

情報爆発時代における新しいIT研究:

超大量多様情報からの価値創出 (情報爆発プロジェクト / 情報大航海プロジェクト)

喜連川優

東大 生研 戦略情報融合研究センタ

※: このマークが付してある著作物は、第三者が有する著作物ですので、
同著作物の再使用、同著作物の二次的著作物の創作等については、
著作権者より直接使用許諾を得る必要があります。
引用情報のない図版は、著作権フリーなもの、
あるいは講演者の有する著作物の中から引用されたものです。

この場所にあった図版は、
著作権処理の都合上、
削除いたしました。
ご了承ください。

知識の爆発！

1900年の人間の知識



この場所にあった図版は、
著作権処理の都合上、
削除いたしました。
ご了承ください。

5

New IT Infrastructure for the Information Explosion Era
Research Project funded by MEXT Grant-in-Aid for
Scientific Research in Priority Areas (FY 2005–2010)

info-plosion

情報爆発時代に向けた新しいIT基盤技術の研究

平成17年発足 文部科学省特定領域研究

喜連川優

東京大学 生産技術研究所



†
自由国民社



†
自由国民社

情報爆発時代の到来

この場所にあった図版は、
著作権処理の都合上、
削除いたしました。
ご了承ください。

info-plosion

988EB = 1ZB
(2010) ゼッタ
161EB バイト
(2006 by IDC07)

情報爆

billion GB

billion GB

5 billion GB

billion GB

Source: Horizon Information Strategies,
cited from Storage New Game New Rules, p.34
(www.horizon.com)

2006年における生成量は過去全ての本の 300万倍 6トンの本/一人、2019年には26トン/人

全人類が情報発信

特定領域研究
情報爆発



72 億ウェブサーチ/月
>> 世界総人口

しかし一人の受信能力は一定

「探す」ことに明け暮れる

特定領域研究
情報爆発



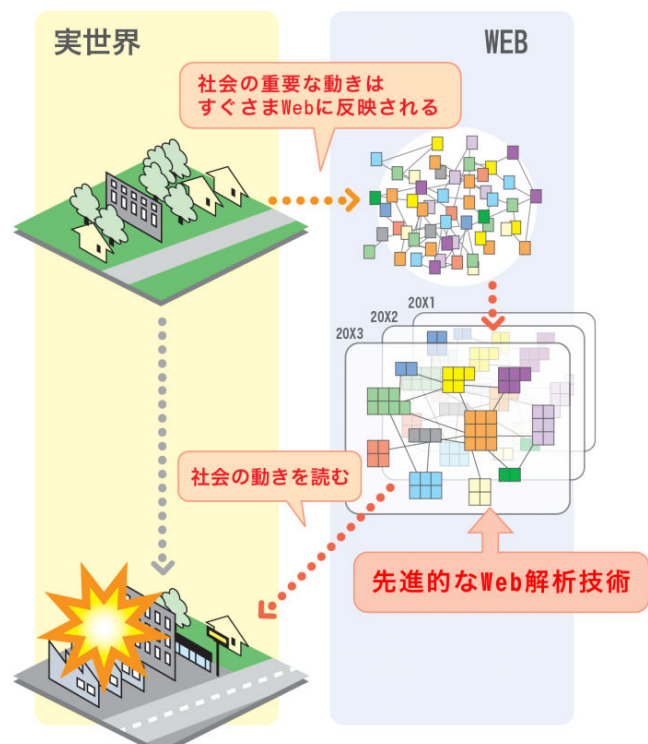
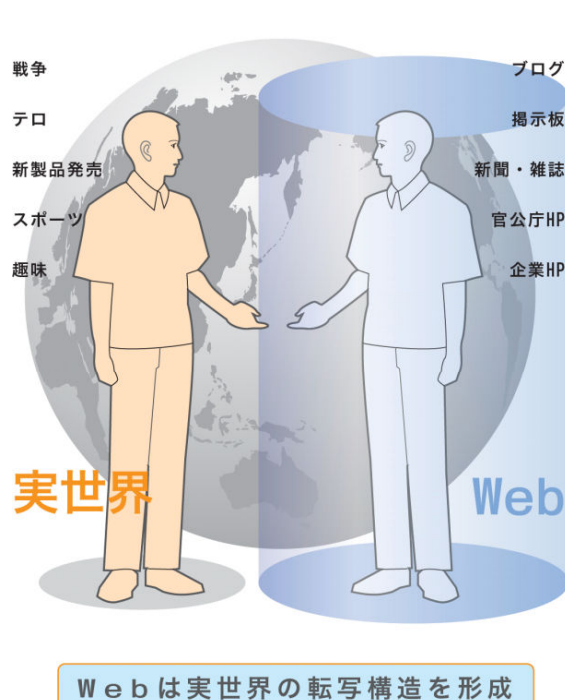
Plenty of information
MSN's own research found **50% of complex queries go unanswered**, while even more simple searches take an average of 11 minutes. Despite this it found that there was real hunger for web-based answers. **Three-quarters of people would rather use the web to answer their questions than their own family members...**

