Intelligences et apprentissages dans les villes numériques

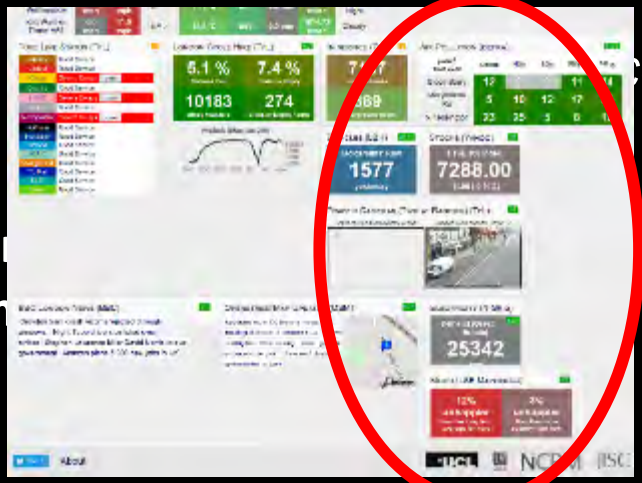
Dominique BOULLIER

Digital Humanities Institute
EPFL

Dijon
Juillet 2019

1. PLURALISME DES DIMENSIONS DE LA VILLE ET DES SMART CITIES

La boussole historique des villes



ertitude

co
et n

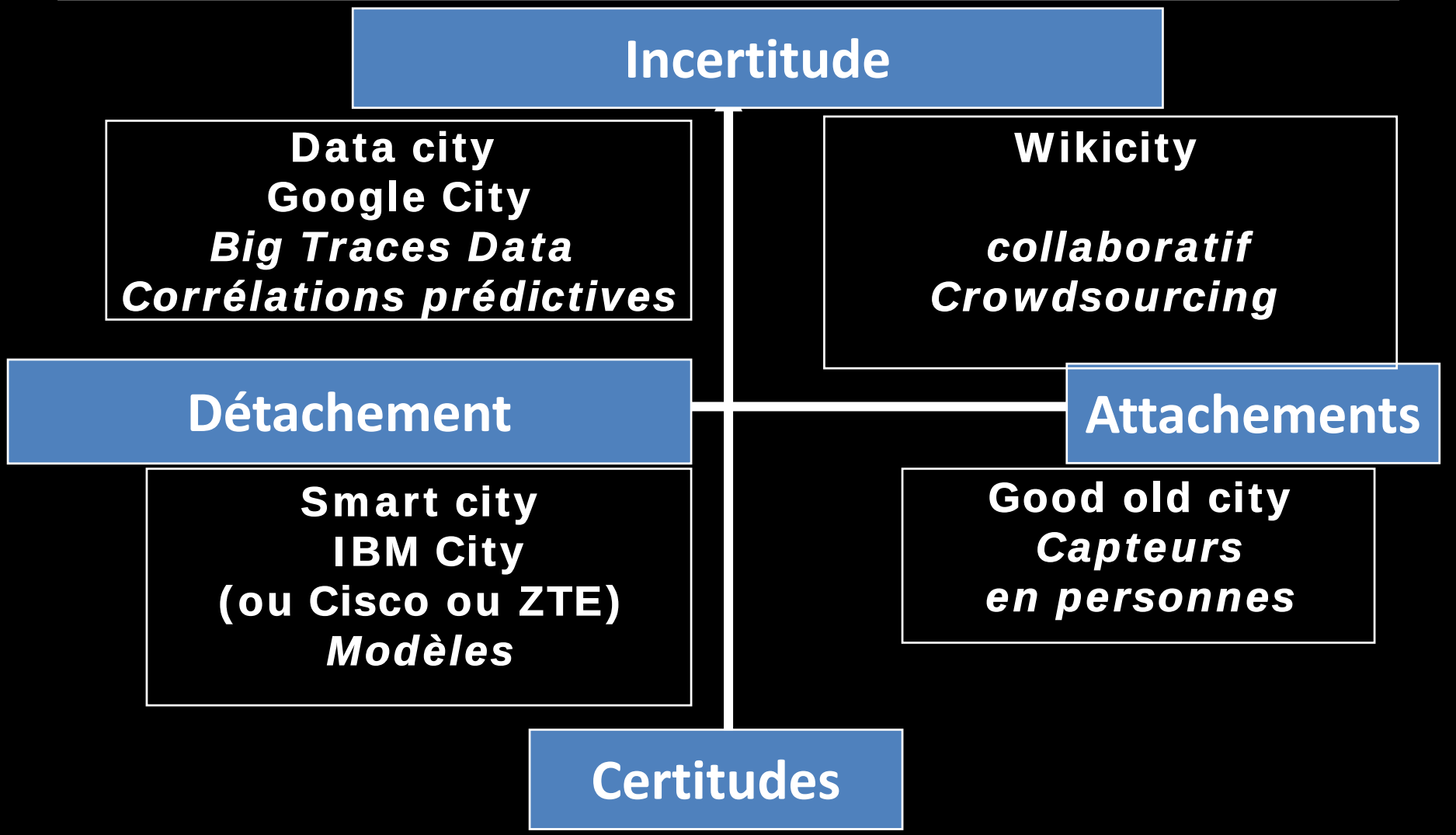
Quelle quatrième ville cosmopolitique?
Re-composer les autres villes

