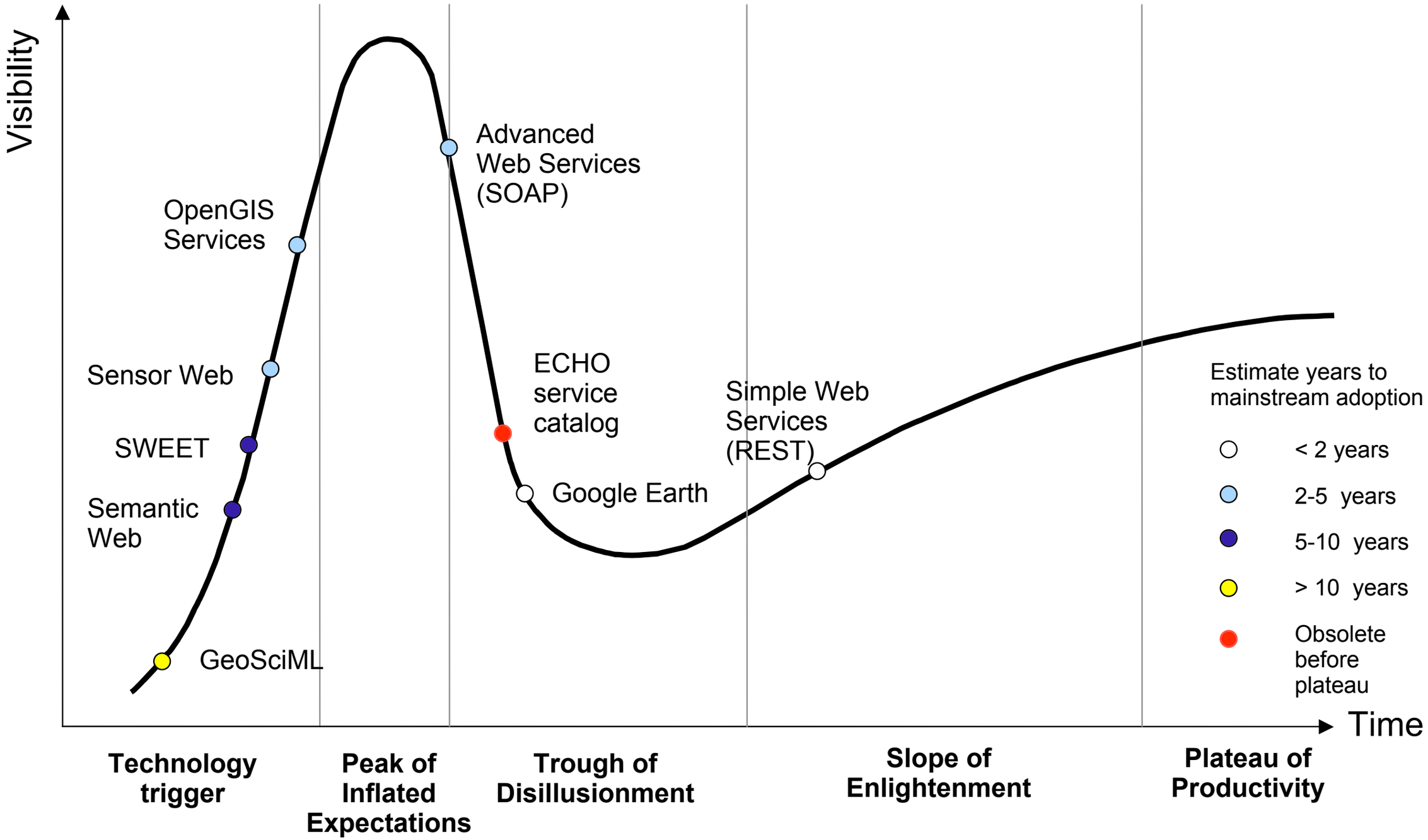
Gartner

Emerging Technologies Hype Cycle 2005