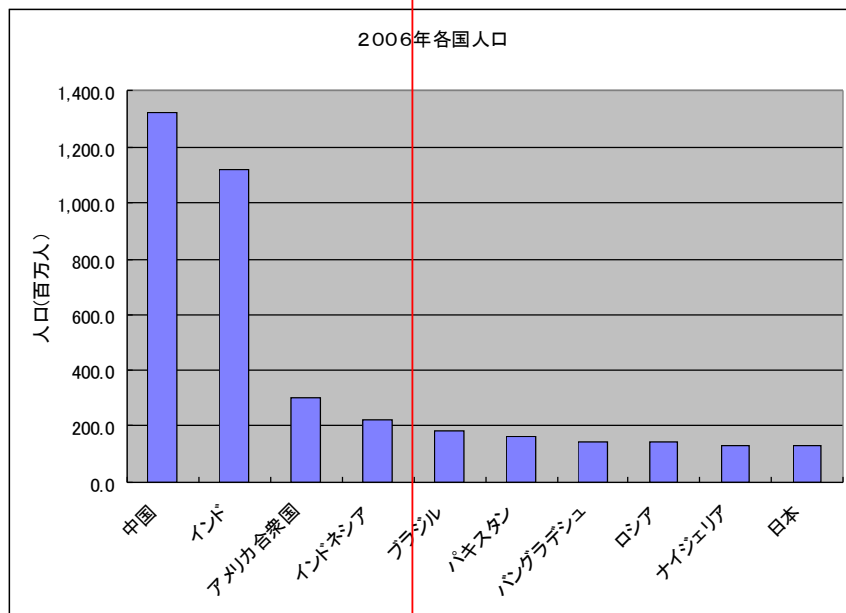
複雑な質問の50%は答えが見つけれない
75%の人々が家族よりもWebに尋ねる

全く新しい「すばやく探し出す手法」が必要

実社会の射影としてのウェブ

e-Society
文部科学省リーディングプロジェクト



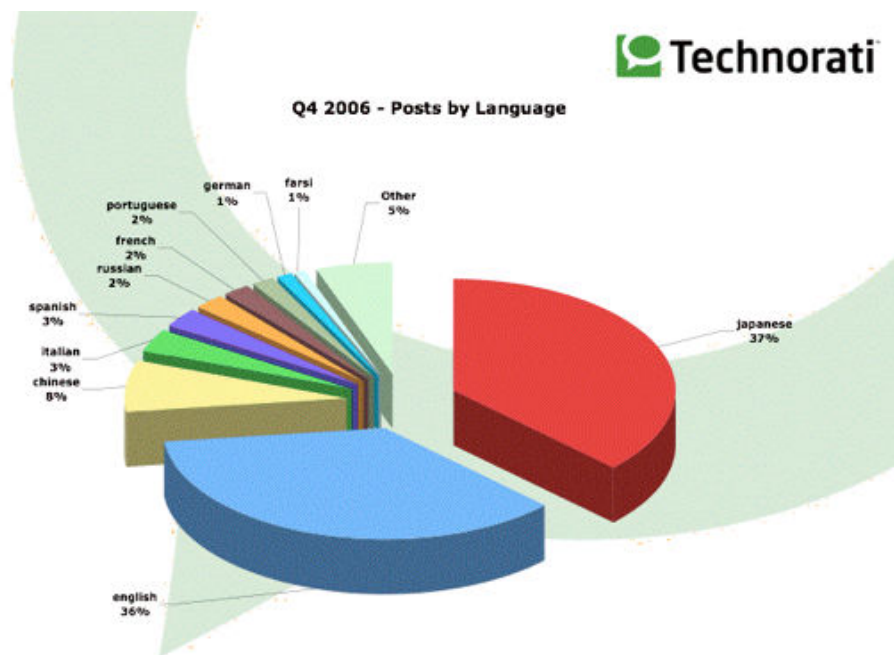


国際連合統計部による各掲載年の7月1日現在の推計人口



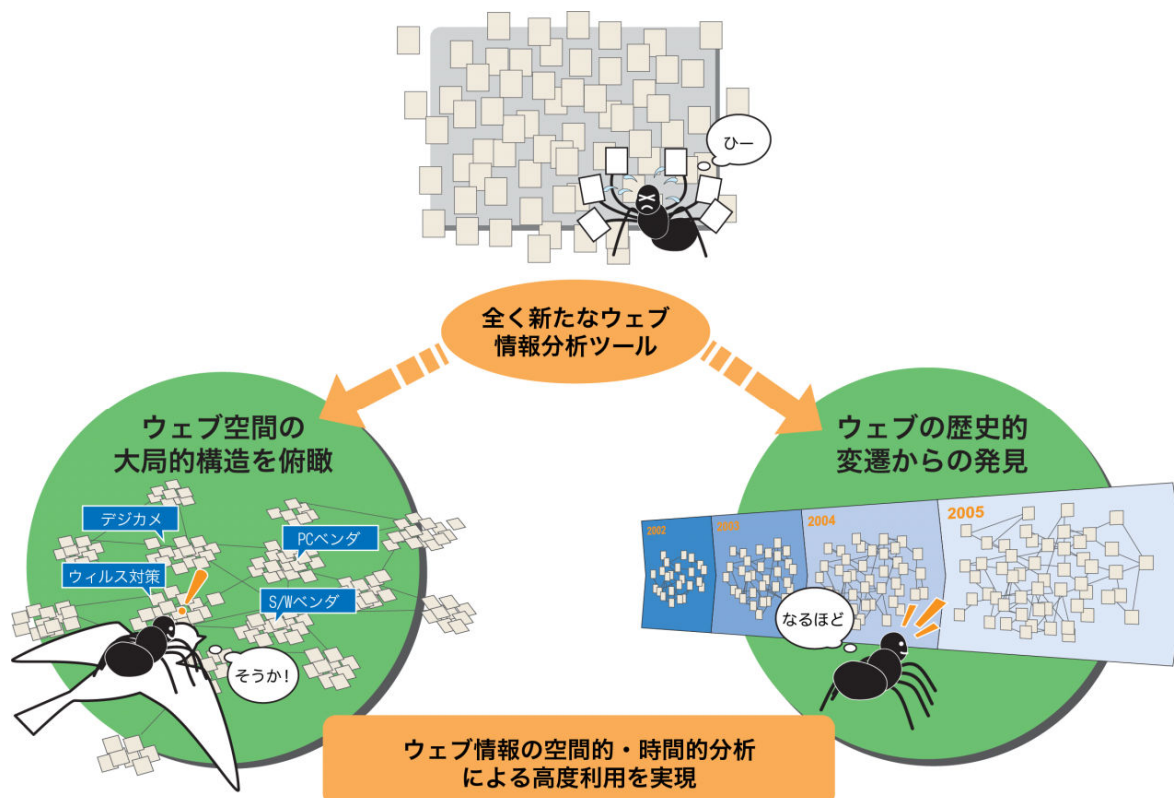
日本人はけっこう しゃべりたがり屋？

特定領域研究
情報爆発



目的: ウェブ情報の高度利用システム の構築 (WEBの時空間解析)

