

Big Data et Statistique Publique

Miracle ou mirage ?

Matthieu Cornec
Mission Innovation



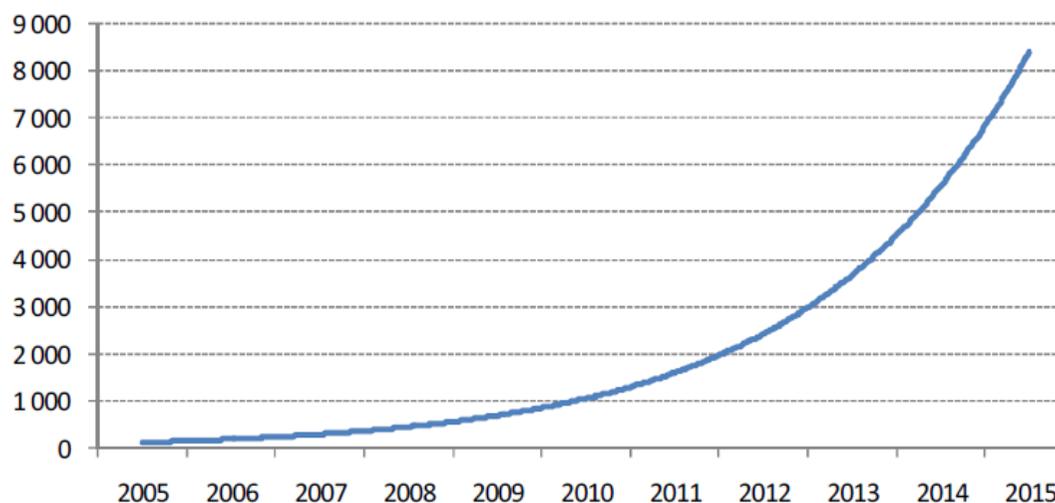
Sommaire

1. Constat : Explosion des données
 1. Constat
 2. Exemples
2. Big Data : nouvelle révolution économique ?
 1. Cas d'usage
 2. Technologies : enjeux, panorama
3. Quels enjeux pour la Statistique publique ?
 1. Enjeux
 2. Exemples : Prix, prix, emploi
 3. La méthodologie de la Statistique Publique, avenir du Big Data Analysis ?

1. Constat : Explosion des données

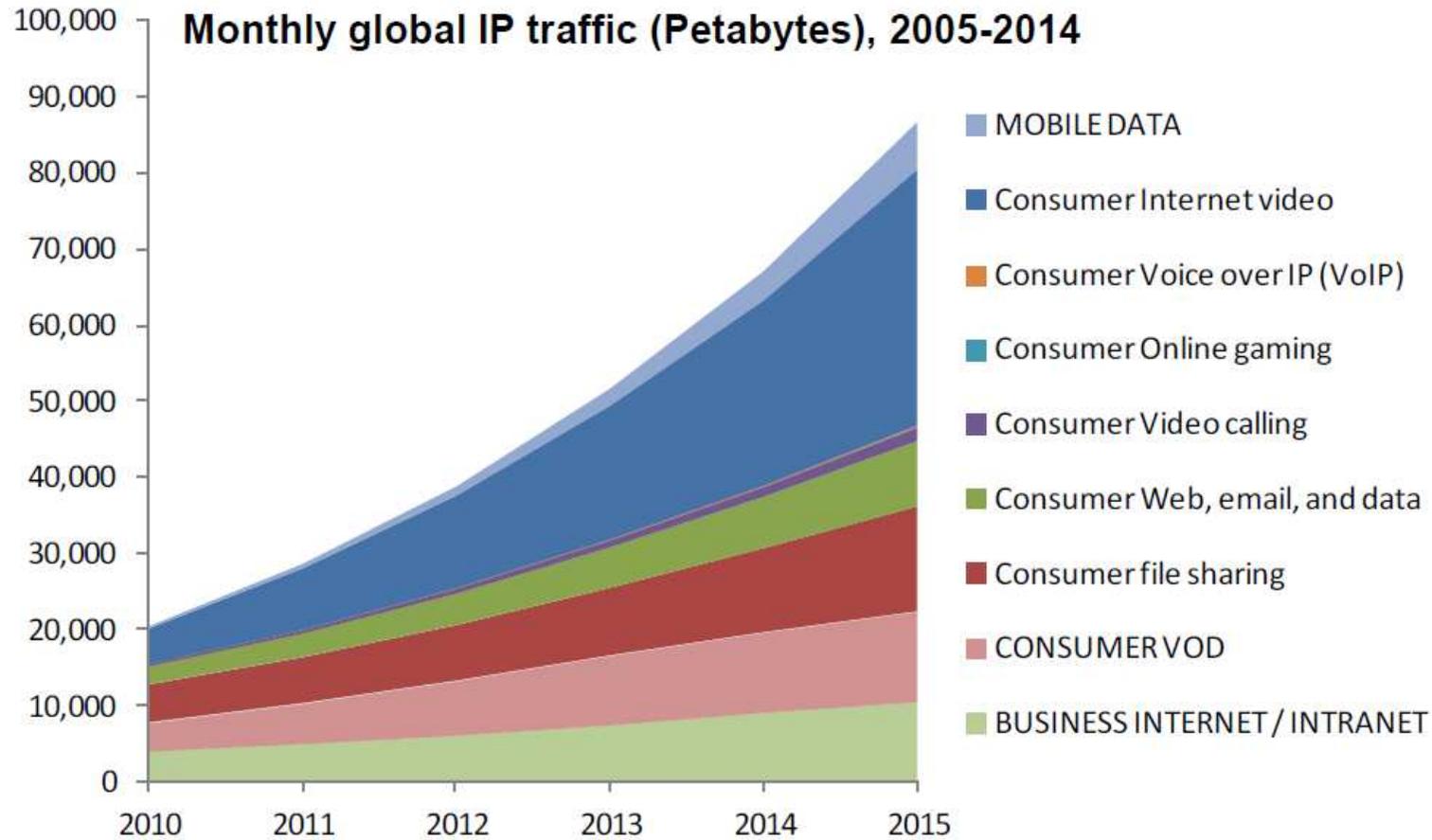
650 % de croissance des données dans les 5 ans à venir, dont **80% non structurées**

Figure 1. Estimated world data storage in exabytes (billions of gigabytes)



Source: OECD based on IDC Digital Universe research project.