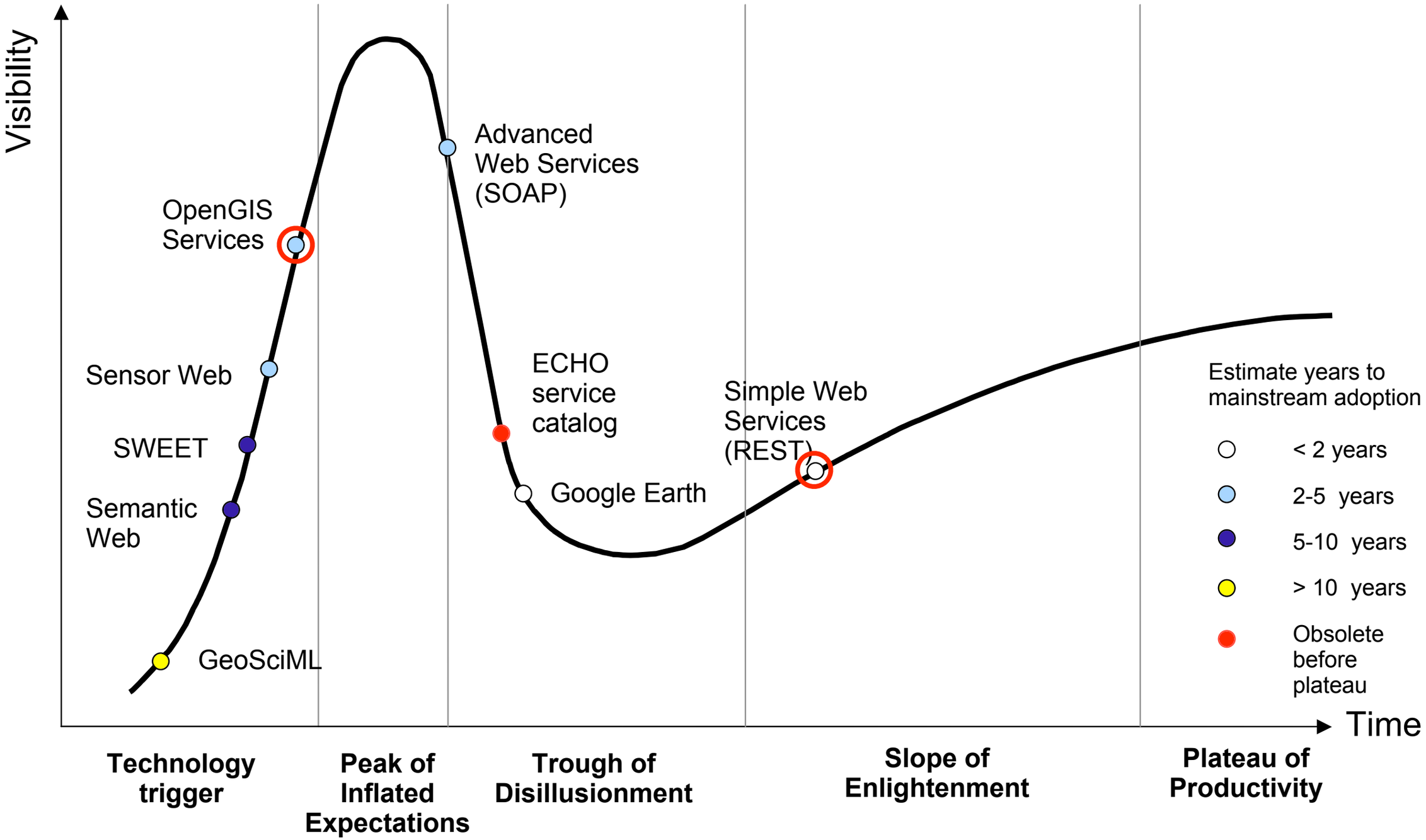
Gartner

2007 Hype Cycle for Earth Science Technologies