e-Society
文部科学省リーディングプロジェクト



成果①ウェブ空間の構造俯瞰

～ウェブ全空間の地図化～

リンク&テキスト解析を用いてウェブの全空間を地図化

産業連関図に相当する地図が得られる

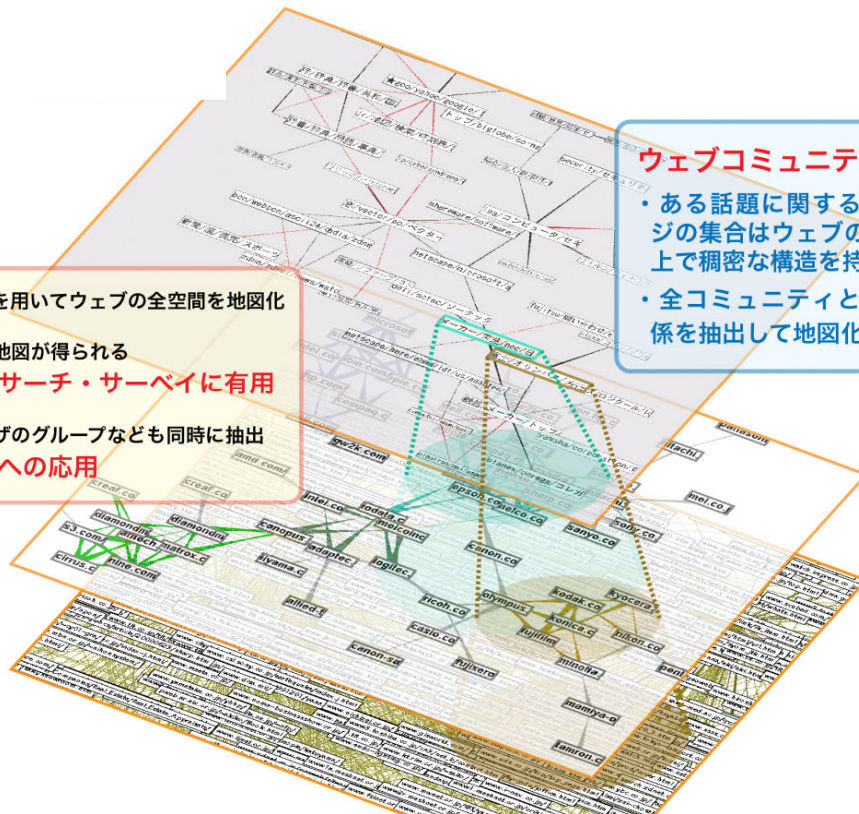
→ 注目分野のリーサー・サーベイに有用

影響力のある製品ユーザのグループなども同時に抽出

→ 広告設置戦略への応用

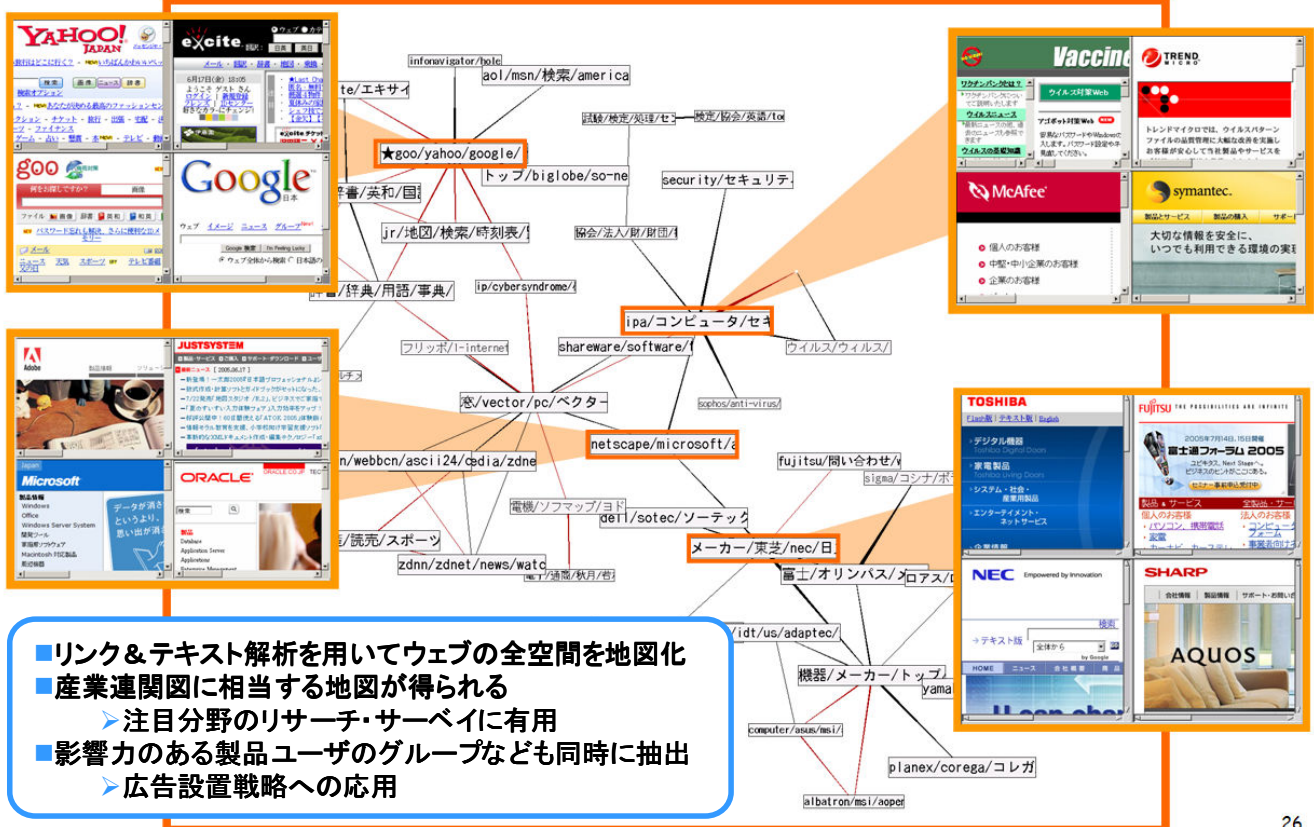
ウェブコミュニティチャート

- ある話題に関する有用なページの集合はウェブのリンク空間上で稠密な構造を持つ
- 全コミュニティとそれらの関係を抽出して地図化



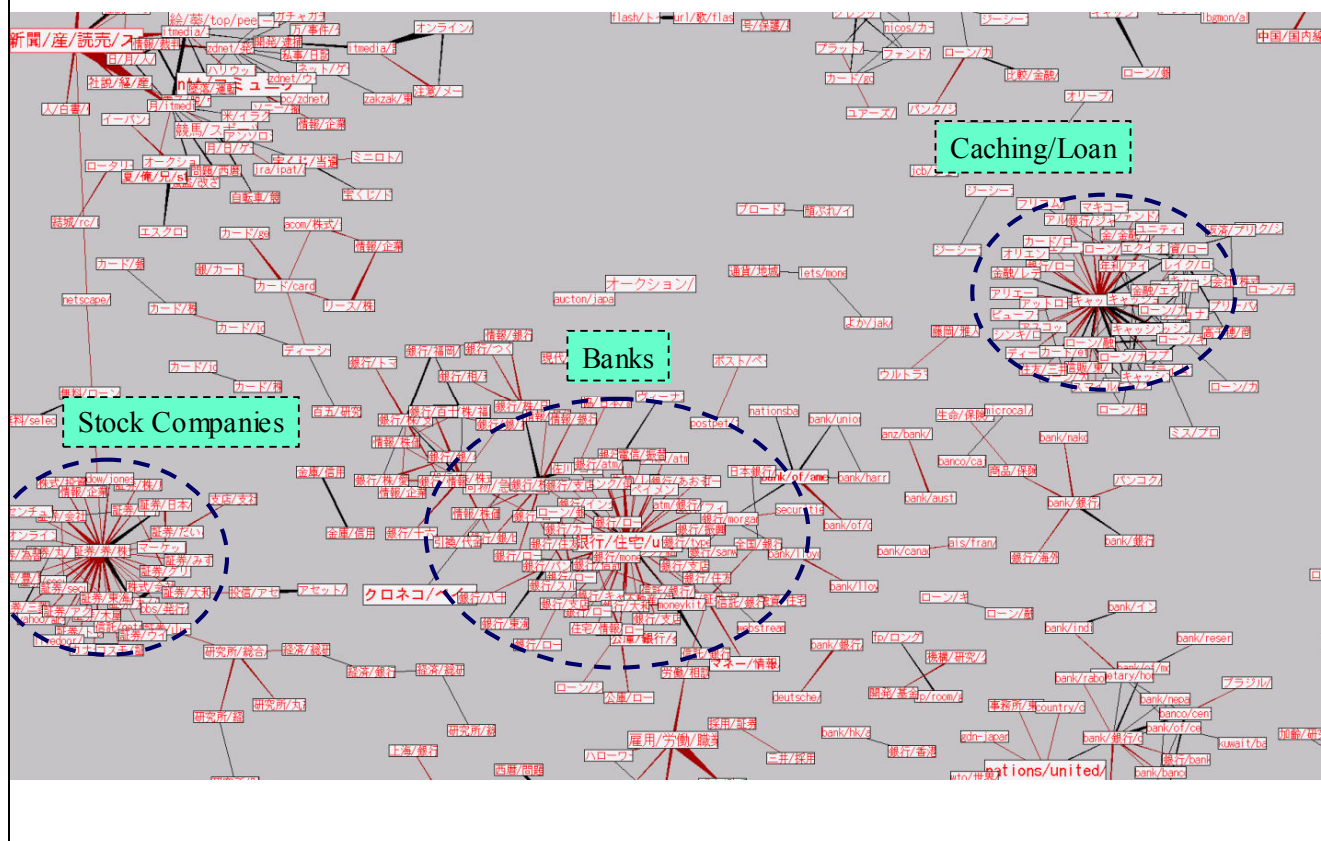
ウェブ空間の構造俯瞰

～コンピューター業界周辺の地図～



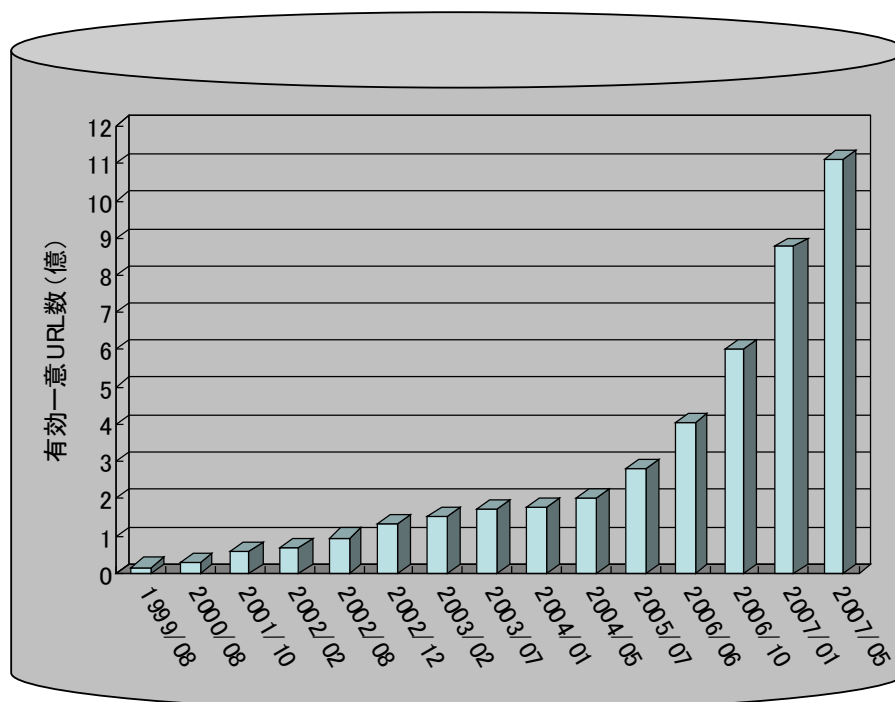
- リンク&テキスト解析を用いてウェブの全空間を地図化
- 産業連関図に相当する地図が得られる
 - 注目分野のリーサー・サーベイに有用
- 影響力のある製品ユーザのグループなども同時に抽出
 - 広告設置戦略への応用

Cyber Map for Financial Business



Web Archive (1999~2007)

特定領域研究
情報爆発



e-Society
文部科学省リーディングプロジェクト



「生協の白石さん」

ひとことカード

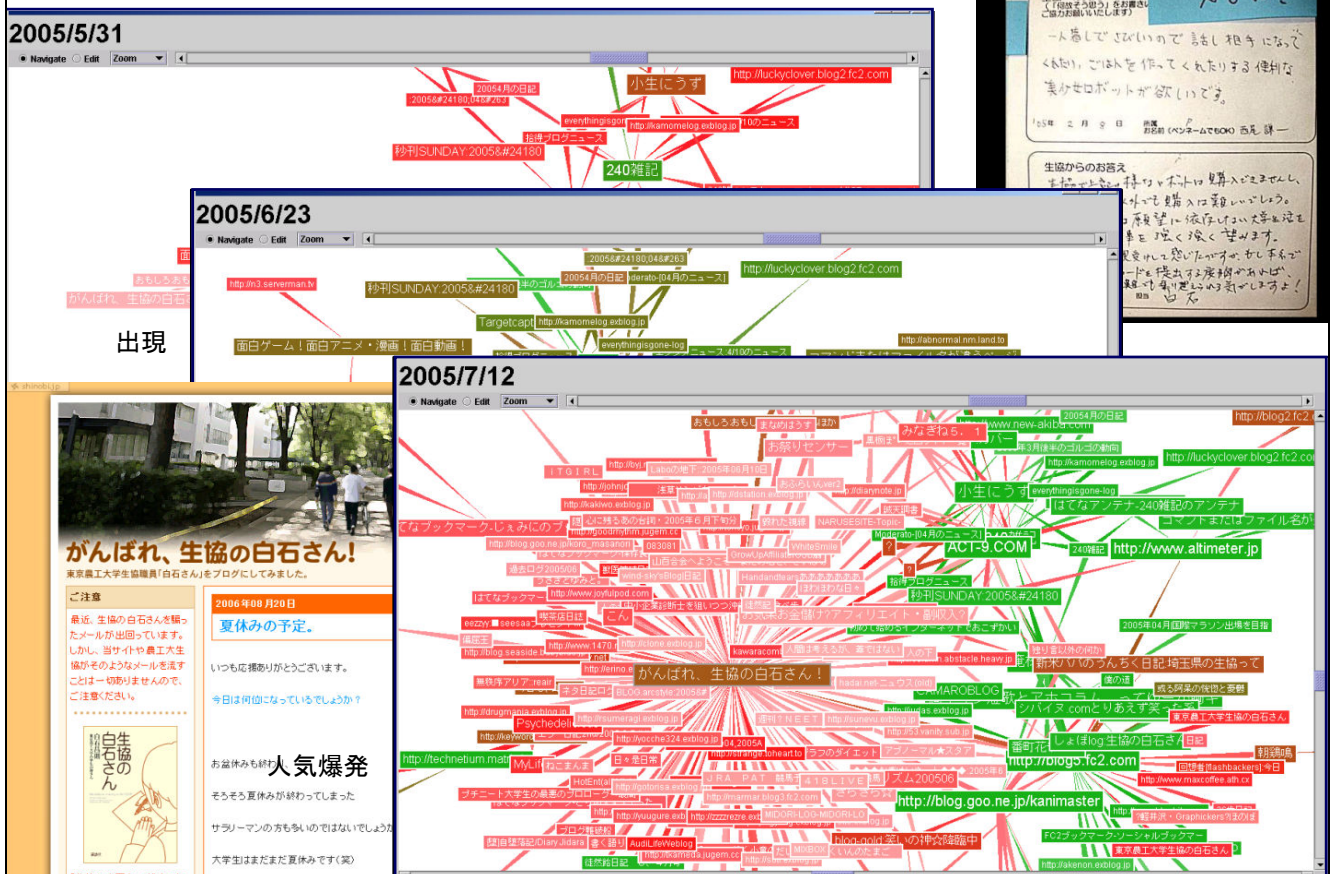
「OOをおいて!」。「OOはどうしてなの?」など
なんでも結構です。簡易書写の導入で生財はわり
です。めだたの言葉が

生活への賛同、慰労、
(感謝や励まし)を容易に
(心から贈る)にします

NEW!