1. Constat : Explosion des données



Source: CISCO VNI 2011

1. Exemples



- 150 millions de Tweets par jour
- 60 millions de mises à jour sur Facebook
- 180 millions de blogs
- 135 millions de comptes LinkedIn
- 63,6 millions de tablettes vendues dans le monde en 2011
- en France :
 - 19 millions d'utilisateurs de smartphone,
 - 20 millions d'usagers Facebook,
 - 20 millions de blogs,
 - 3 millions de comptes sur LinkedIn,
 - 5 millions de comptes Viadeo...

2. Discours marketing : Big Data is Big Money

- *3ème volet de la révolution marketing*
- *Pétrole du 21ème siècle*
- *Les entreprises de demain seront « data centric ».*



2. Cas d'usage : Jeux vidéos

JEUX EN LIGNE

(IsCool Entertainment, PARIS)

- Détecter et classifier les comportements de joueurs : habitudes, addictions, fraudes...
- Traiter 50 Go de données par jour
- Données essentiellement dans le Clc



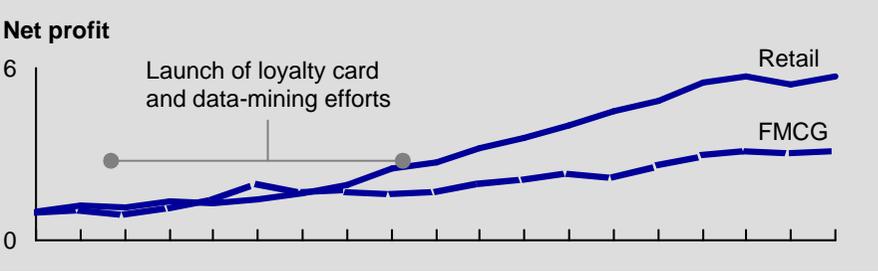
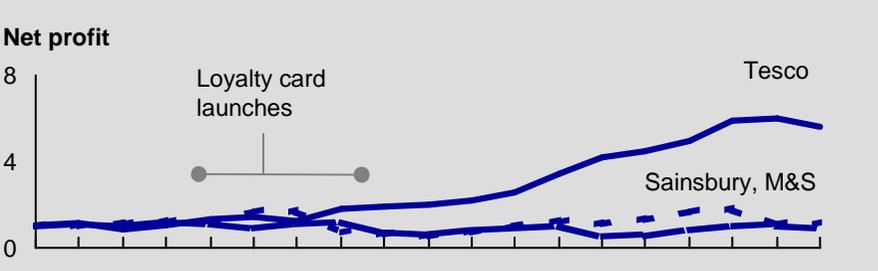
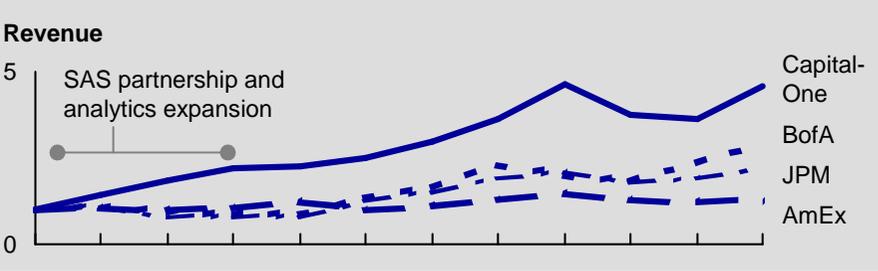
2. Cas d'usage : Transport

PILOTAGE DE LOGISTIQUE EN TEMPS RÉEL

- FedEx suit en temps réel les commandes d'expéditions.
- Utiliser le moyen de transport le plus économique à un instant donné.
- Détection des comportements commerciaux atypiques des clients.



2. Plusieurs acteurs de la distribution ont capturé de la valeur en utilisant le big data (source MacKinsey)

Actions	Battle and outcome	Impact (index 1991 = 1)
 	<ul style="list-style-type: none"> Walmart gained upper hand in negotiations with P&G by mining 600 terabytes of data for customer preferences and buying behavior 	
 	<ul style="list-style-type: none"> Tesco increased basket size by 28% and took U.K. share leadership by understanding consumer behavior and linking purchases to a loyalty card program 	
 	<ul style="list-style-type: none"> CapitalOne outgrew the competition by running 60,000 yearly product configuration experiments to understand consumer behavior and create new winning card offers 	

2. Technos : le discours marketing

les 4V du Big Data

Volume : les technologies actuelles sont inadaptées à cette croissance effrénée.

Vélocité : L'accès et le partage des données doit se faire en temps réel.