Détachement

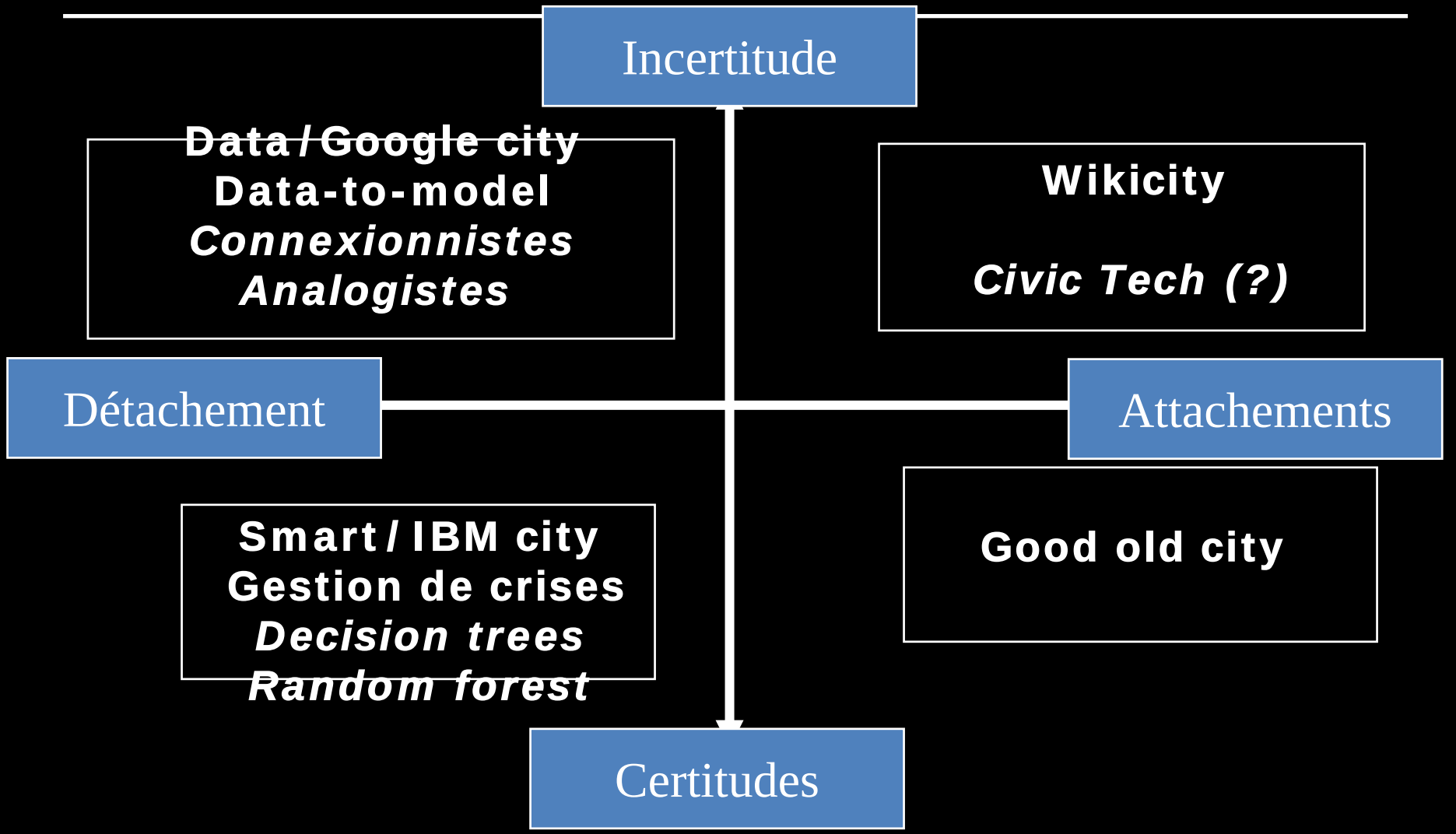
Attachements

ertitu

Politiques des données urbaines

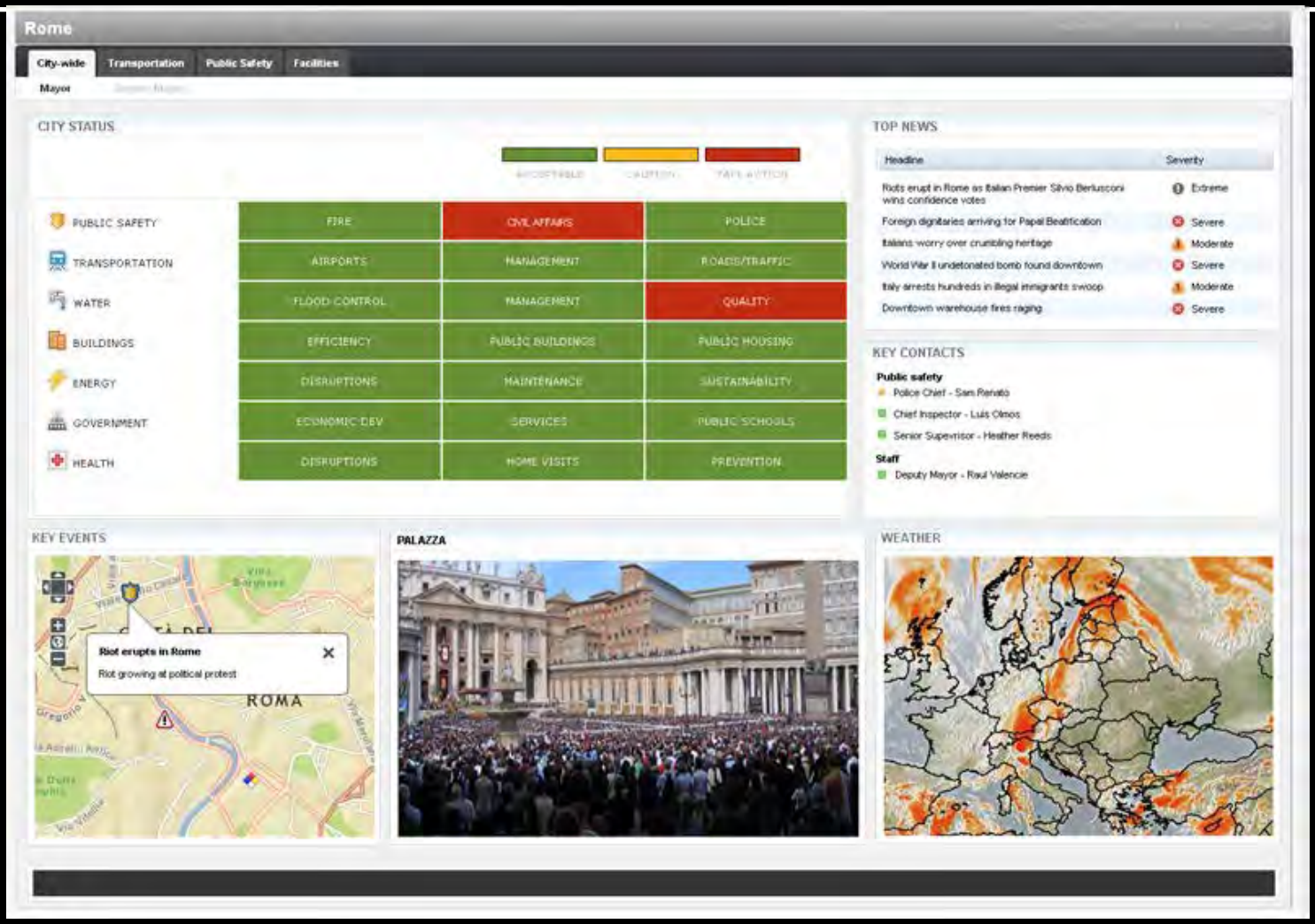


Politique des algorithmes des données urbaines et choix de Machine Learning



2. IBM CITY : LES MODÈLES

City Command Center IBM




CENTRO DE OPERAÇÕES
PREFEITURA DO RIO

Les nouvelles villes nouvelles comme modélisation absolue