一歩進んでほしいので、あきらめずに、
くわい、では作って、くれりする便利な
集かさばボタンが谷川いぞう。

1985年 2月 8日 東京 豊洲(マンハッタンビル)西元 講一



情報大航海プロジェクト

Information Grand Voyage



> お問い合わせ



プロジェクト概要

情報大航海プロジェクトが目指すゴールやプロジェクトの推進体制、実施スケジュールなど、プロジェクトの全体像をご説明しています。

- ▶ プロジェクトの背景
- ▶ プロジェクトの目的
- ▶ プロジェクトの原則
- ▶ プロジェクトの体制
- ▶ プロジェクトのスケジュール

- ▶ リンク:
経済産業省プロジェクトページへ



取り組み

情報大航海プロジェクトの現在の取り組みについてご説明しています。

- ▶ 次世代技術の実証
- ▶ 共通化

イベント情報:



11/1(金)、2(土)、ビジネスショウKANSAIに出展します。



TOPICS

情報大航海プロジェクトからのお知らせやイベントへの出展などをご紹介します。

メディア掲載:

2007.10.16

- ▶ CEATECでの活動がメディアに取り上げられました。

ニュース:

2007.7.25

- ▶ 「情報大航海プロジェクト」モデルサービス実施企業の決定について(経済産業省発表PDF)



情報大航海プロジェクト
【マイ・ライフ・アシスト サービス(仮称)】

ユーザーの日々の行動の情報を収集・蓄積

NTT DoCoMo

家



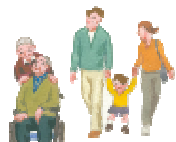
NWアクセス
(時刻・属性・嗜好・対象)

駅



各出口へのヒトの流れ
動線傾向、移動者数・
属性、NWアクセス

近所・公園



散歩経路による
歩行距離・心拍
データなどの
生体データ

情報家電・無線LAN

・車内NW・カメラ
・GPS・RFIDタグ
・FeliCa・POS

鉄道・バス



乗降者数・属性
ヒトの流れ・動線傾向
車内でのアクセス動向

レストラン



事前のNWアクセスと
実来店の関係
途中の行動特性

デパート

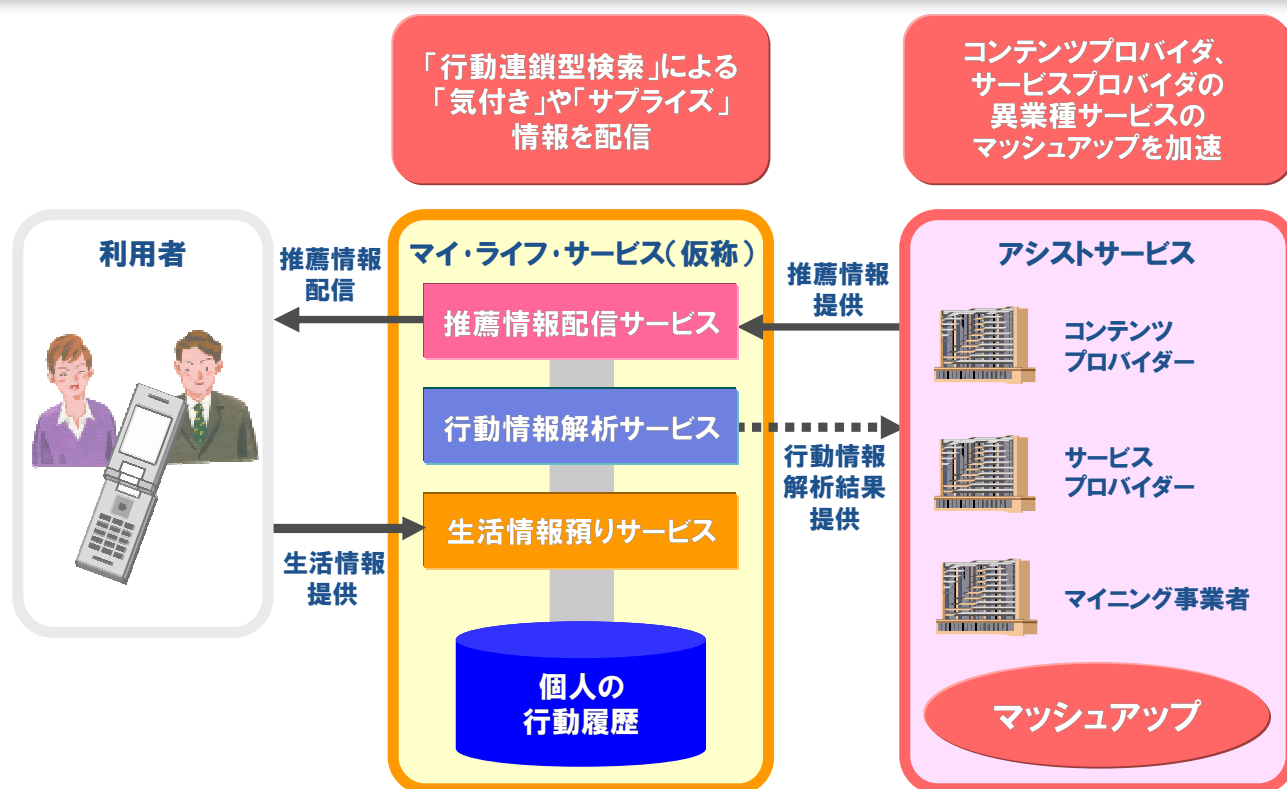


デパートでの行動特性
属性と消費の関連
セールへのレスポンス

コンビニ



店舗POSデータの
統計的なリアルタイム
消費動向特性



「こんな情報、ちょうど欲しかったんだ！」を実現する サービス名: プレビュー・チャンネル

過去の蓄積から次の行動を予測
必要な情報を先行調査

過去の行動履歴に無い
意表をつく提案

何気なくケータイを覗くと、自分の行動や潜在的欲求にマッチした内容を表示。



今週は頑張ったし、
ちょっと自分にご褒美
あげてもいいかな！



最近、残業で
帰りの遅い日が
続いていたらなあ。
今日は、久々の
定時退社。

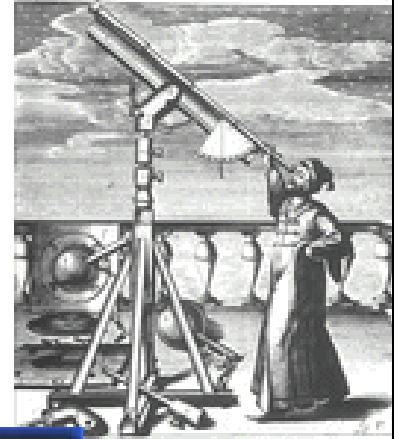


The Evolution of Science

- **Observational Science**
 - Scientist gathers data by direct observation
 - Scientist analyzes data
- **Analytical Science**
 - Scientist builds analytical model
 - Makes predictions.
- **Computational Science**
 - Simulate analytical model
 - Validate model and makes predictions
- **Data Exploration Science**

Data captured by instruments
Or data generated by simulator

 - Processed by software
 - Placed in a database / files
 - Scientist analyzes database / files