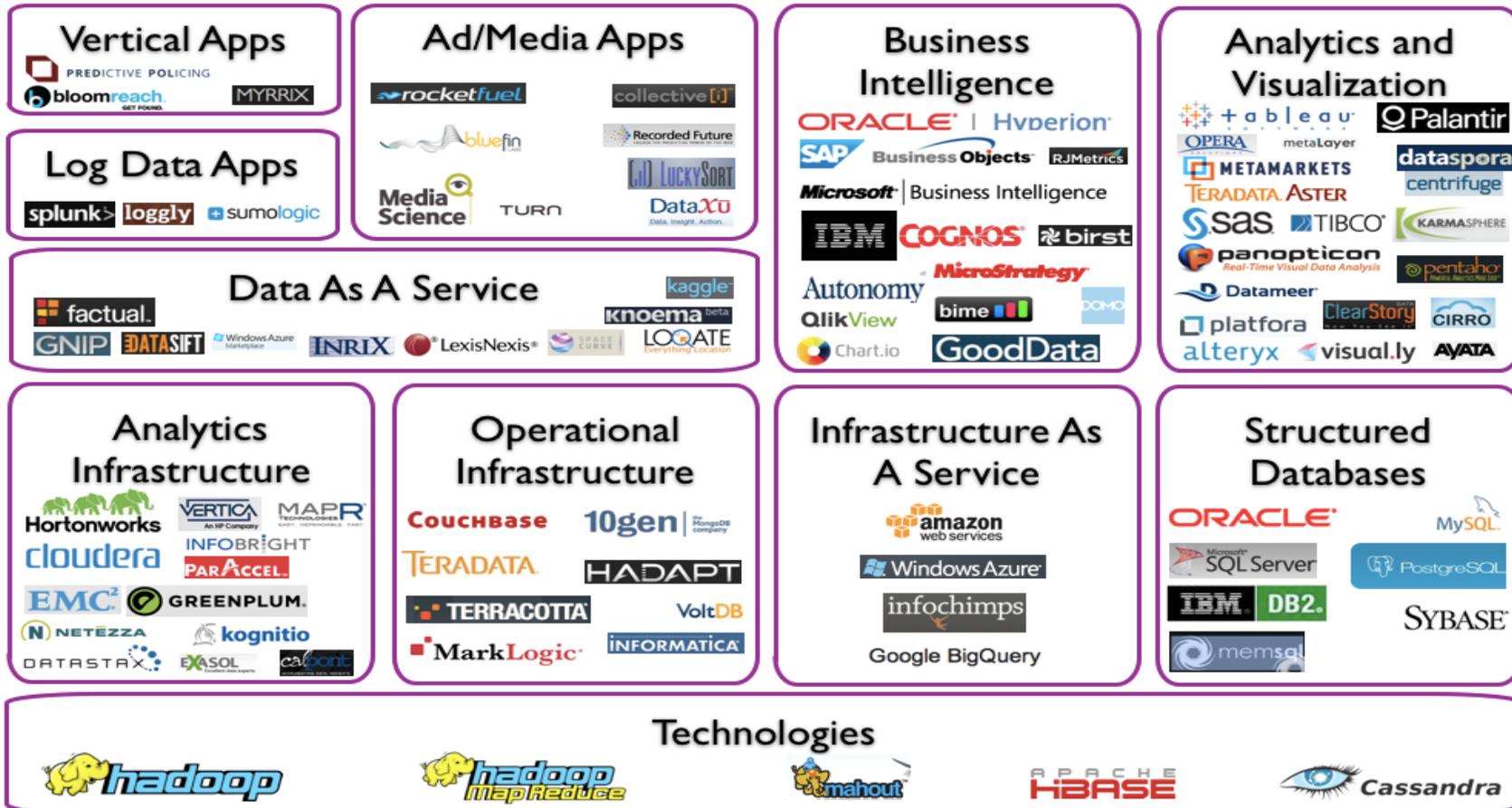
Variété : l'entreprise est confrontée à des données non structurées : emails, web, réseaux sociaux, son, image, video...

Variabilité : On ne sait pas prévoir l'évolution des types de données



3. Big Data : quelles technos ?

Big Data Landscape



Copyright © 2012 Dave Feinleib

dave@vcldave.com

blogs.forbes.com/davefeinleib

4. Big Data et la Statistique Publique ?

Aujourd'hui, sommes-nous dans le « Big Data » dans la Statistique Publique ?

Recensement + Sources fiscales + < 1 To/an

 Non

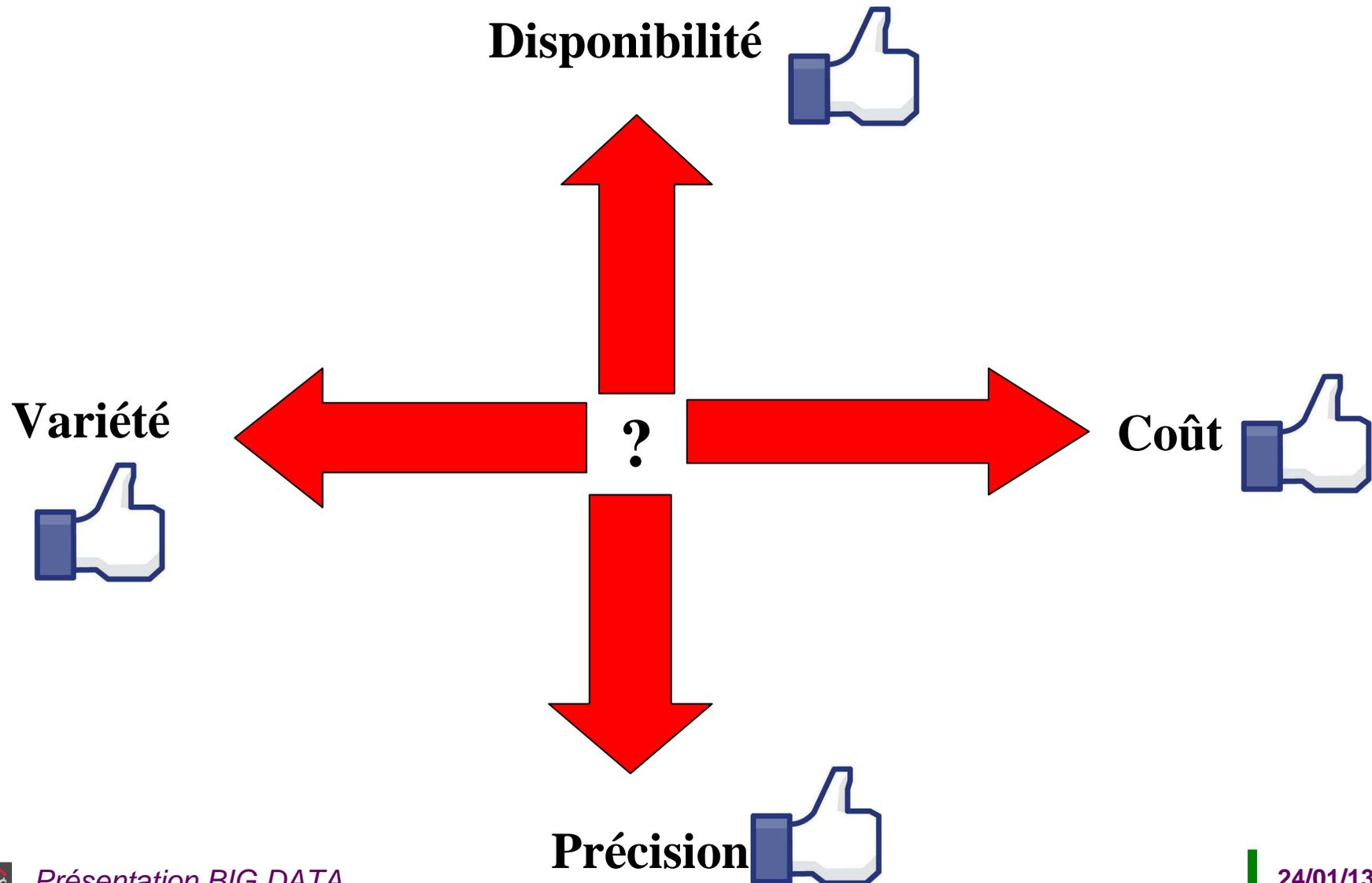
4. Big Data et la Statistique Publique ?