POPULAR SCIENCE
WANT MORE?
PREMIUM

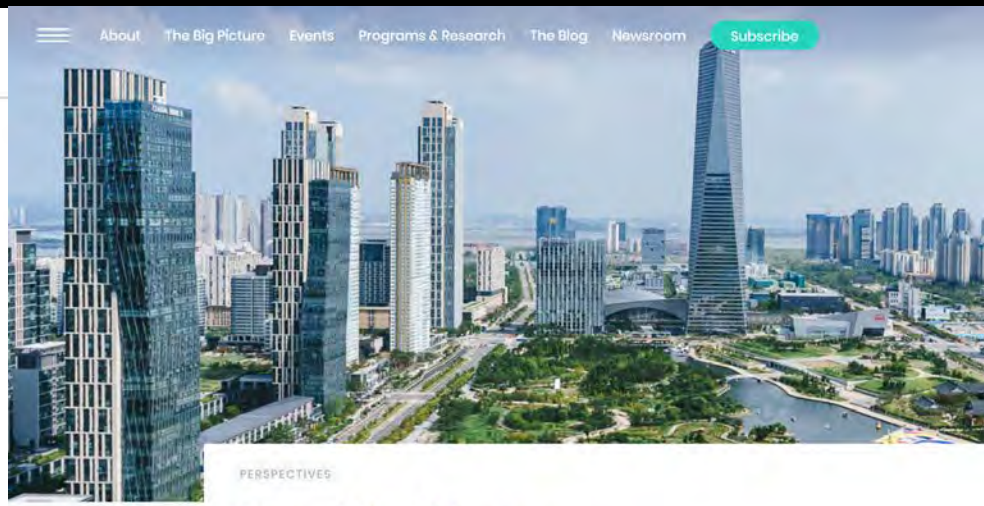
POPSCI EXPEDITIONS

ENVIRONMENT

Is Masdar City a ghost town or a green lab?

Tour Abu Dhabi's ambitious, and incomplete, eco-utopia.

By Molly McArdle April 24, 2018



PERSPECTIVES

South Korea Conceptualizes the Ultimate Smart City

Model to data:

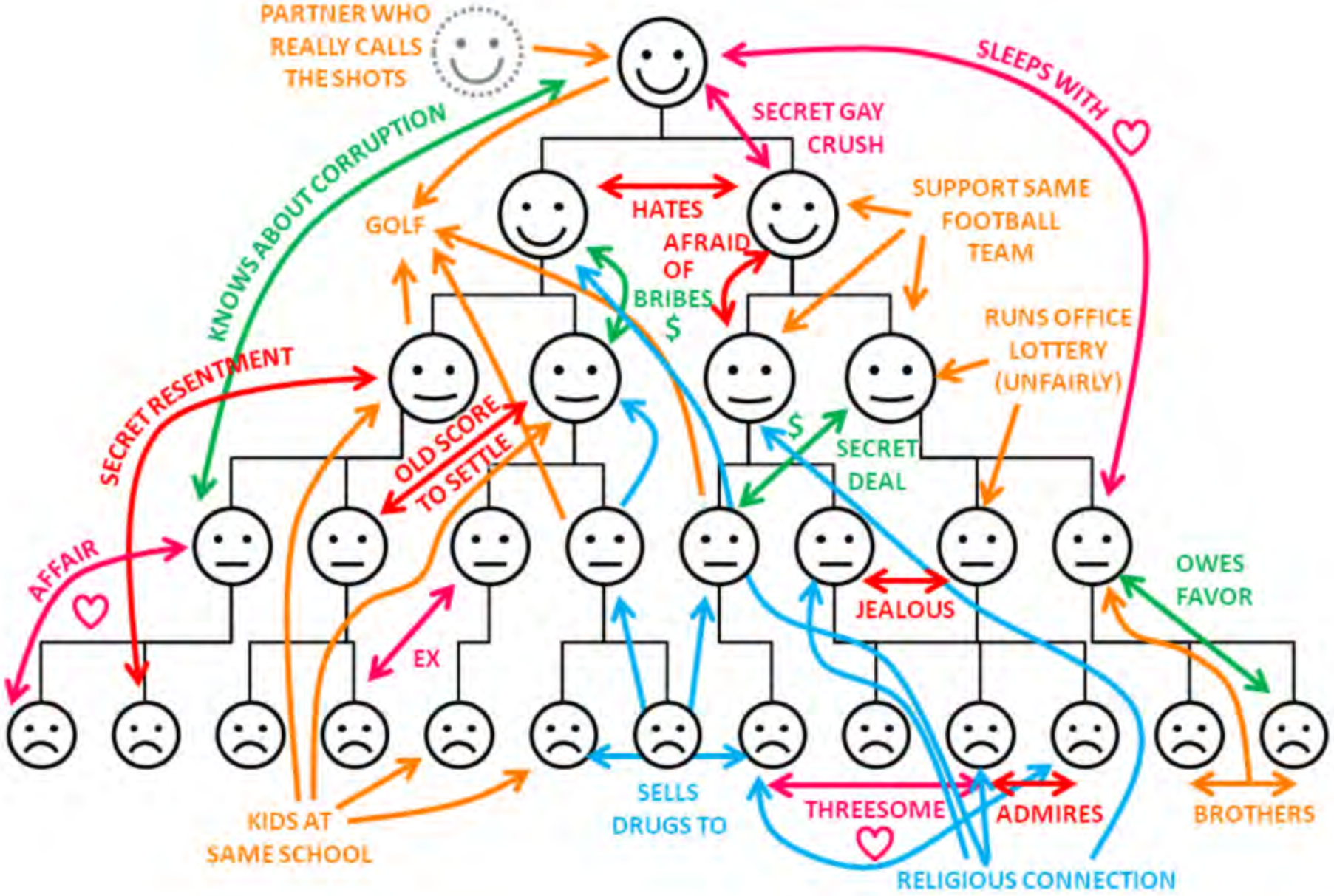
Modèles du monde

Les ontologies

faciles à critiquer

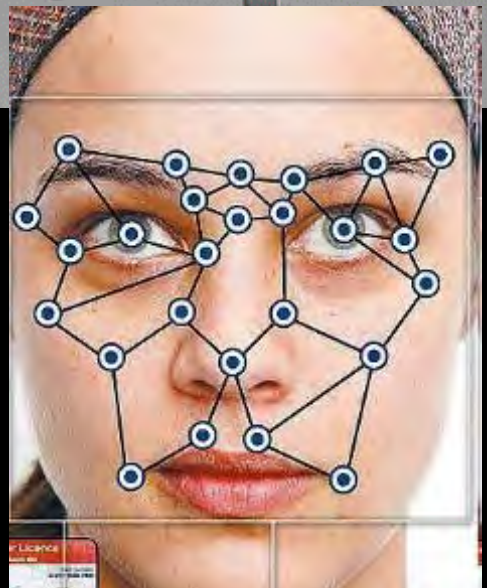
par les sciences sociales

L'époque de l'image fixe

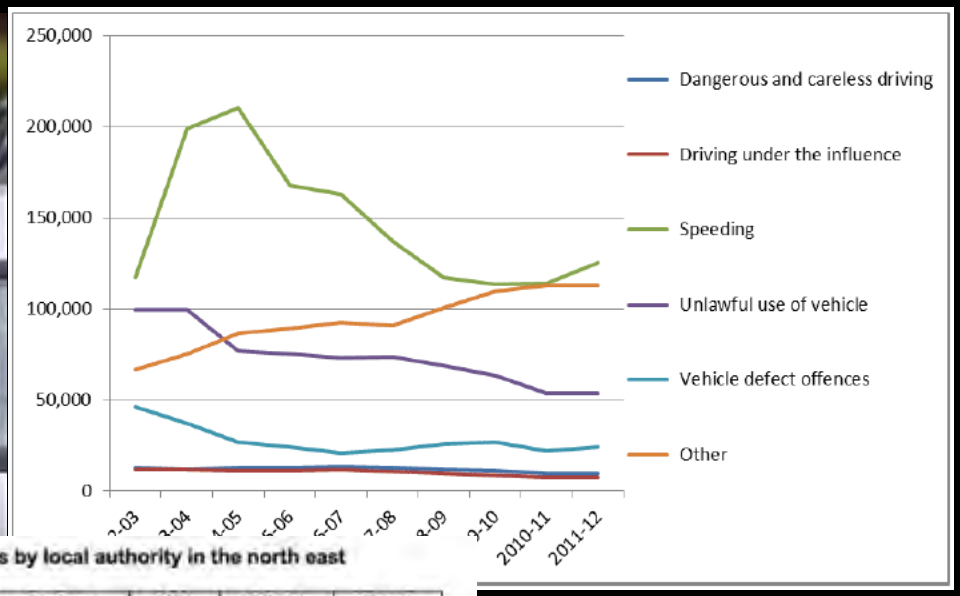

PARTNER WHO REALLY CALLS THE SHOTS
SLEEPS WITH
SECRET GAY CRUSH
KNOWS ABOUT CORRUPTION
HATES
SUPPORT SAME FOOTBALL TEAM
GOLF
AFRAID OF
BRIBES $
RUNS OFFICE LOTTERY (UNFAIRLY)
SECRET RESENTMENT
OLD SCORE TO SETTLE
$
SECRET DEAL
AFFAIR
OWES FAVOR
JEALOUS
EX
KIDS AT SAME SCHOOL