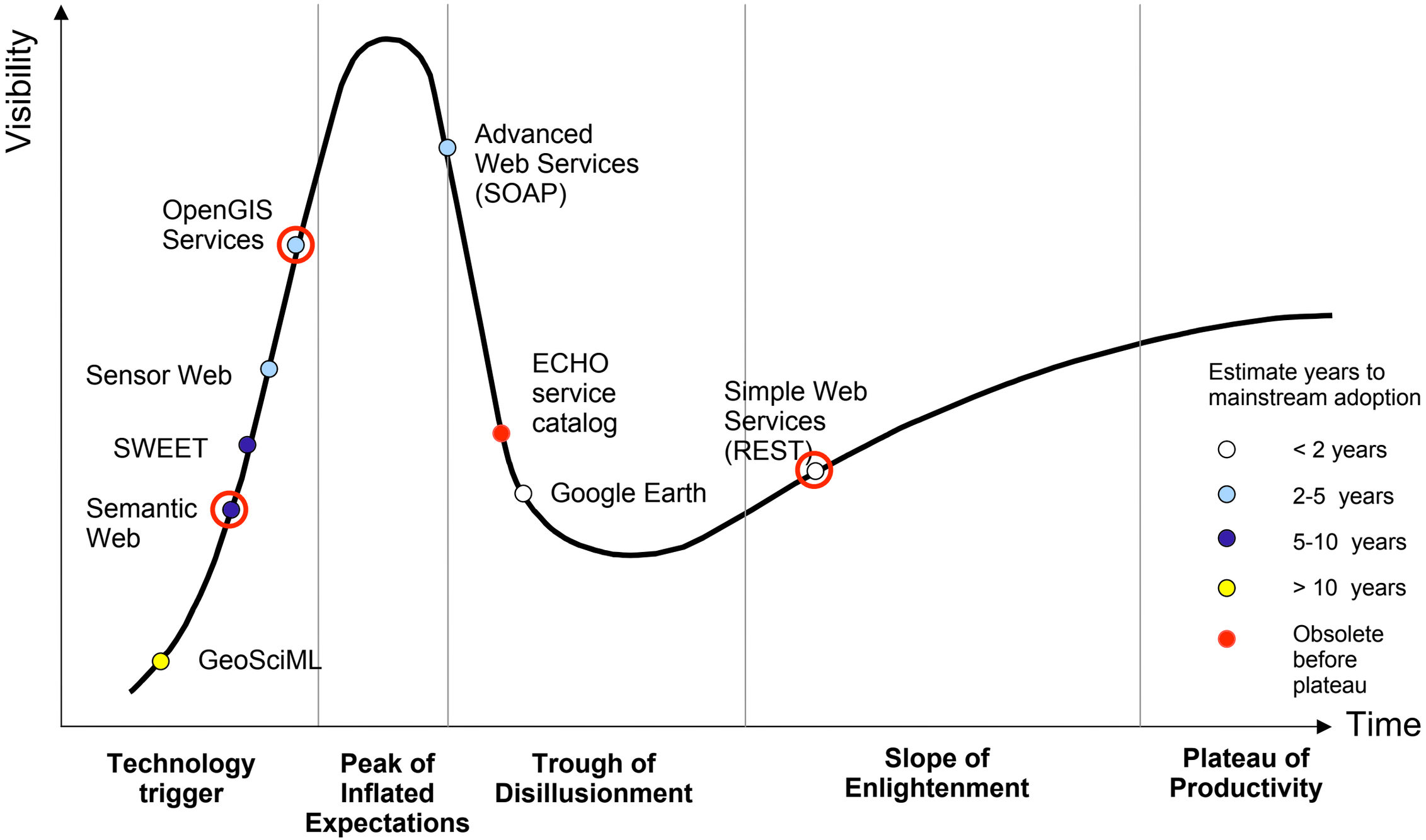
courtesy NASA Technology Infusion Working Group