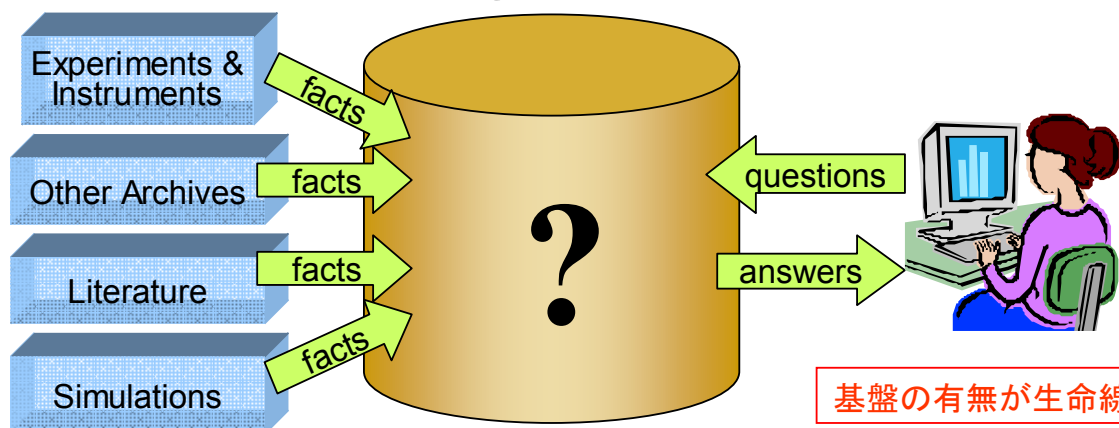


Jim Gray氏より



The Big Picture

Jim Gray氏より

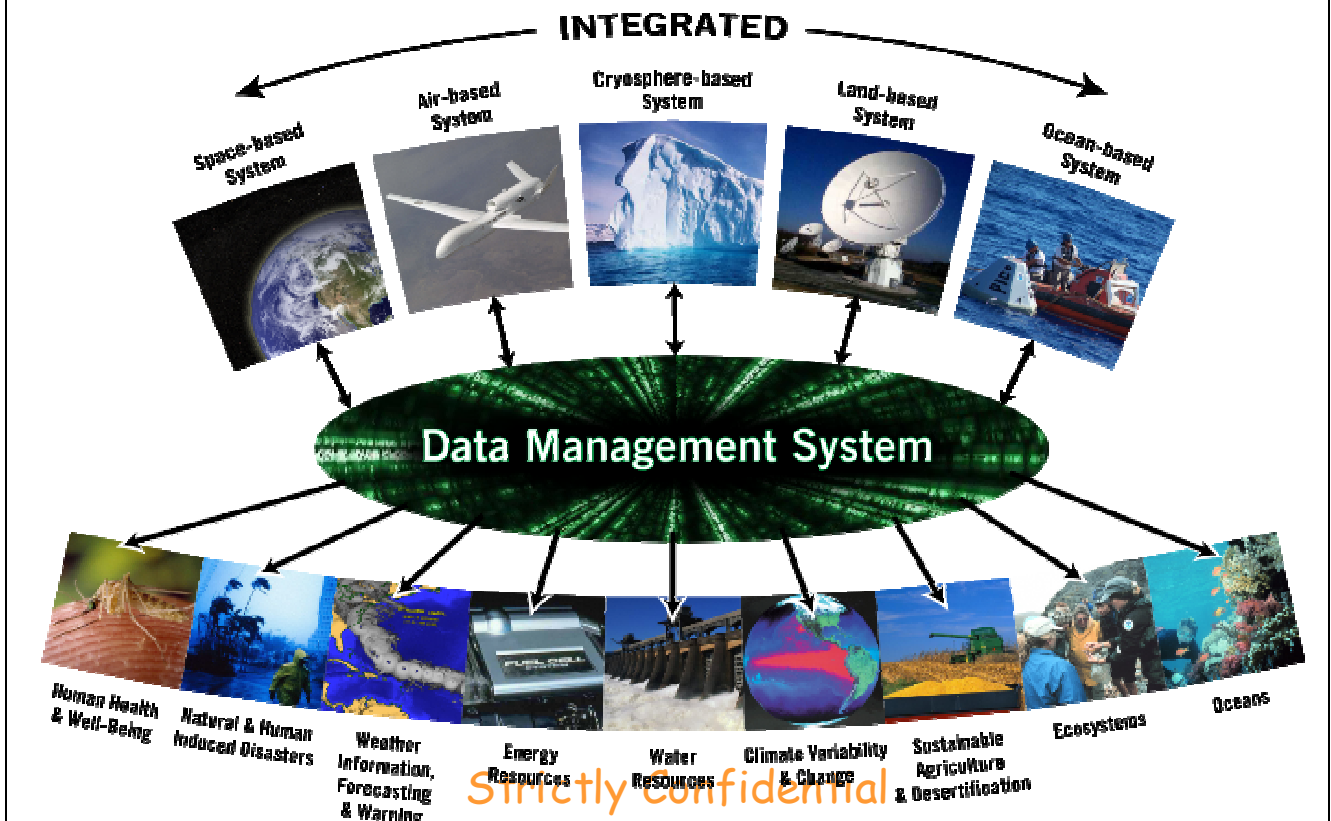


基盤の有無が生命線

The Big Problems

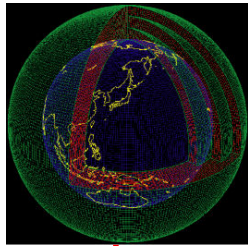
- Data ingest
- Managing a petabyte
- Common schema
- How to organize it?
- How to reorganize it
- How to coexist with others
- Query and Vis tools
- Support/training
- Performance
 - Execute queries in a minute
 - Batch query scheduling

Global Earth Observation System of Systems (GEOSS)



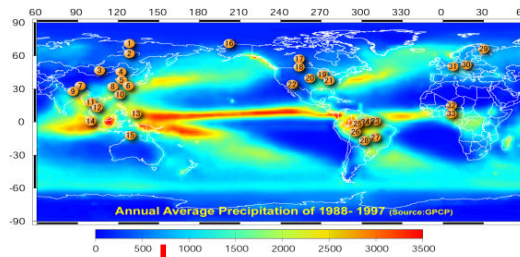
CEOP The First Global Integrated Data Sets of the Water Cycle

Model Outputs by Numerical Weather Prediction Centers



MODEL Output Data Archiving Center at Max-Planck Institute of Germany
<http://www.mpg.de/>

Surface Observational (*in-situ*) Data from the 36 CEOP Reference Sites



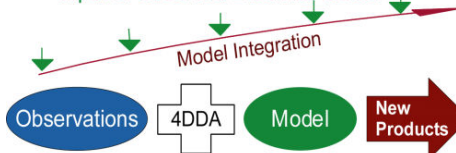
In-Situ Data Archiving Center at UCAR (Center at University Corporation for Atmospheric Research) of USA
<http://www.ucar.edu/>

Satellite Remote Sensing Data



Data Integrating/Archiving Center at University of Tokyo and JAXA or Japan
<http://monsoon.t.u-tokyo.ac.jp/ceop/>

Input of Observed Data into Model

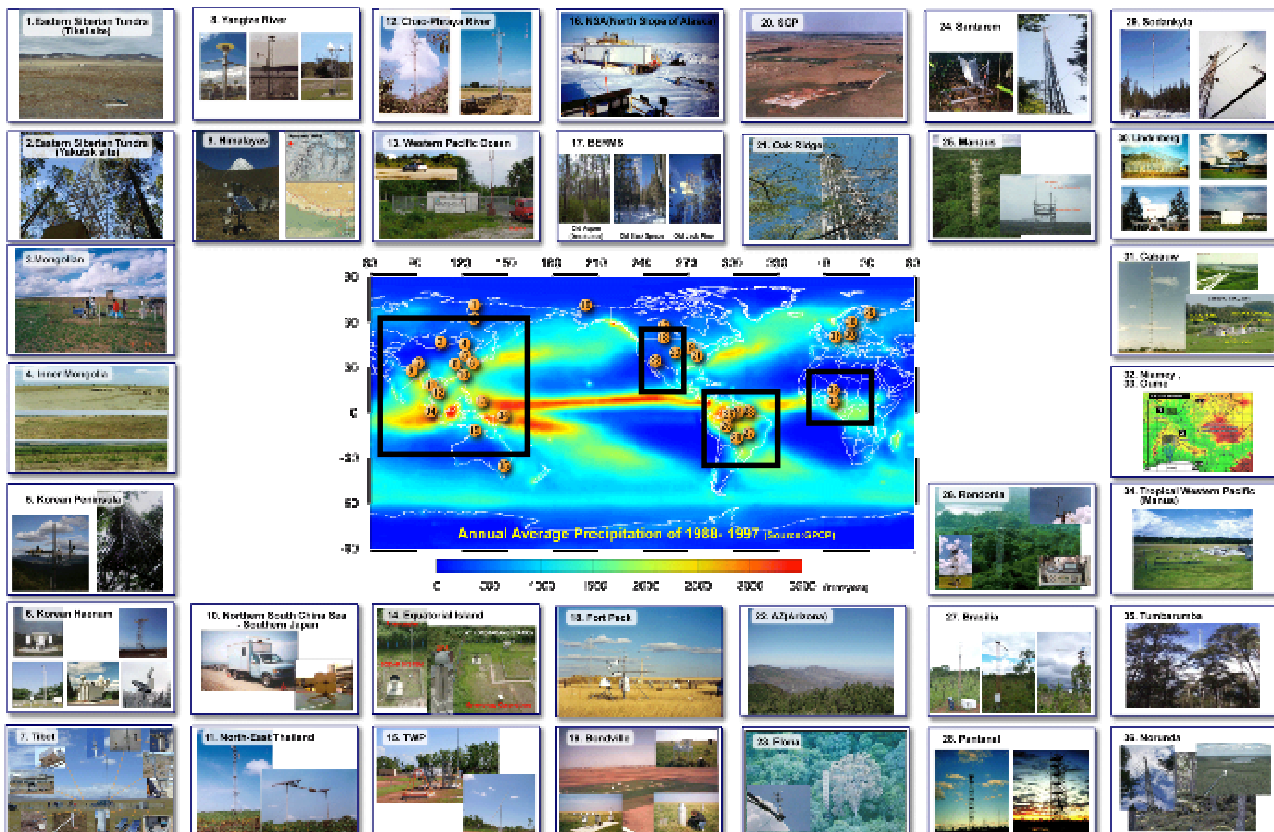


Global Land Data Assimilation System at NASA Goddard Space Flight Center of USA
<http://das.gsfc.nasa.gov/>