SELLS DRUGS TO
THREESOME
ADMIRES
BROTHERS
RELIGIOUS CONNECTION


Model-to-data
Modèles d'apprentissage
L'apprentissage permet de réviser à partir de tout input
L'époque de l'image animée

Toujours plus de capteurs
Des modèles toujours plus optimisés
mais quel design organisationnel?



Le *design organisationnel* du traitement des données dans la traffic police anglaise

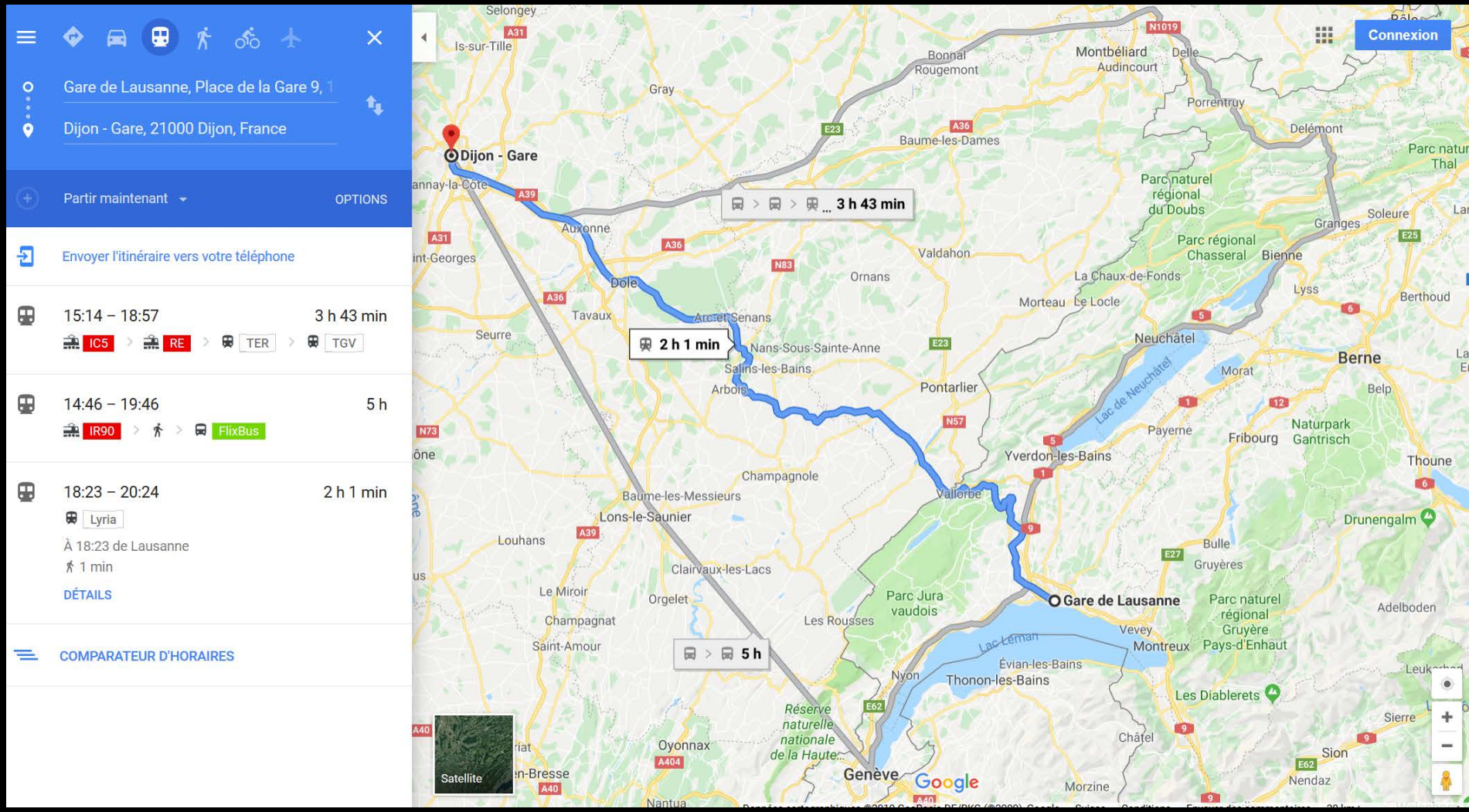


Number and severity of young driver casualties by local authority in the north east

