

Le déluge de la big data appelle une *science of data* au-delà de la *data(-based) science*

Présentation à l'Institut de l'Économie

Frédéric Lefebvre-Naré

T16013

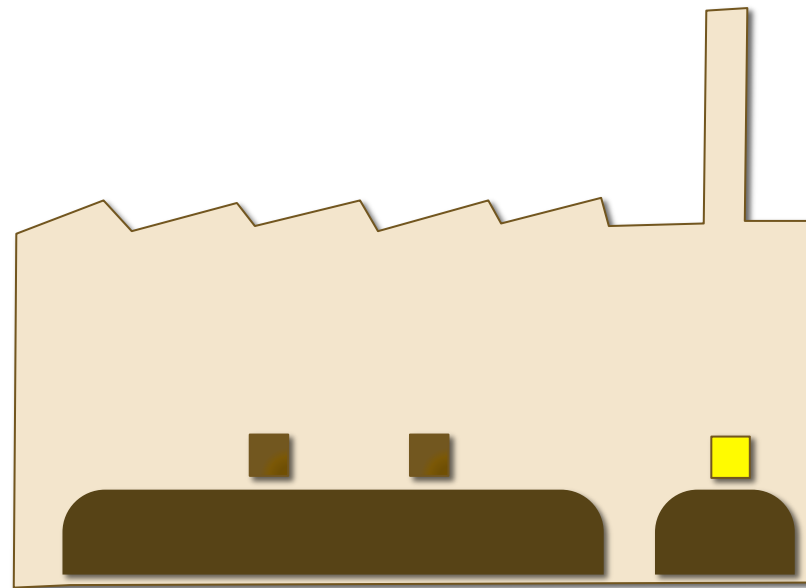
10 juin 2016

La big data selon 43 hommes et femmes de l'art

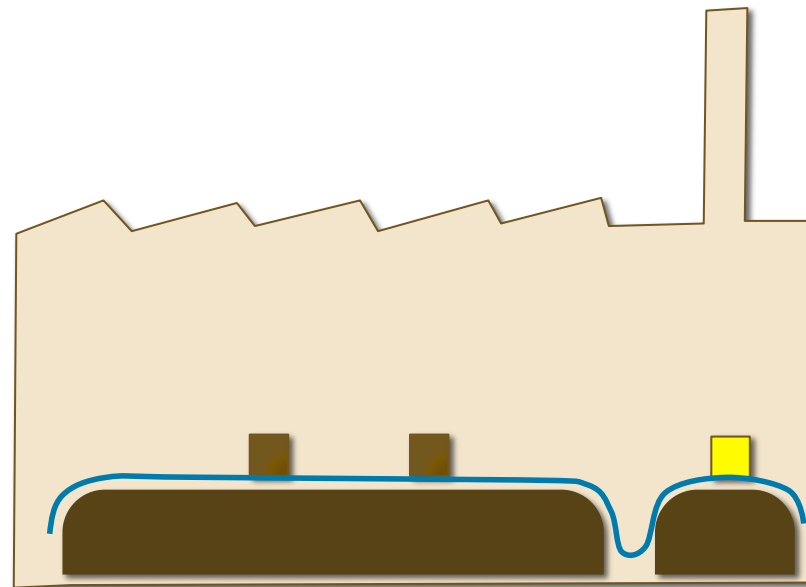
Accent	#	Nuances	Nouveau ?
Une rupture par rapport à la donnée qui n'était pas <i>big</i>	18	<ul style="list-style-type: none"> Nouvelles technologies Nouvelles compétences 	+ « 3V... » incrémental ?
Utilisation de données (de faits) dans la décision (humaine)	10	<ul style="list-style-type: none"> Analytics : traitements Dataviz : interface humaine 	+/- (réinventions) <i>data science</i>
Stockage et traitement continu de (beaucoup de) <i>data</i>	8	<ul style="list-style-type: none"> Capteurs Traitements temps réel Actionneurs / API 	+++
Nouvelle culture	5	<ul style="list-style-type: none"> Nouveaux enjeux éthiques Nouvelles opportunités 	+/-- (réinventions)
L'absence de l'être humain : décisions prises par des algorithmes	2	<ul style="list-style-type: none"> Aussi un changement culturel 	+