« The sample survey, –the most important invention of the social sciences in the 20th Century, is in trouble. «

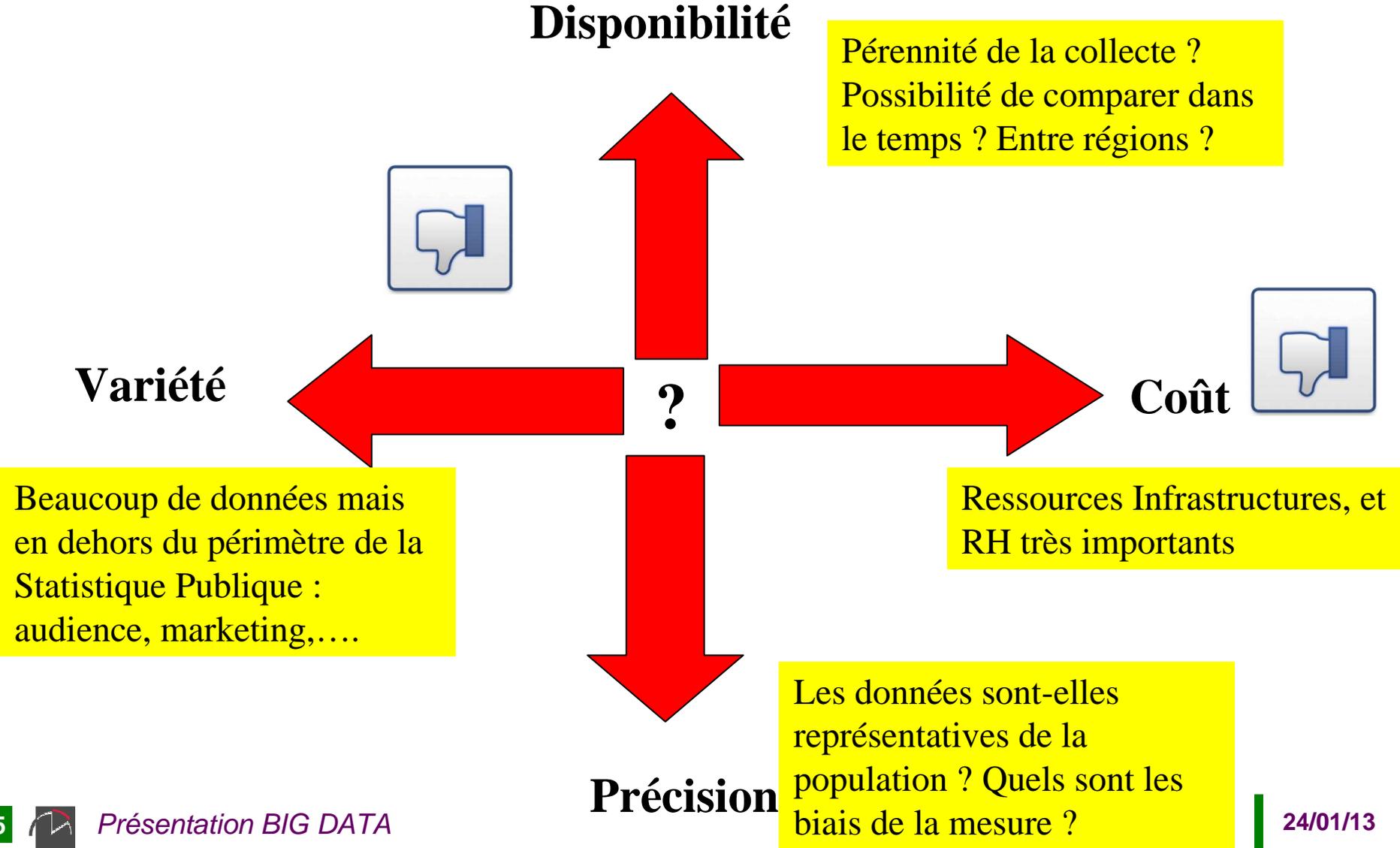
« Current practices are unsustainable ».