Local Authority	Fatal	Serious	KSI	Slight	Total
County Durham	16	102	118	789	**907**
Darlington	2	13	15	171	**186**
Gateshead	2	34	36	348	**384**
Hartlepool	0	18	18	115	**133**
Middlesbrough	0	20	20	225	**245**
Newcastle upon Tyne	0	45	45	471	**516**
North Tyneside	1	17	18	277	**295**
Northumberland	7	57	64	580	**644**
Redcar and Cleveland	3	34	37	159	**196**
South Tyneside	1	21	22	173	**195**
Stockton-on-Tees	1	40	41	230	**271**
Sunderland	5	39	44	433	**477**
North East Total	**38**	**440**	**478**	**3,971**	**4,449**

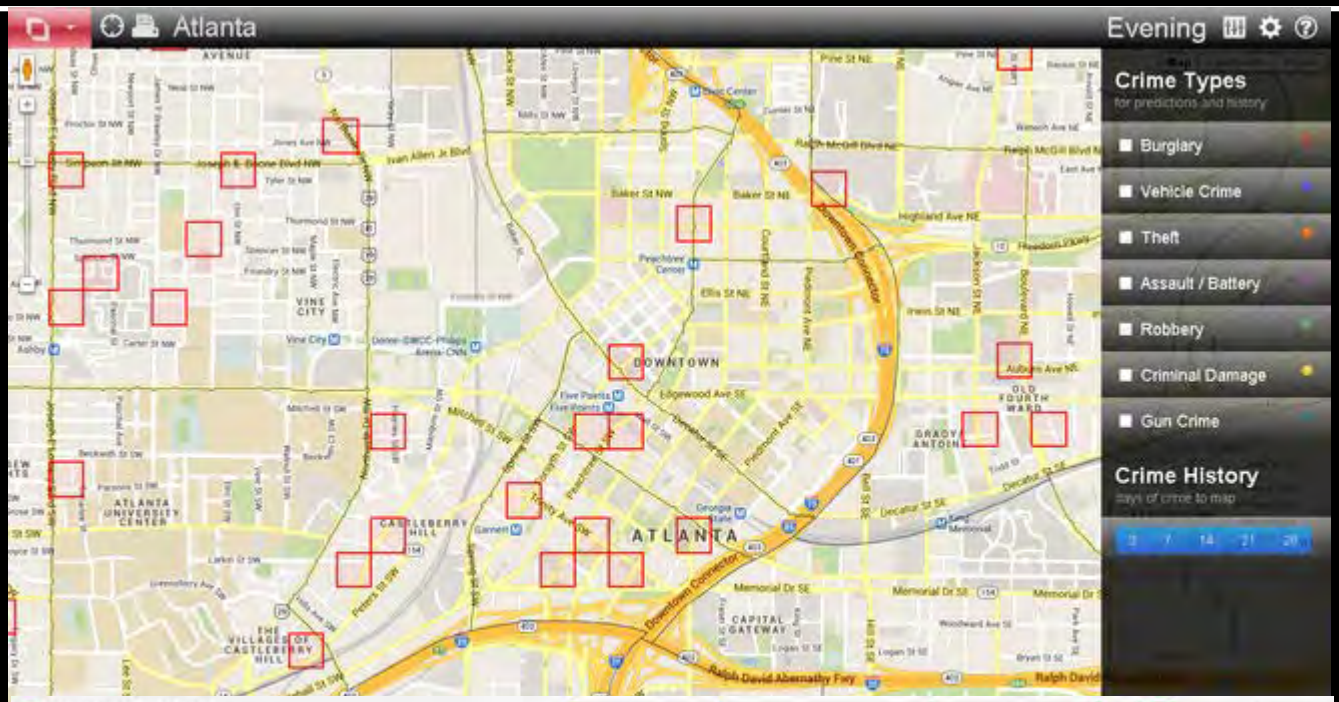
3. GOOGLE CITY :
VÉLOCITÉ DES TRACES

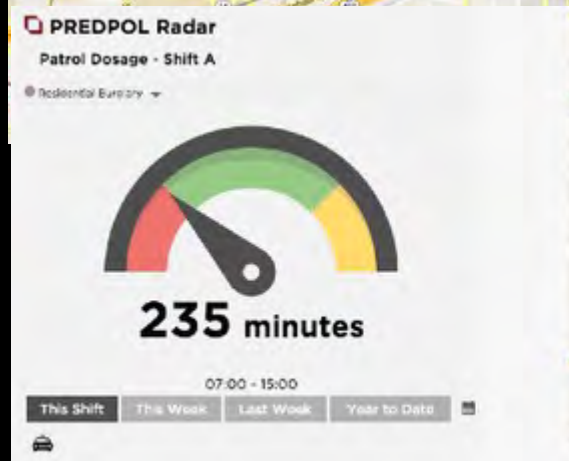
L'empire de Google maps
Data-to-model
Approche par les flux de données

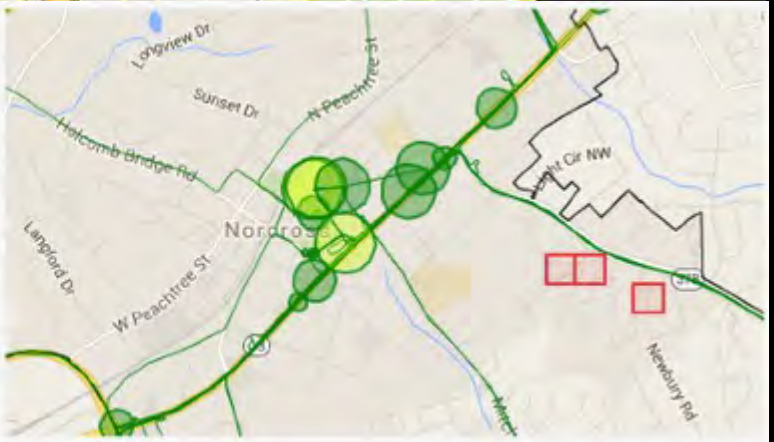


Mais qui apprend et qui désapprend?