Dutcher, J. (2014). "What is Big Data?," récupéré le 29 novembre 2014 de <http://datascience.berkeley.edu/what-is-big-data/>

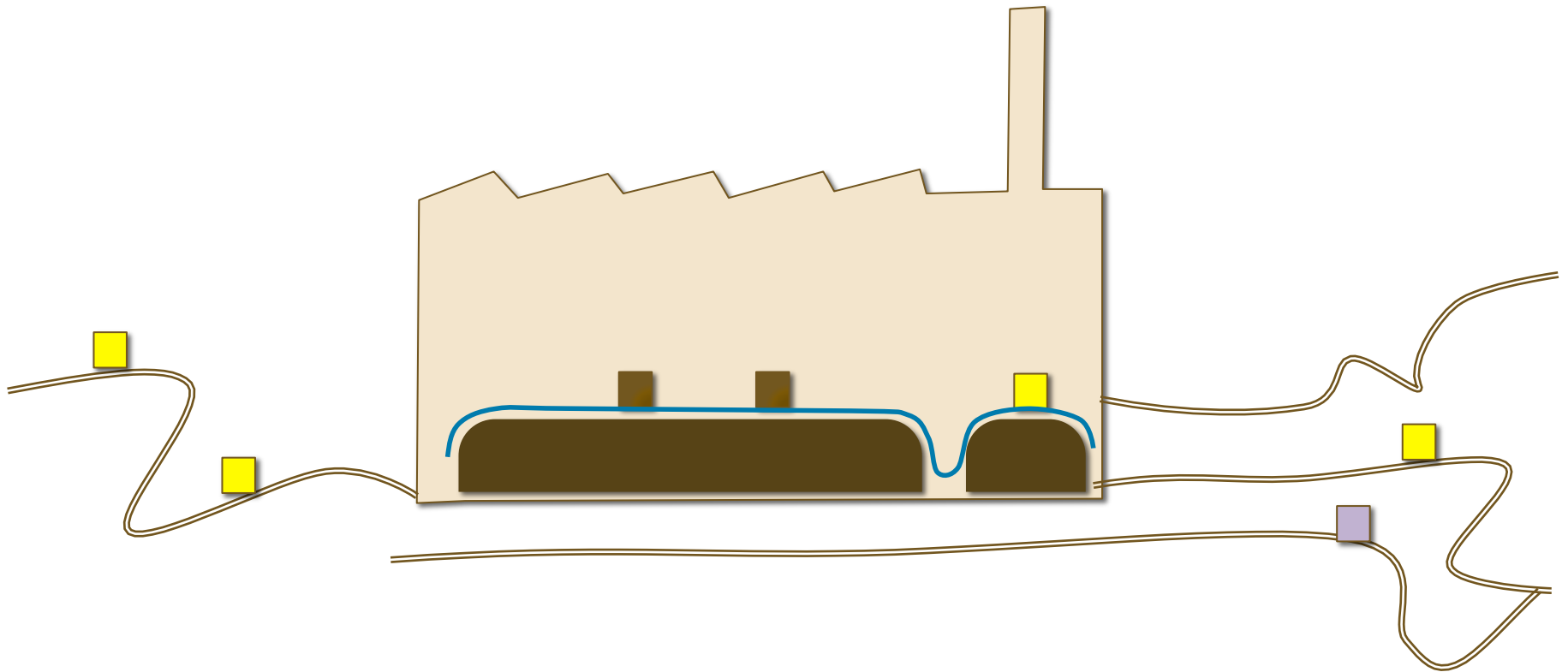
1^{ère} étape : l'organisation de l'espace et de la production. Plan, schéma de process.