- ◆ **Robert Groves, directeur du U.S. Census Bureau**

4. Les 4 promesses du Big Data pour la Statistique Publique



4. La Statistique Publique, c'est de la good data avant d'être de la Big Data



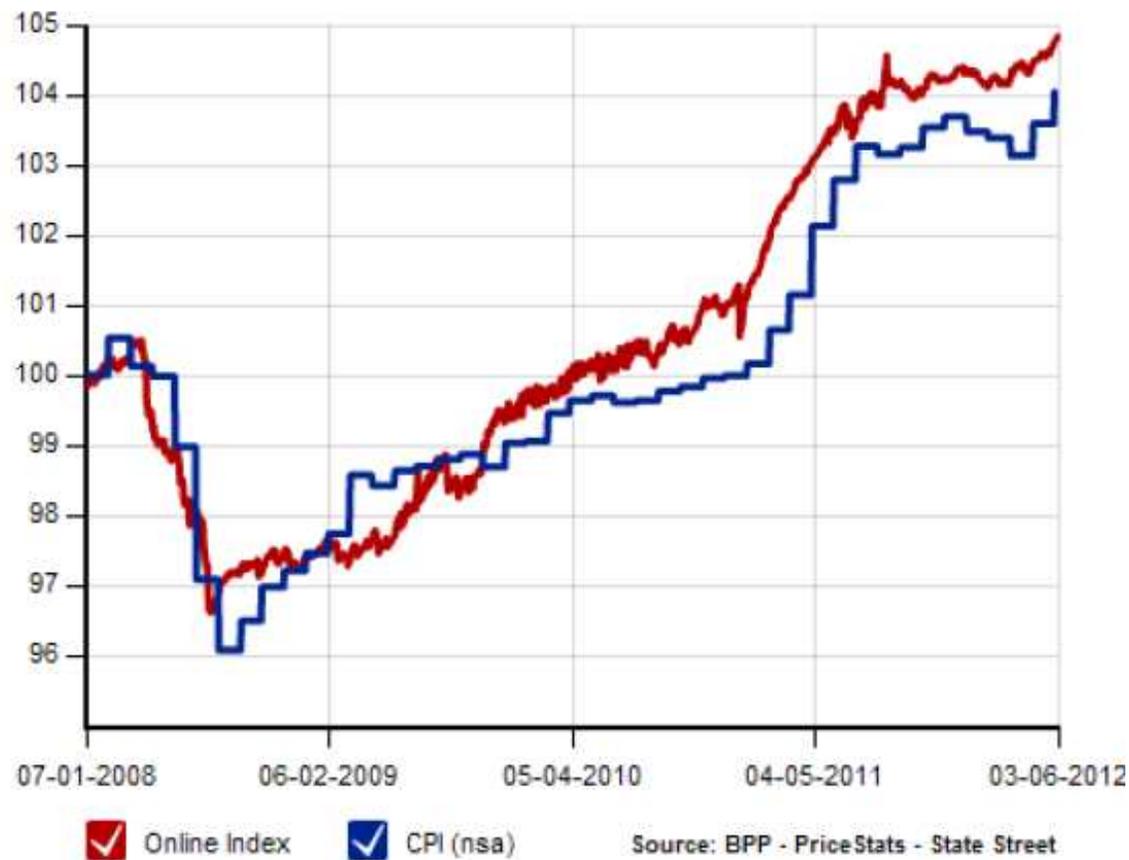
4. Stratégie Big Data pour la Statistique Publique

- Prouver le concept sur quelques exemples ciblés
- Sur les données, commencer par des indicateurs avancés plutôt que de la mesure
- Sur les logiciels, tester les technologies (Hadoop,...)
- Mettre en œuvre des partenariats public-privé

4. Big Data et la Statistique Publique : Prix

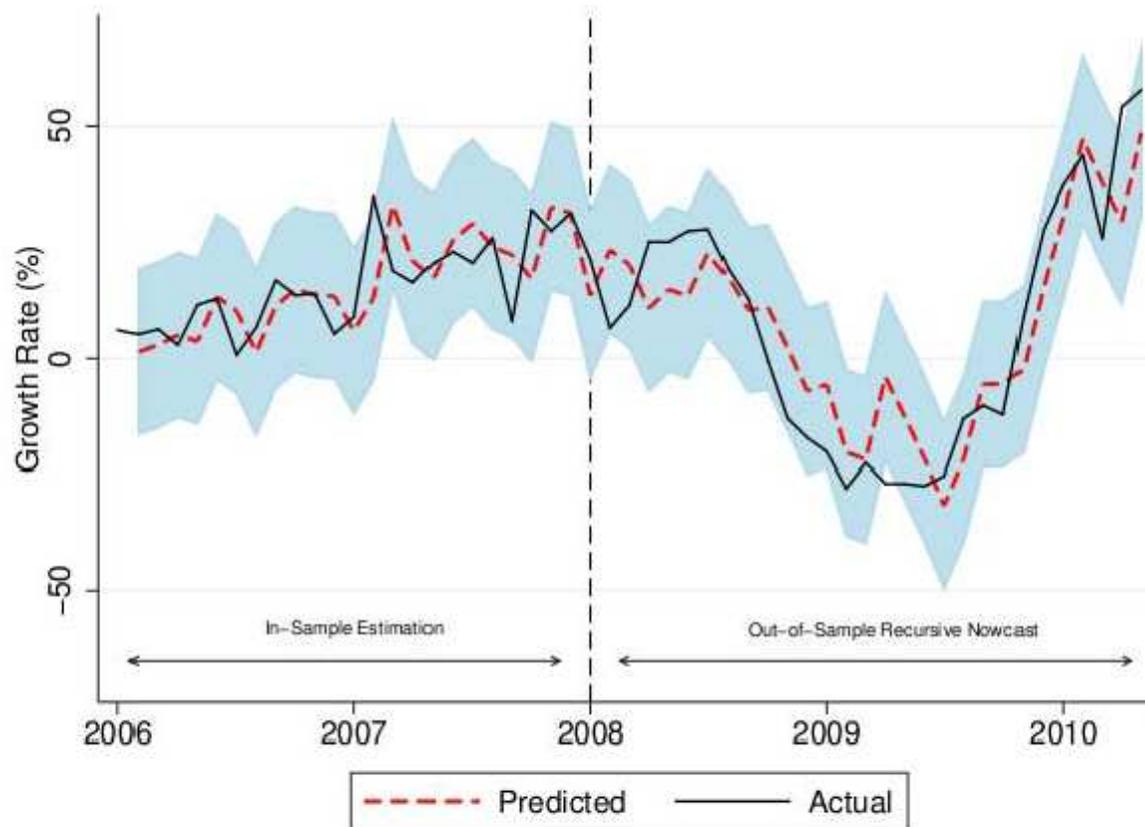
Figure 3. Daily online price index, United States, 2008-2012