Police prédictive: boite noire ou machine learning responsable







12:00:00-04:00	LARCENY (OTHER)	-83.116	42.3312	01/01/11T12
12:00:00-04:00	LARCENY (OTHER)	-83.1406	42.401	01/01/11T12
0:00:00-04:00	INFORMATION	-83.1484	42.4222	01/01/11T0:
0:00:00-04:00	INFORMATION	-83.138	42.4134	01/01/11T0:
0:00:00-04:00	TELEPHONE USED FOR HARASSM	-83.2257	42.3534	01/01/11T0:
0:00:00-04:00	DAMAGE TO PROPERTY - PRIVAT	-83.1265	42.3706	01/01/11T0:
0:00:00-04:00	INFORMATION	-83.0952	42.362	01/01/11T0:

HunchLab European Overview

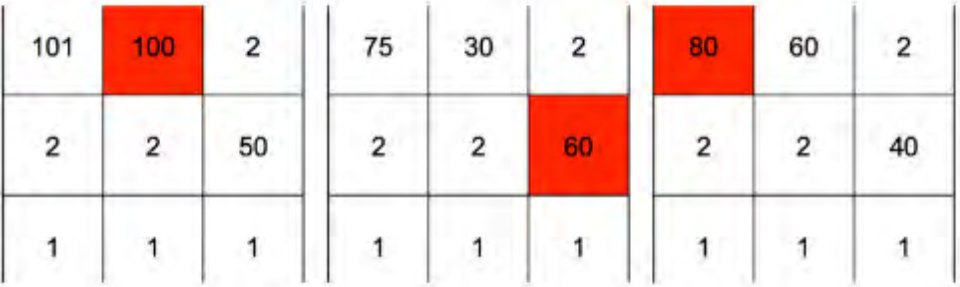
Have you wanted to explore a crime forecasting tool but hesitated because you were unsure of how it could work for a European law enforcement agency? In this webinar our team discusses how HunchLab can support the European market by providing a translated user interface and data residency within the EU.

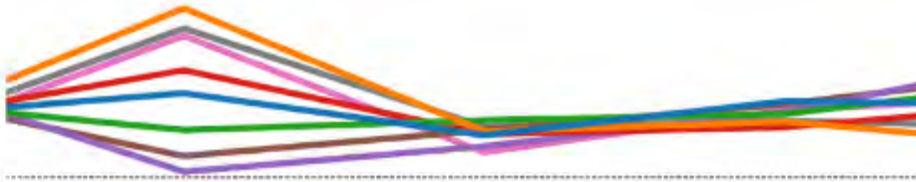
Recording

Guidelines for Data Sets in HunchLab

Given that HunchLab can use many different types of data within its models, how do you prioritize the data to use? This webinar outlines the types of data that HunchLab can use and ways to think about prioritizing data sets.

Recording





Allocation Engine

Is the best allocation of resources always going to the highest risk locations? This introduces the concept of metered dosage to improve HunchLab recommendations.

Recording

Insights from HunchLab: Comparing Crime Correlates Across Four US Cities

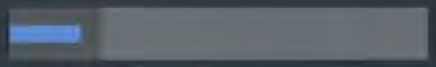
What can we learn from the variables used to predict high risk? This webinar uses HunchLab to build predictive models across four U.S. cities using the same data sets and then compares the resulting models.

Arrival Time (avg)

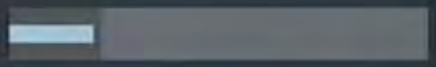
6m49s
Agency Value

6m12s
Local Value

Total Time (avg)



1h30m
Local Value



1h46m
Agency Value

Response Time (avg)

58m3s
Agency Value

43m30s
Local Value

Scene Time (avg)

47m23s ▲9.3%

48m32s
Agency Value

Event Count

3.7K
Agency Value

1.3K
Local Value

Dispatch Time (avg)

37m19s ▲3.6%

51m16s
Agency Value

What modeling methods does HunchLab use to generate forecasts of crime?

In some ways, the model building process in HunchLab mimics the thought process of an experienced analyst. For instance, consider asking an analyst to decide where to place patrol resources for a given upcoming time period. She may start by looking at where crimes have occurred in concentration previously and delineate hotspots of activity. Based on her past experience, she may know that during this particular time period, schools dismiss their students, which increases petty crimes around the schools in the neighborhood. She builds up many such layers of knowledge and balances these various concerns to form a plan. After the time period concludes, she may go back and look at where activities occurred to see if she can determine additional insights into the crime patterns to include in future plans. HunchLab incorporates machine learning concepts to help the software "think" like a crime analyst by imitating years of experience drawn from a police department's own data.