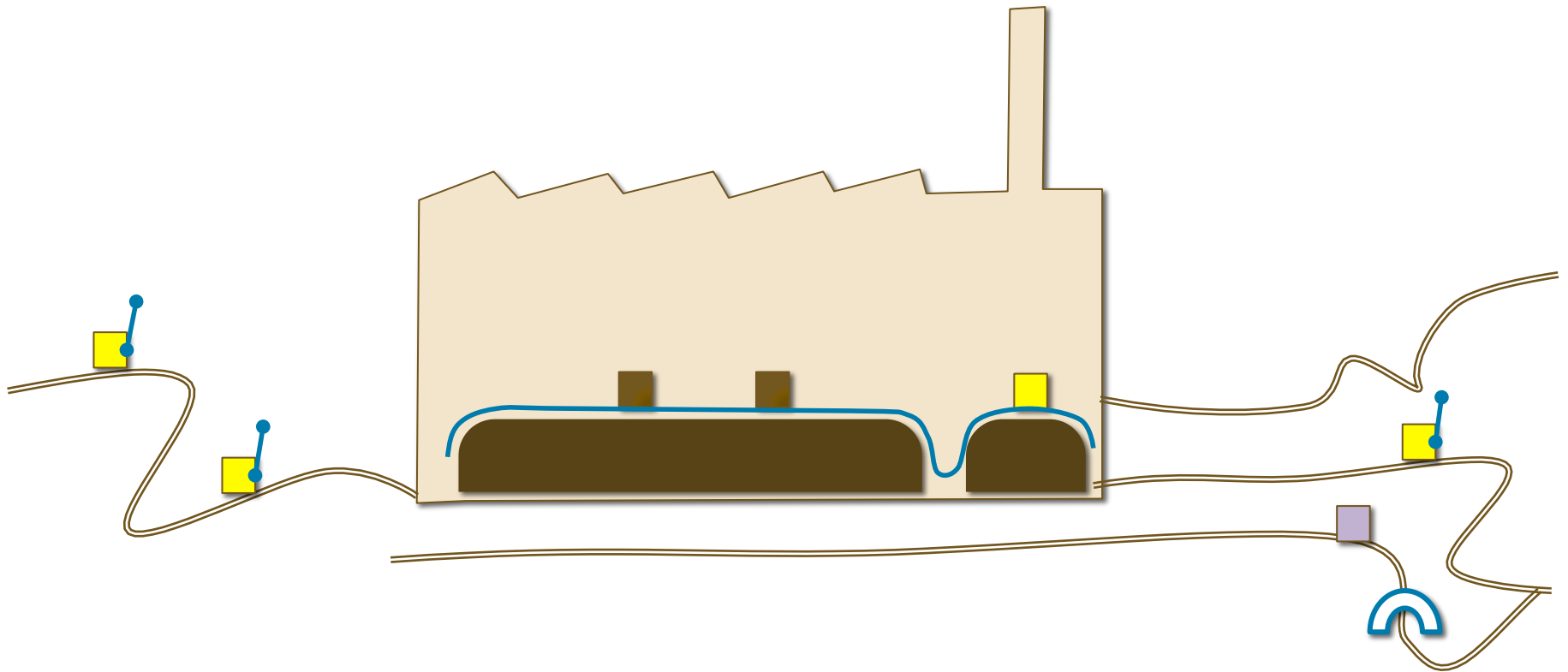
2^{ème} étape : informatisation de la production (ERP...) exigeant au passage de mieux modéliser la production, de détecter / combler les « trous »...



Puis le produit vit sa vie dans le monde réel (hors de l'environnement de production). Les produits concurrents aussi.



3^{ème} étape : tracer la vie du produit dans son univers concurrentiel.
P.ex. avec des enregistreurs GPS ou des enregistrements en points de passage.



Pour une *science of data*

- Les systèmes de gestion de bases de données relationnelles (SGBDR) décrivent bien des processus qu'on peut planifier à l'avance et qui s'exécuteront à peu près comme prévu. Les différents types de noSQL couvrent le reste du monde.
- La logique des SGBDR suppose de pouvoir anticiper la structure des données utiles avant de collecter celles-ci (avant toute action / tout traitement dans le SGBD), ce qui peut coûter des années de données perdues.
- Ne pas l'anticiper implique que la gestion de la qualité des données, l'unification des différentes sources de données, etc., vont être des cauchemars : sur-travail, méta-modélisation, pseudo-spécifications-ex-post, etc.
 - *garbage in, garbage out* —
- **La constitution de la data** doit donc
 - Décrire le mieux possible un monde que (l'entreprise) ne structure pas et qui évoluera,
 - Anticiper le mieux possible des traitements / utilisations futures de cette data.
- Cela demande une **connaissance générique et opérationnelle**
 - Du monde sous l'angle de sa mise en données
 - De l'utilisation des données

= *Science of data*