Index, 100 = 01 July 2008



4. Big Data et la Statistique Publique : Production

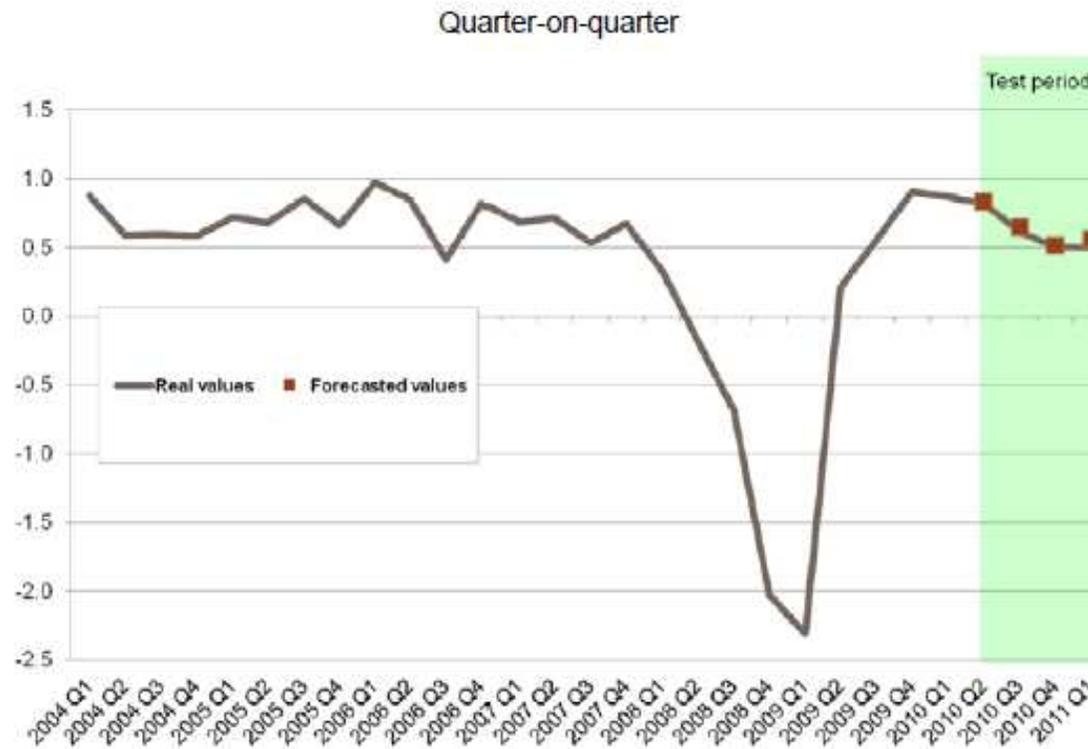
Figure 9. Monthly prediction of growth in automobile sales in Chile, January 2006-July 2010



Source: Carrière-Swallow and Labbé (2010).

4. Big Data et la Statistique Publique : Production

Figure 10. SWIFT Index predictive power: OECD GDP quarterly growth, Q1 2004-Q1 2011

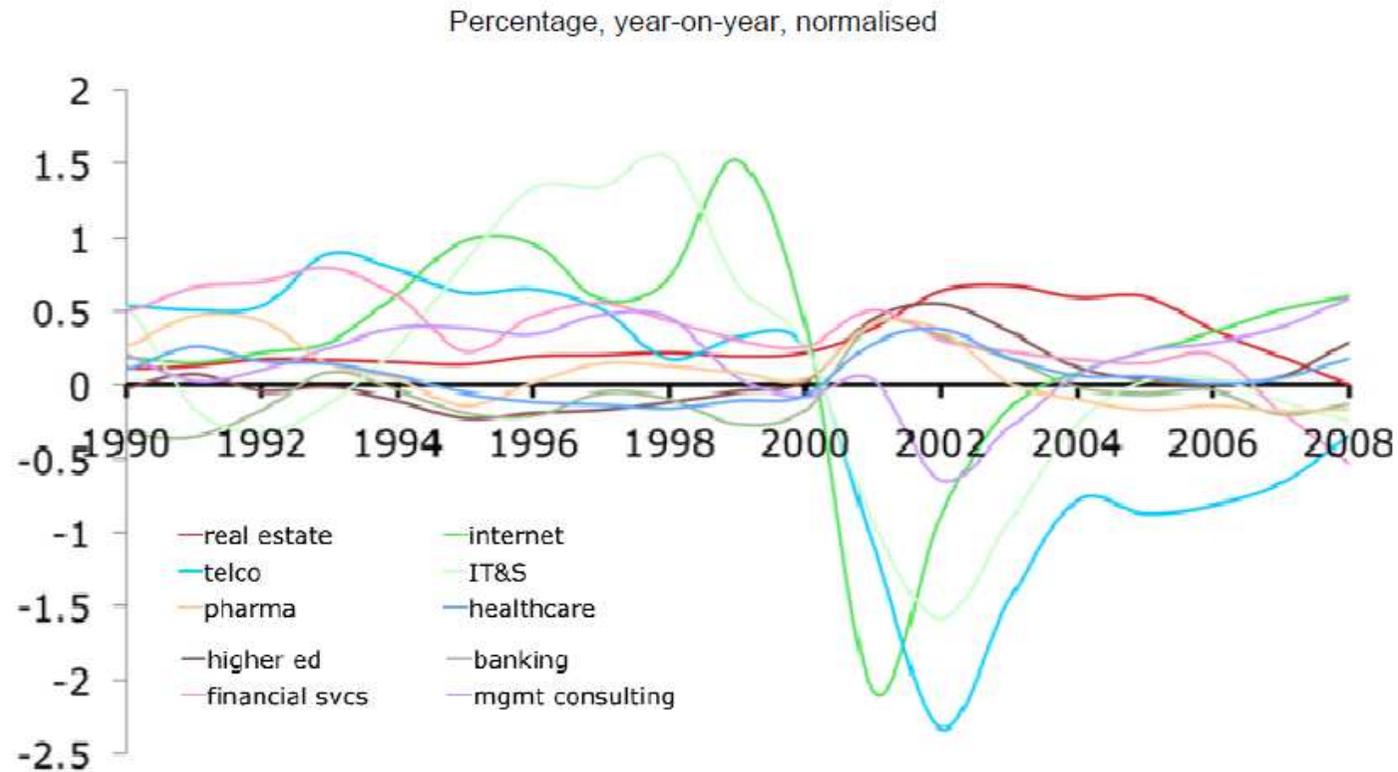


Note: Test period forecasts based on OECD GDP data published in May 2011 and SWIFT Index total aggregate series.