2007 Hype Cycle for Earth Science Technologies



courtesy NASA Technology Infusion Working Group

2007 Hype Cycle for Earth Science Technologies



courtesy NASA Technology Infusion Working Group

O'REILLY radar

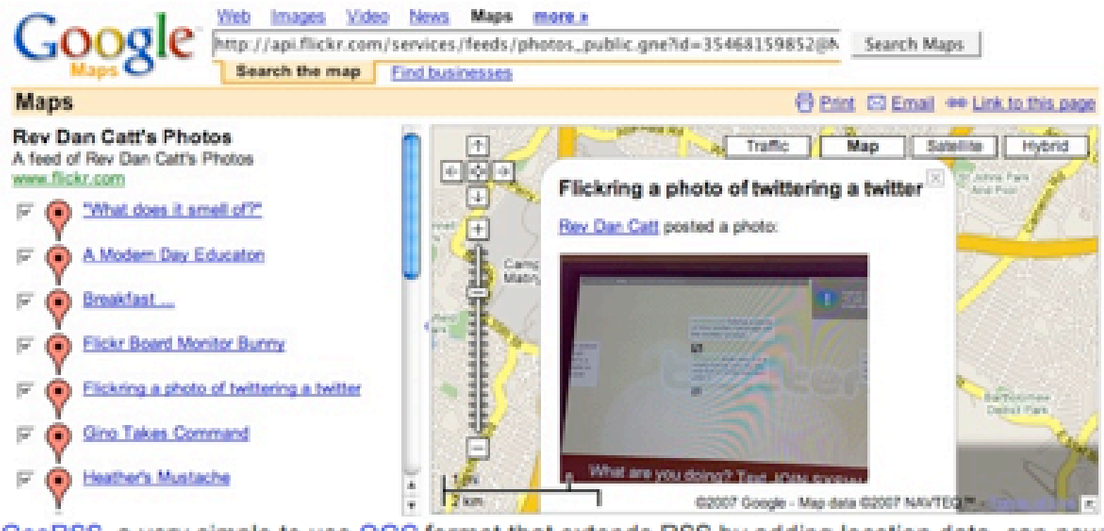
Geo

Mapping and location data are driving some of the most intriguing new web applications. Geohacker alpha geeks are building wickedly clever mashups. Established companies are integrating location data into all manner of workaday applications. And the location industry is growing into the Web 2.0 era. We're tracking it, and showcasing the most significant work at Where 2.0.

Brady Forrest

Thu
Mar 22
2007

Google is Supporting GeoRSS



GeoRSS, a very simple to use OGC format that extends RSS by adding location data, can now be consumed via Google Maps and the Google maps API. Congrats to Mikel and the OGC! As Google stated on the Google Maps API blog:

> To start we now support GeoRSS as a data format for geographic content in Google Maps. We want to enable users to create data in whatever format is most convenient for them, and feel that by supporting both KML and GeoRSS we can enable a wider variety of people and applications to contribute content to Google Maps. We've built support for the Simple, GML, and W3C Geo encodings of GeoRSS -- all you have to do is enter the full URL of a GeoRSS file into the Maps query box to load the file. For example, take a look at SlashGeo's GeoRSS on Google Maps.
>
> Most importantly, we've extended support for displaying geographic data -- both KML and GeoRSS -- into the Google Maps API. Now in addition to programatically adding content to a Maps API site, you can create your content as KML or GeoRSS and load it into the Map with a simple function call. This means that the more than 1 million KML files that are available from all over the web can easily be mashed up with the map on your site.

Additionally, KML is on its way to becoming an OGC standard (and as you can see from this

Radar RSS feed

MOST ACTIVE | MOST RECENT

- The Future of Web 2.0
- Support for Kathy
- Don't Miss Nat's Last Post
- From Subprime Loans to Failing Newspapers
- SF Chronicle in Trouble?

RADAR TEAM

REPORTS

Web 2.0 Principles and Best Practices
What does Web 2.0 mean for your company? Get the latest on the why, what, who, and how of Web 2.0 in this report.
Read more

CONFERENCE

MySQL Conference & Expo

Scale to new heights at MySQL.
Read more

Best Practices

- Openness and collaboration
- Use standards including standard tools and infrastructure
- Efficient, iterative development—bite size pieces
- Documentation (and maybe metadata)
- Hardware/Technology independence
- Design for durability
- Use professionals
- User driven