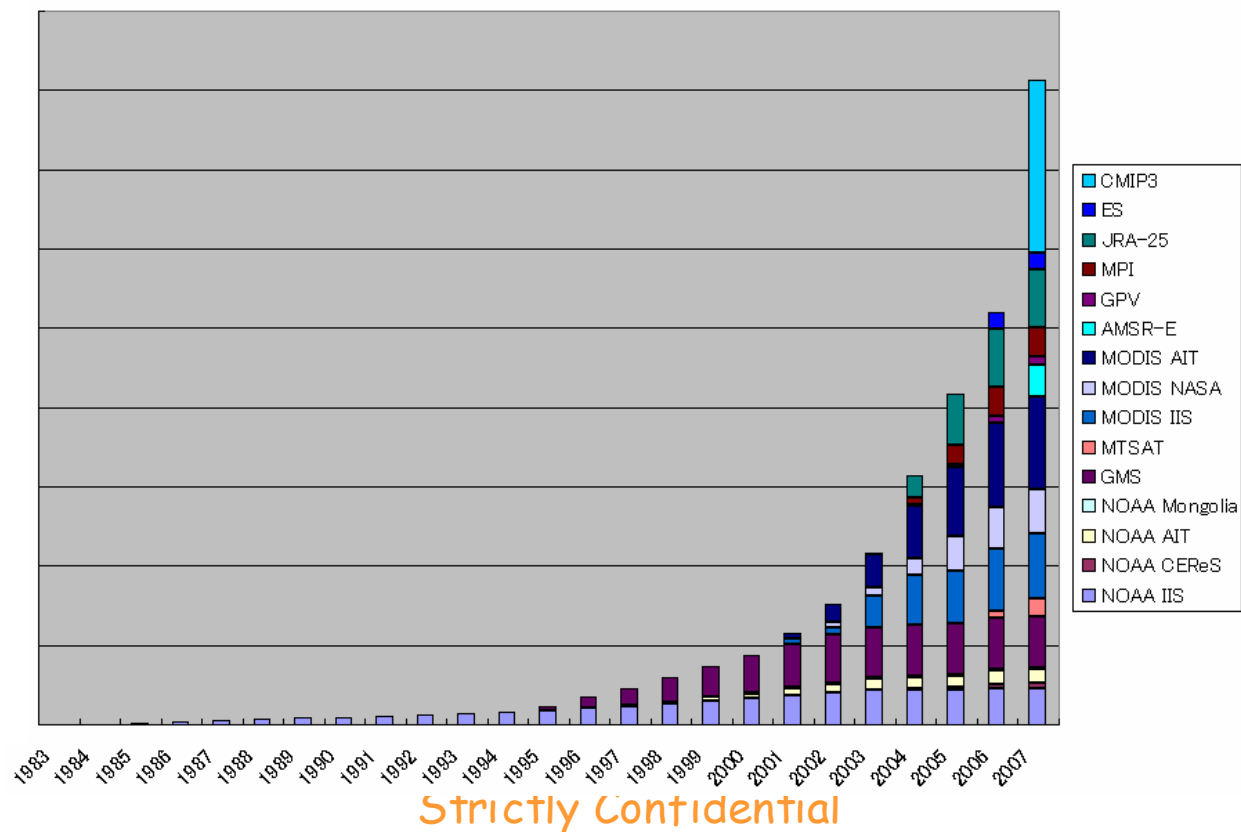
Data Archive Center

CEOP International Cooperation for the Global Coverage

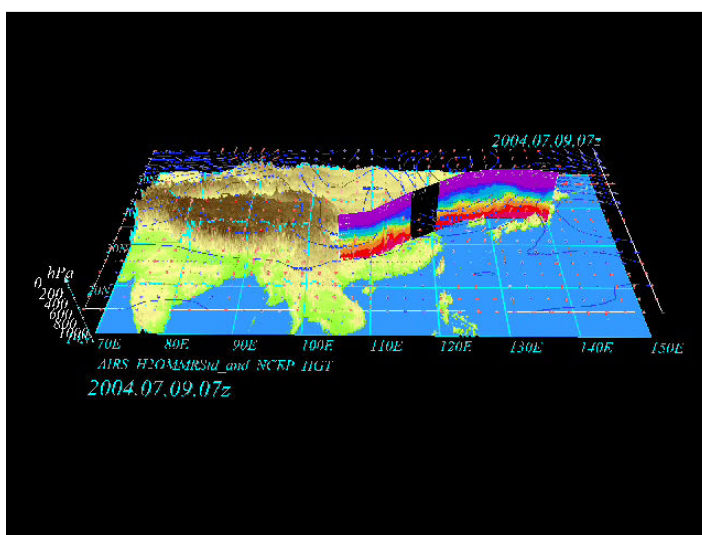


STRICTLY CONFIDENTIAL

喜連川研 地球環境デジタルライブラリ



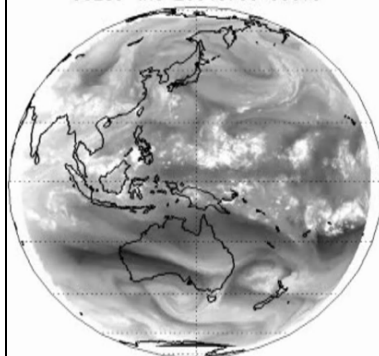
2004年7月13日 新潟豪雨



極軌道衛星(水蒸気・気温三次元分布)

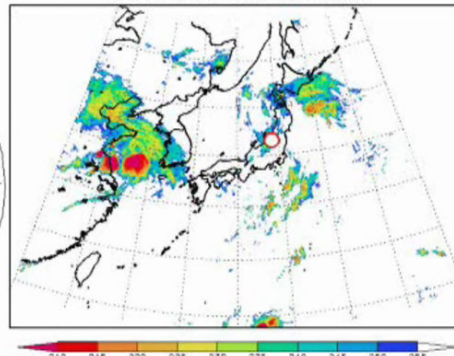
静止衛星 (水蒸気の流れ)

GOES9-IR3 20040709 00UTC



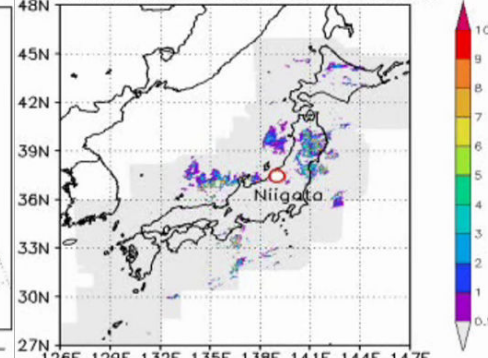
静止衛星 (雲の動き)