Source: SWIFT.

4. Big Data et la Statistique Publique : Emploi

Figure 7. Net job starters by top industries by volume, 1998-2008

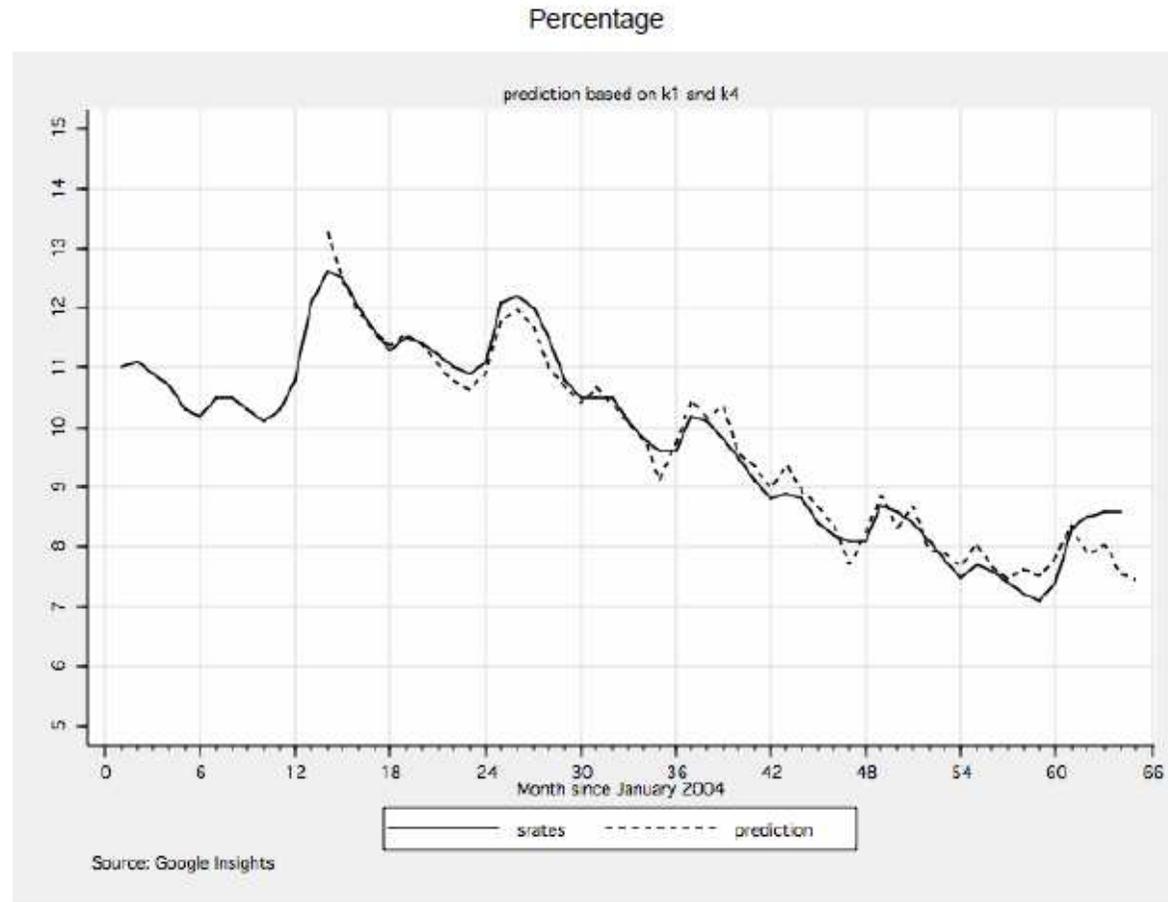


Note: 6 out of 147 industries represent 25% of new positions in 2011.

Source: LinkedIn Analytics.

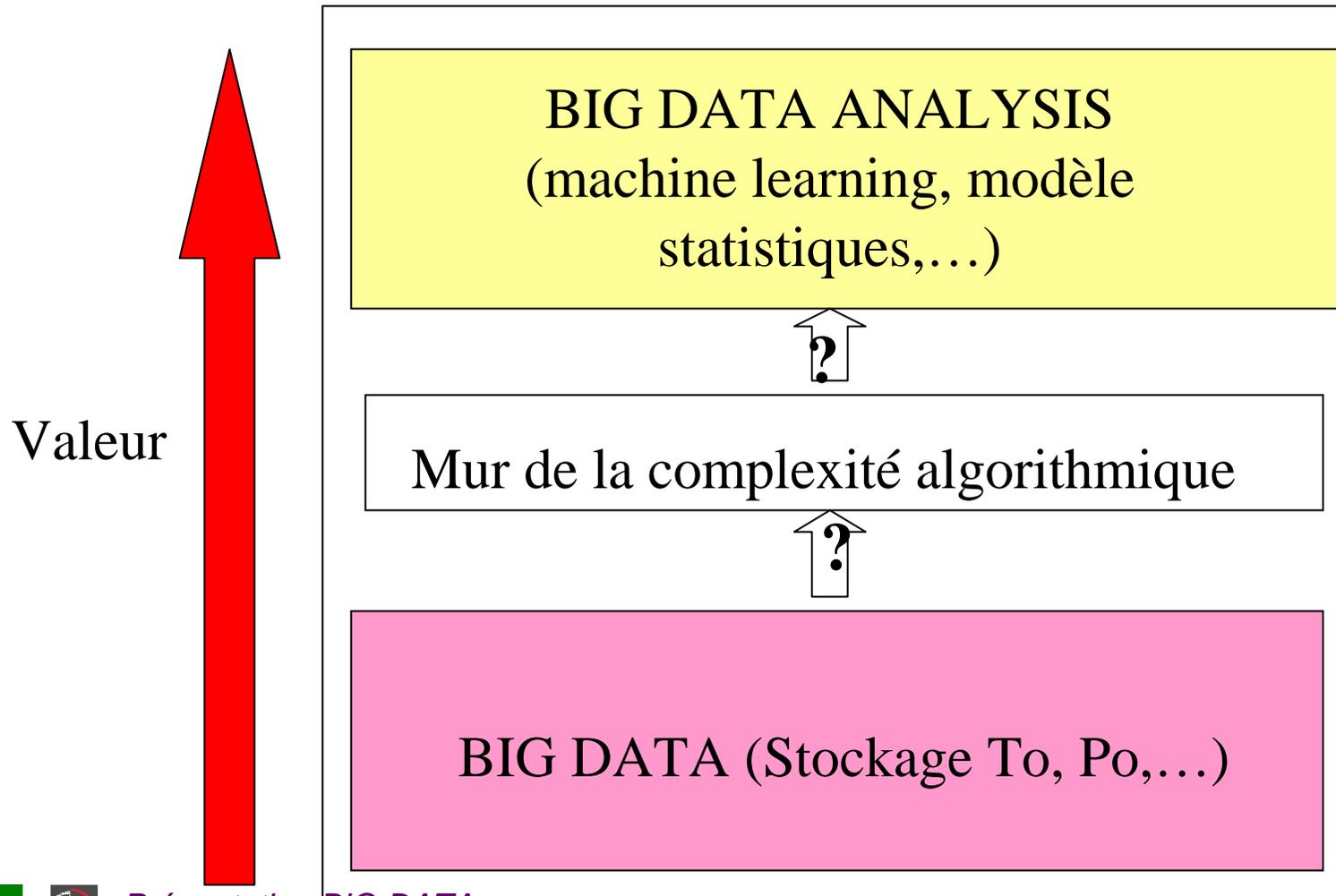
4. Big Data et la Statistique Publique : Emploi