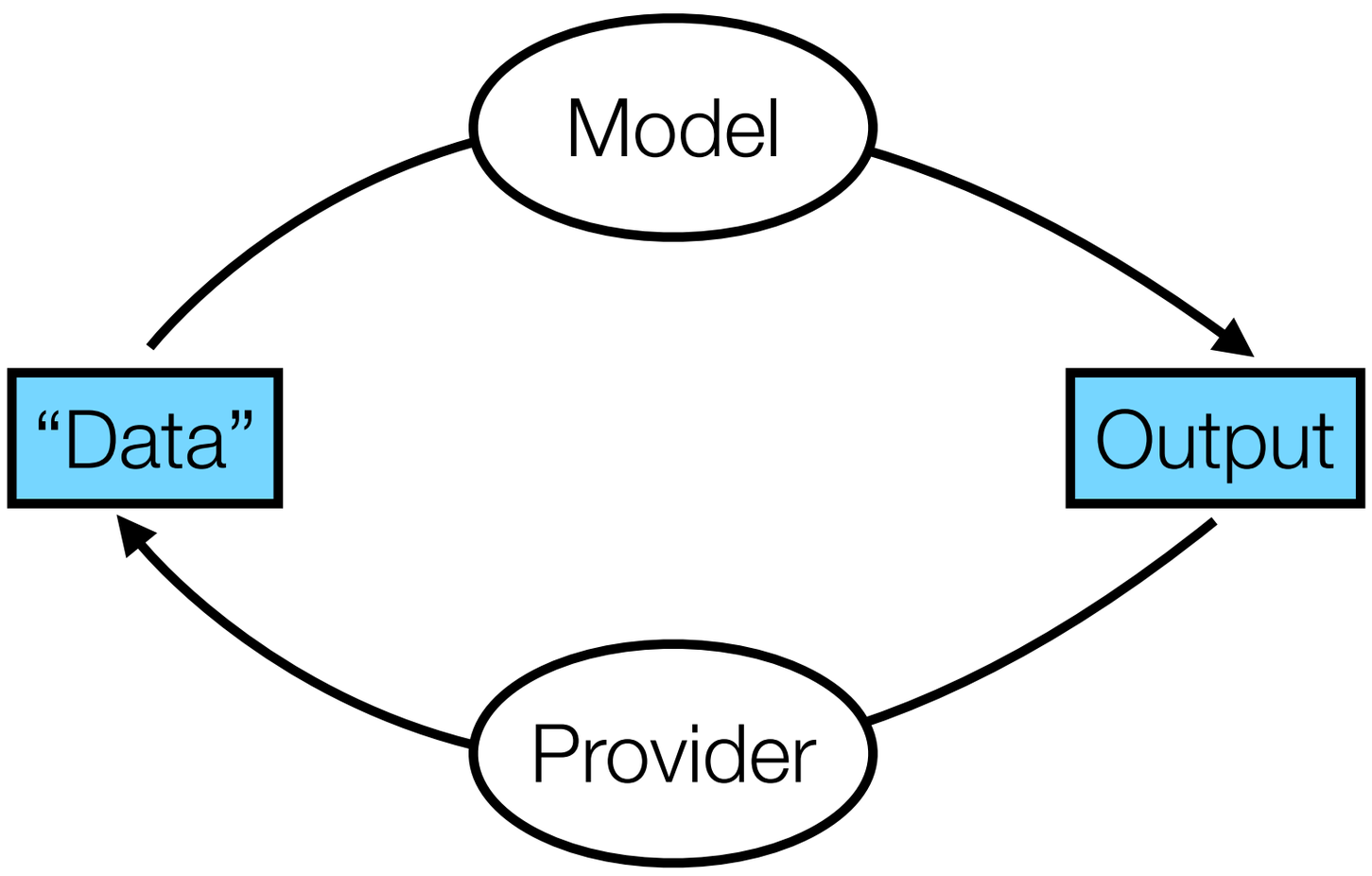
NSIDC
National Snow and Ice Data Center

Best Practices

Simple
Predictable
Reliable
Extensible
Accessible
Durable

- Openness and collaboration
- Use standards including standard tools and infrastructure
- Efficient, iterative development—bite size pieces
- Documentation (and maybe metadata)
- Hardware/Technology independence
- Design for durability
- Use professionals
- User driven