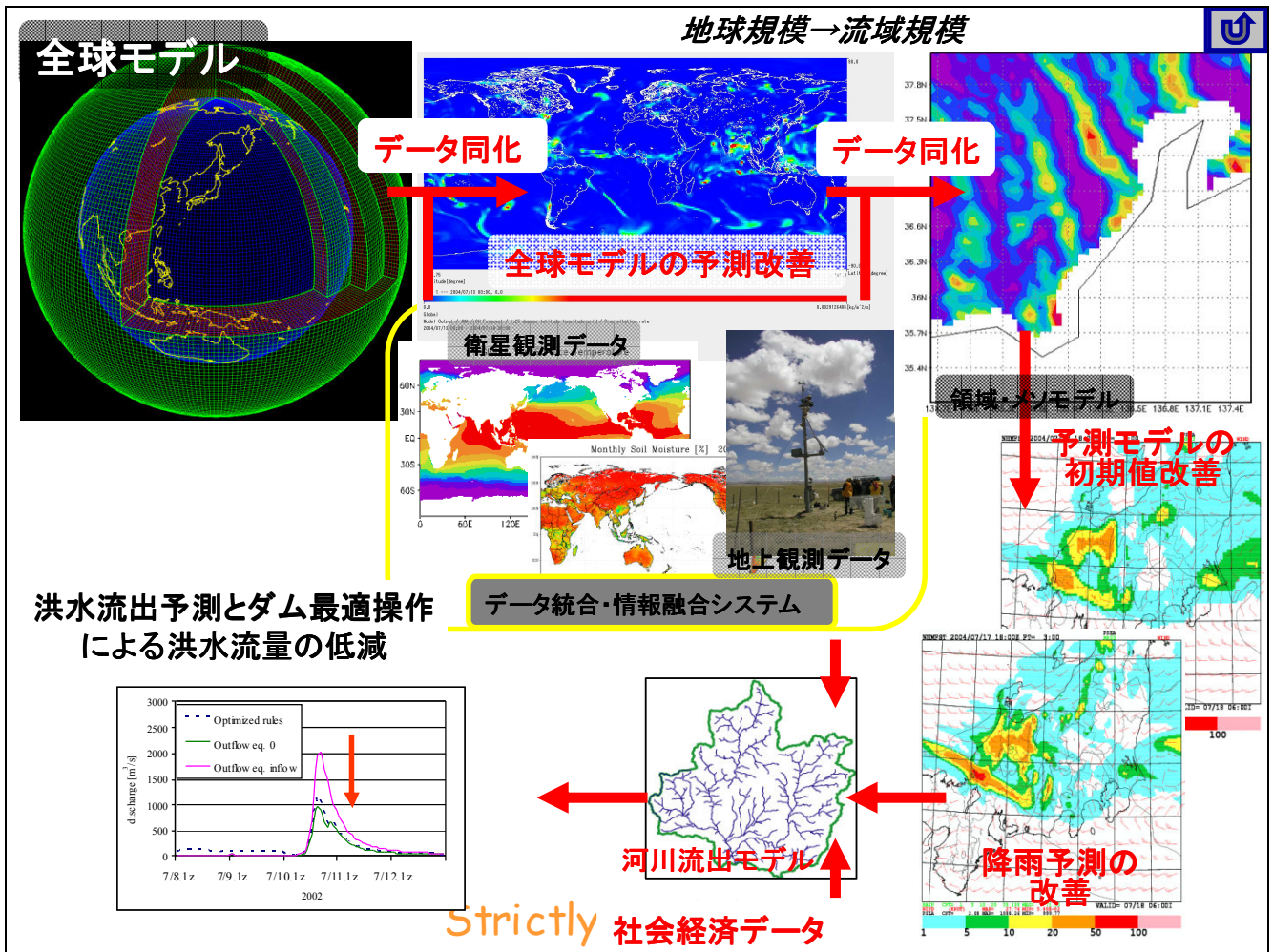
GOES9 IR1 TB 2004 0711 09JST



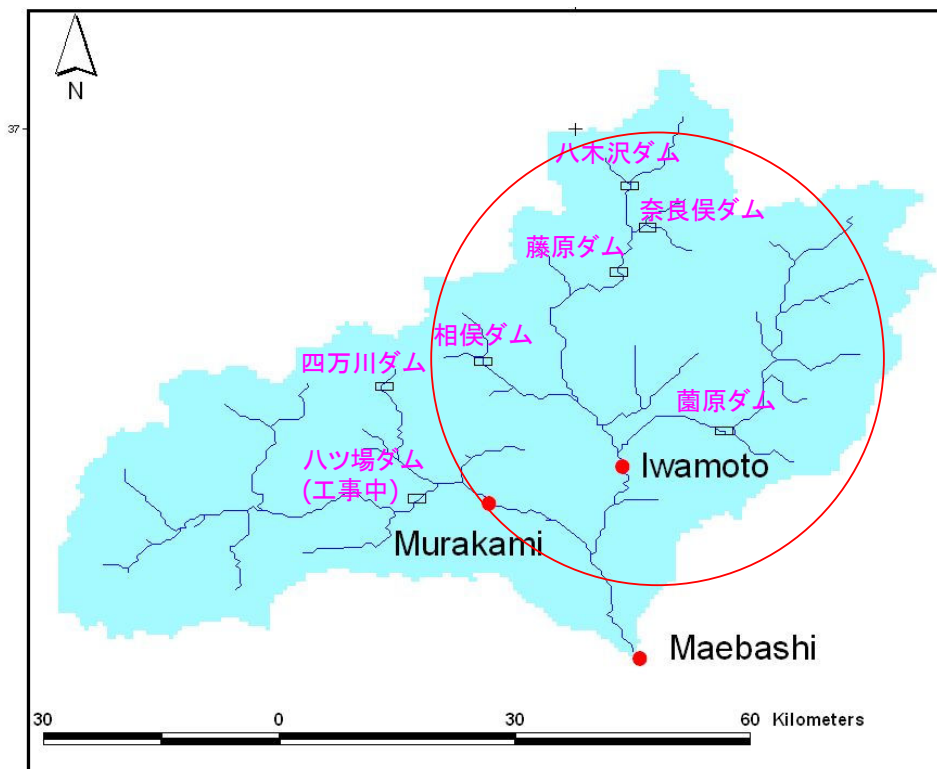
レーダー・アメダス合成情報 (降雨分布)