Figure 5. Unemployment prediction for Germany, January 2004 – May 2009

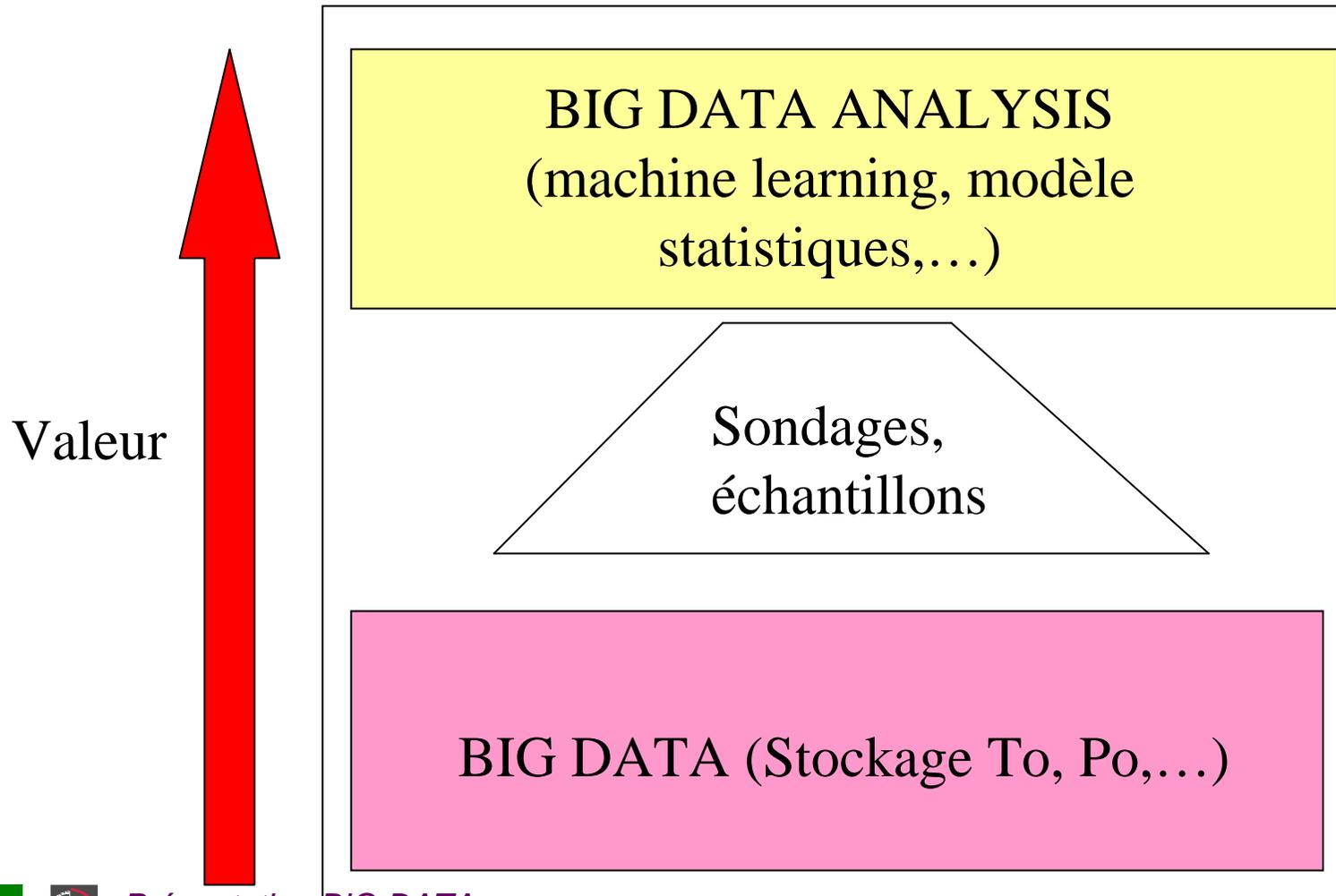


Source: Askitas and Zimmermann (2009).

4. La méthodologie de la Statistique Publique, avenir du Big Data Analysis ?



4. La méthodologie de la Statistique Publique, avenir du Big Data Analysis ?



4. Discussion