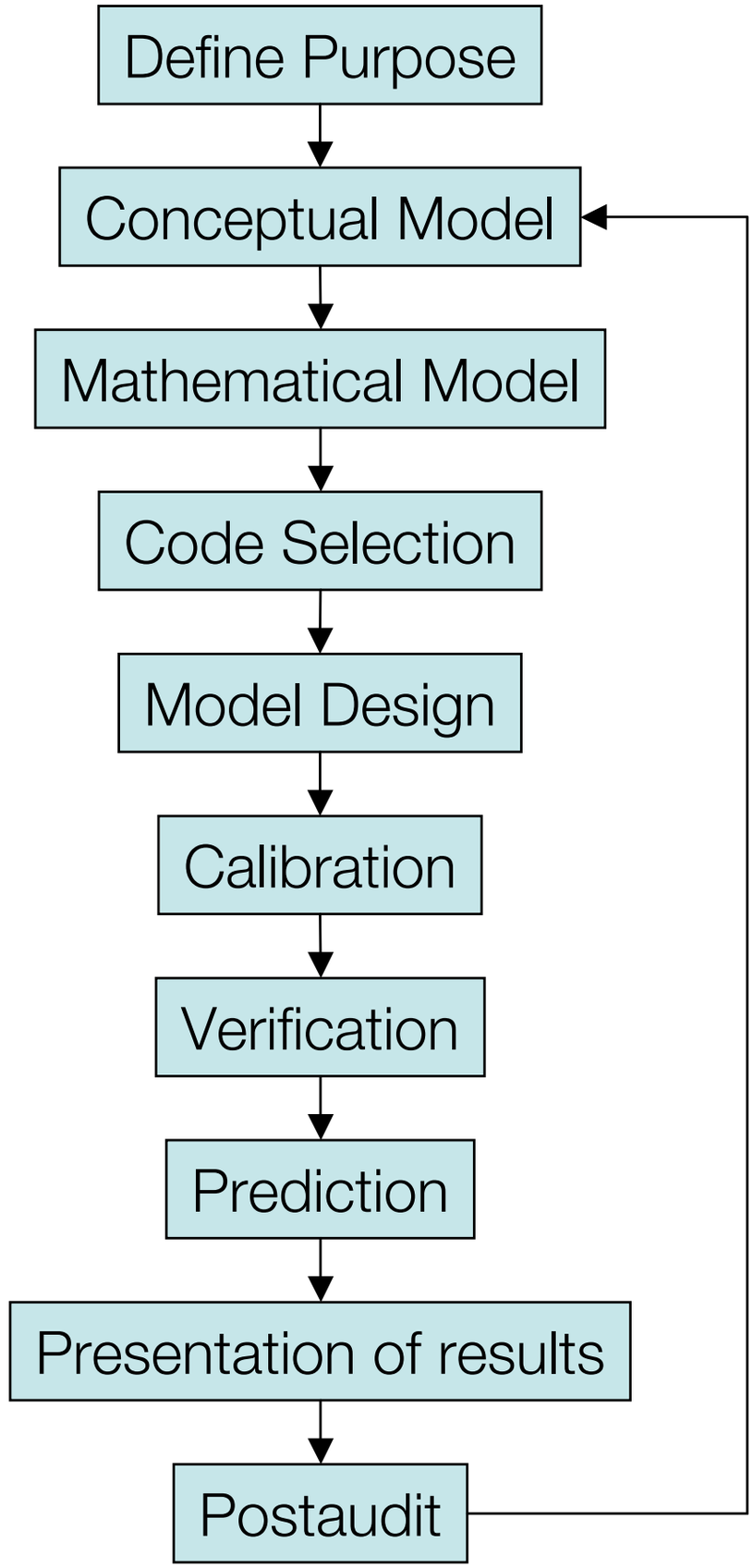
NSIDC
National Snow and Ice Data Center


Model
“Data”
Output
Provider