JMA RADAR-AMeDAS 2004 0711 00:30JST

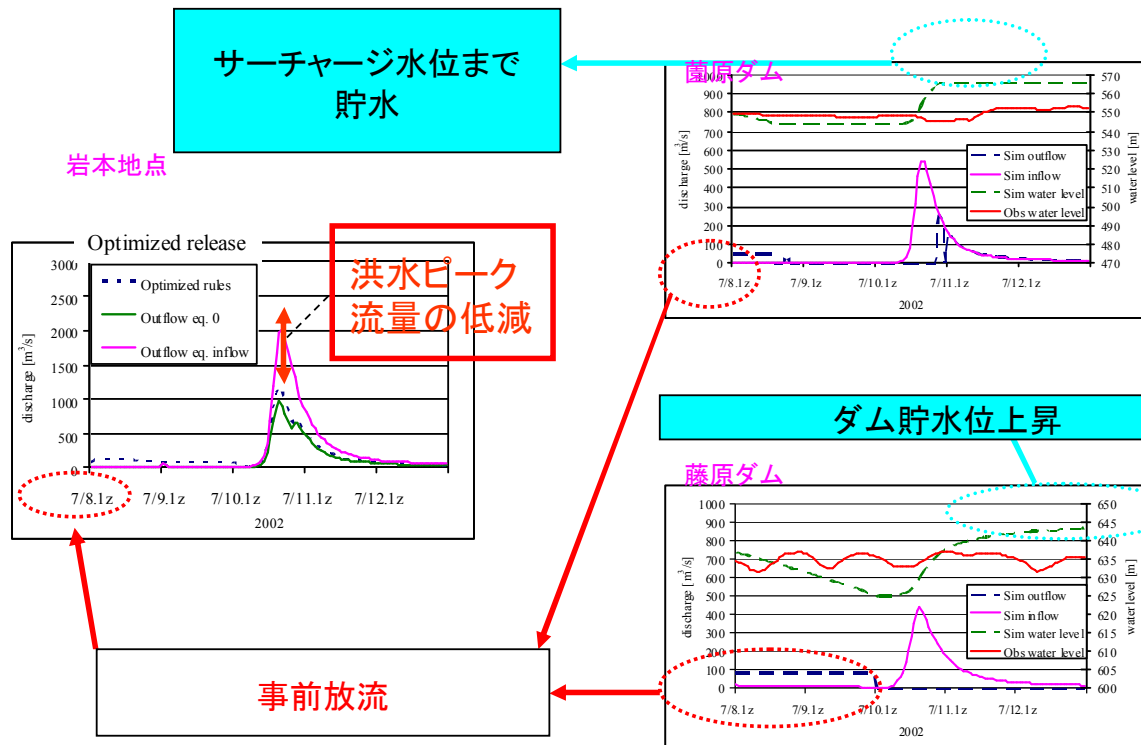




利根川上流域

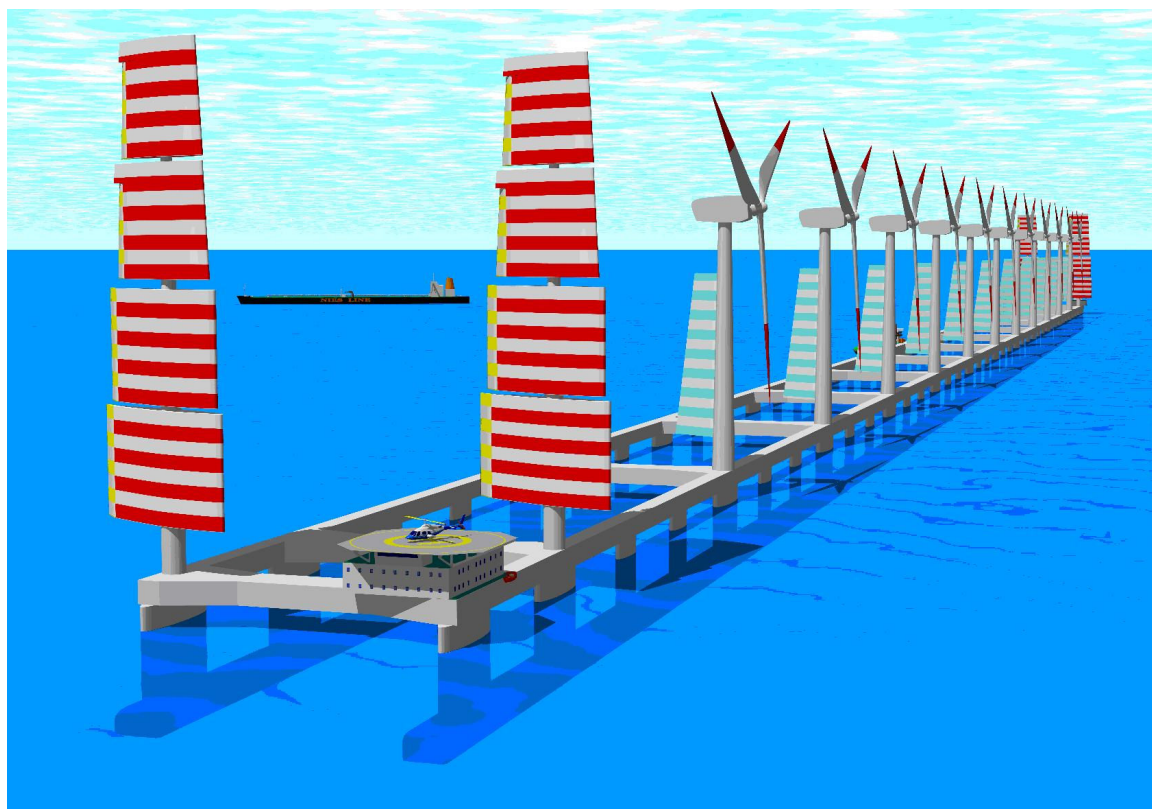


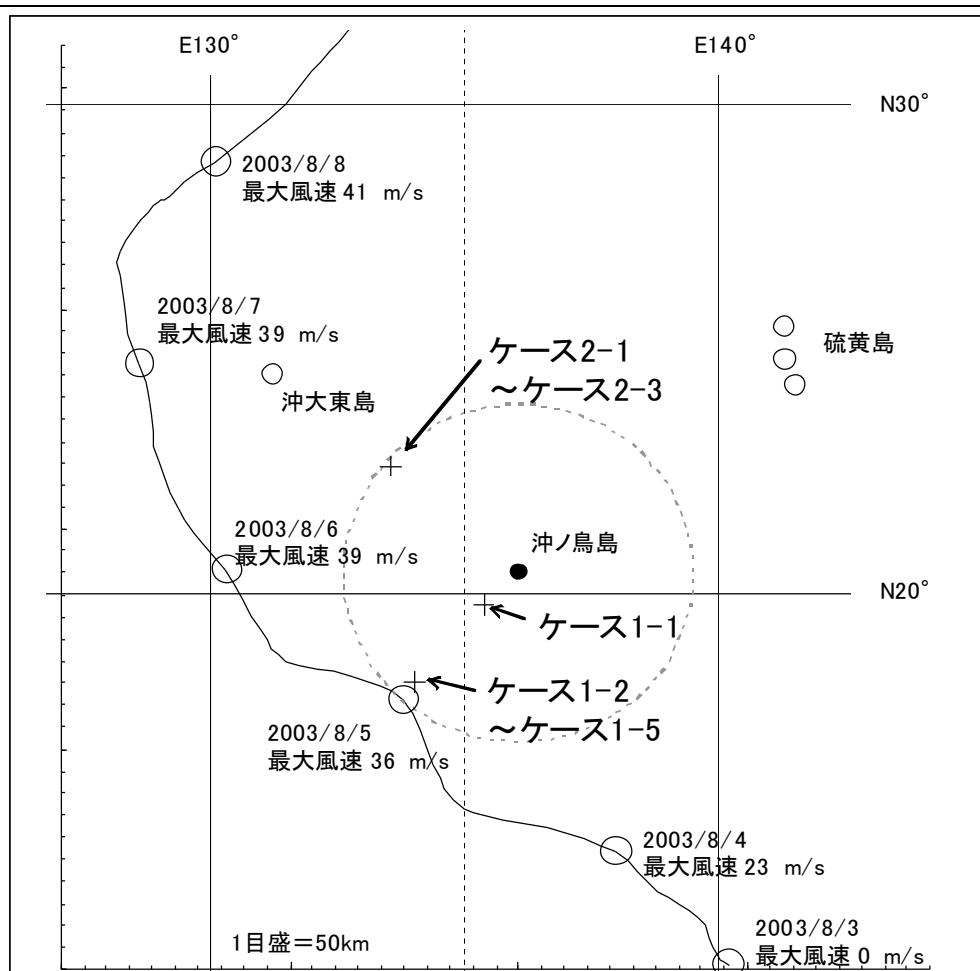
気象庁13~18時間降雨予測を用いた最適ダム操作

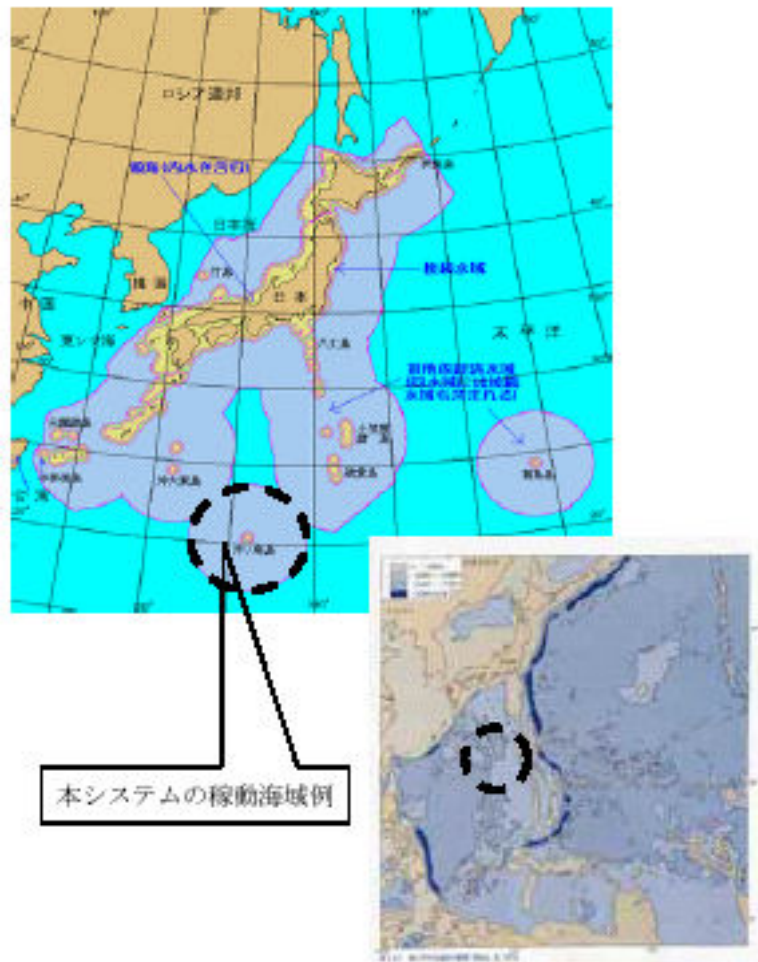


Strictly Confidential

洋上発電





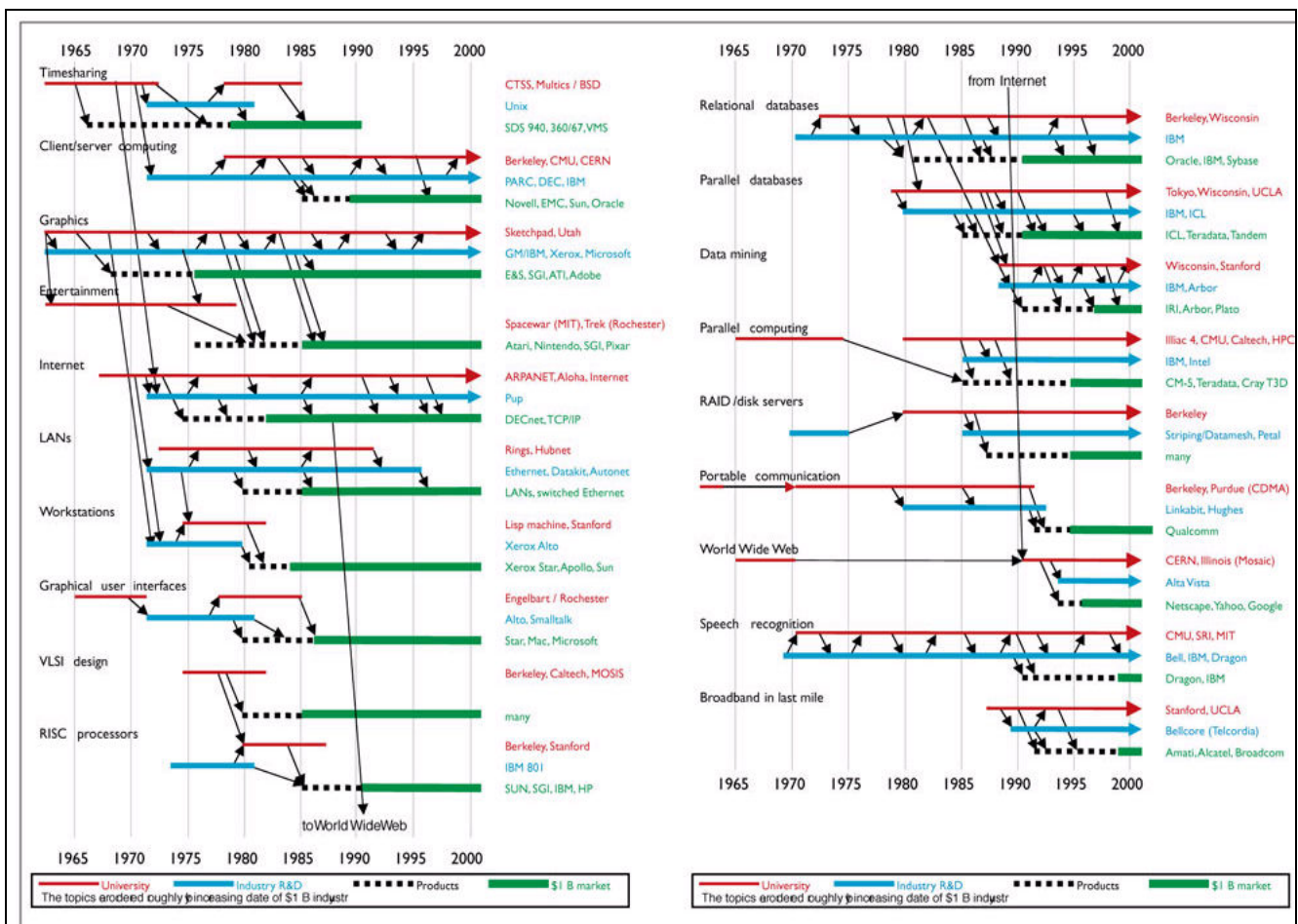


Goal

- 循環型エネルギー系は自然エネルギーの基幹エネルギーとしての利用によってのみ成立する
- 経済性に加えて環境負荷を最小にすべき
- 水素社会では外洋はエネルギーの宝庫
- 帆走メガフロートは適度な風を求めて移動し、暴風を回避できる
- 1400ユニットで日本の総発電量の18%を賄え、CO2排出を約10%削減可能
- 副次的に洋上に情報ネットワークが形成され軍事的のみならず資源管理、環境管理に有益

情報爆発(大量・異種情報) からの価値創出

情報融合炉



Source: From [6], reprinted with permission from the National Academy of Sciences, courtesy of the National Academies Press, Washington D.C. © 2003.