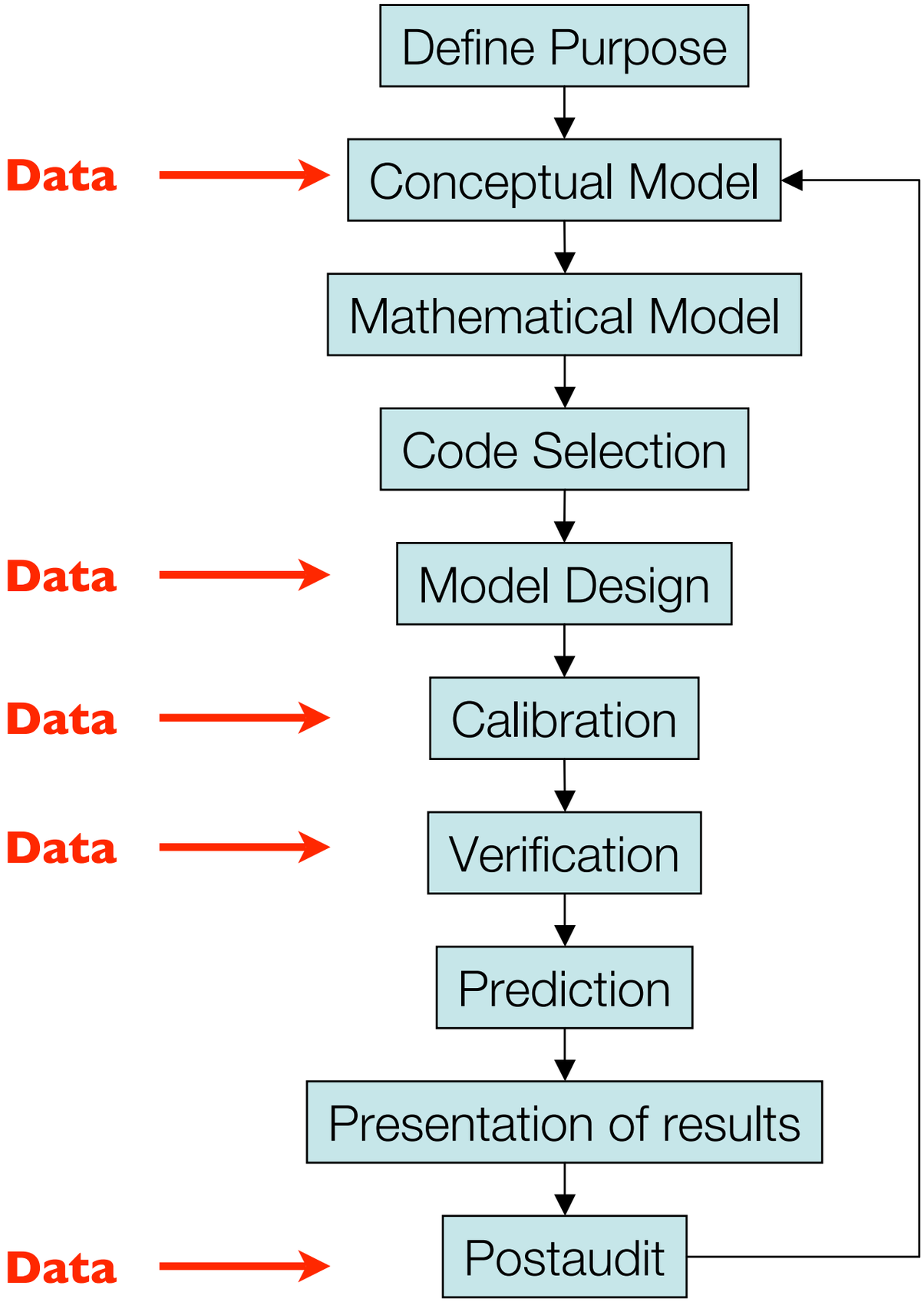
Two (Over-Simplified) Worldviews

(borrowing from Ben Domenico & Stefano Nativi)

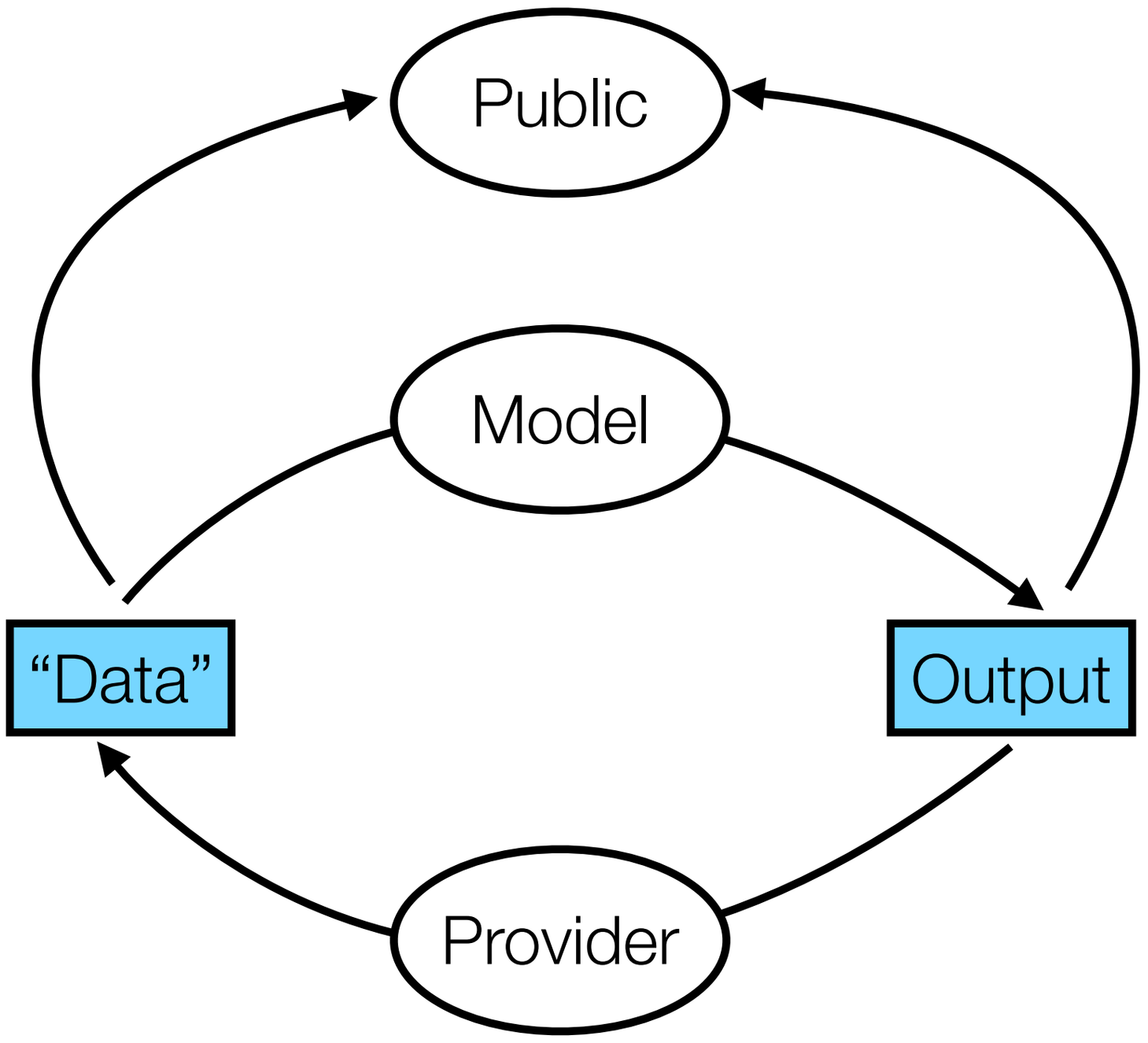
- **To the GIS community, the world is:**
  - ✓ A collection of **features** (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface).
  - ✓ The **features** are **discrete objects** described by a set of (typically 2-D) characteristics such as a **shape/geometry**
- **To fluid-earth scientists, the world is:**
  - ✓ A set of observations/measurements described by **parameters** (e.g., temperature, velocity) that vary as **continuous functions** in (4-D) space-time
  - ✓ Parameter behaviors are governed by a set of **equations**.



from Anderson and Woessner 1995



from Anderson and Woessner 1995


Public
Model
"Data"
Output
Provider

Possible approaches

- Establish communities of practice around “cycles” disciplines, model types
- A virtual modeling observatory
- Identify benchmark data sets or base maps and gaps for key parameters, and make them broadly usable
- Case studies
    - How do data providers use model output to plan new field data collections
    - How do knowledge brokers discover, access, and use data
    - Look at different models and study how modelers assess, acquire, and prepare data

National Snow and Ice Data Center

Create uses cases to identify issues

- Assess
    - context, semantics, metadata, uncertainty...
- Acquire
    - spatial and temporal requirements, timeliness, local or remote analysis, connection to tools...
- Prepare
    - formats, compilation, semantics, grids, projections, scales...
- Present
    - outreach, connection to other models and data collection efforts, value-added/ compiled products, downscaling and upscaling...

National Snow and Ice Data Center

Thank you

National Snow and Ice Data Center