The primary model HunchLab currently uses is a stochastic gradient boosting machine (GBM) comprised of decision trees trained using the AdaBoost loss function. This model is built to forecast whether a crime event will occur or not in a given space-time raster cell (a binary outcome). The general way this model works is as follows:

- Begin by selecting a random portion of the training examples.
- Build a decision tree that separates examples of where crimes occurred from ones that didn't based upon the variables.
  - For instance, the first decision within the tree might be interpreted as: "if no event happened in the last year in this location, it is very unlikely for a crime to occur today". The decision tree then splits the examples into two sets: (1) where a crime occurred during the past year and (2) where no crimes occurred during the past year.
  - Within each set, the process repeats. For example, the next decision for the set of locations with crimes in the past year might be interpreted as: "if an event happened in the last week, it is more likely for one to occur today". This set of examples would again be split based upon this decision rule.
  - This process continues to build out a decision tree that describes why crimes occur where they do.
- The decision tree is then used to make predictions of how likely crimes are for each observation in the entire training data set.
- This completes one training iteration within the boosting machine.
- The modeling process then begins again.
  - We start by selecting another random portion of the training examples. This random sampling process is why the model is stochastic.
  - In this next iteration, we build another decision tree (in the same manner as above). This time, however, we build the tree to predict the errors from applying the first decision tree model to this new sample of observations. In this way we are attempting to correct our mistakes. This concept is called boosting.
  - We then use these two trees to make predictions across the entire data set.
  - As we conduct this process, we can keep track of how many training iterations within the machine have made incorrect predictions for each training example. We increase the importance (via weights) of observations that we continue to get wrong and decrease the importance of observations that we continue to get correct. This process is called adaptive boosting (AdaBoost). When we build the next decision tree, we tell it to focus on the observations that we continue to get wrong via these weights.
- Training iterations continue several hundred times. The resulting model represents tens of thousands of decision rules of why crimes occur where they do.
- We conduct this entire process several times, each time holding back a portion of the example data. We can then use each of these models to make predictions for this held-out set of data to

Mais l'interprétabilité est possible et rend le Deep Learning plus performant !

The Annals of Applied Statistics
2015, Vol. 9, No. 3, 1350–1371
DOI: 10.1214/15-AOAS848

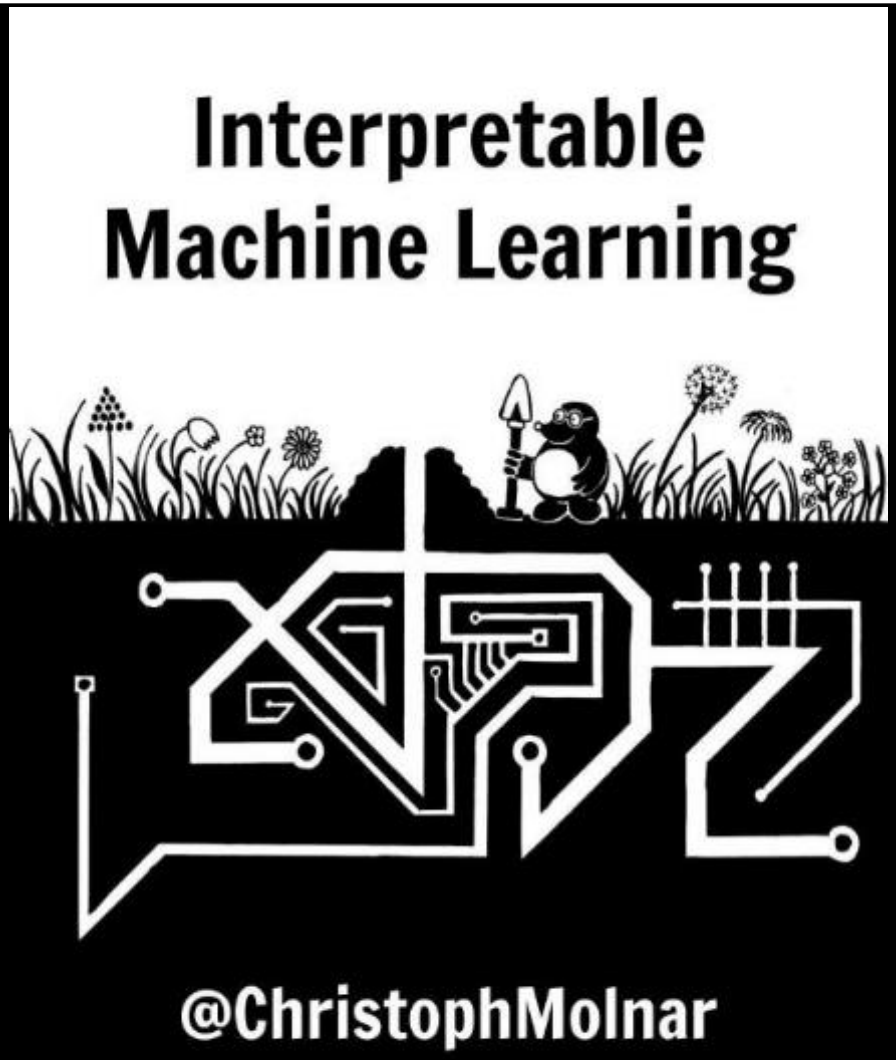
INTERPRETABLE CLASSIFIERS USING RULES AND BAYESIAN ANALYSIS: BUILDING A BETTER STROKE PREDICTION MODEL

BY BENJAMIN LETHAM[*,1], CYNTHIA RUDIN[*,1], TYLER H. MCCORMICK[†,2] AND DAVID MADIGAN[‡,3]

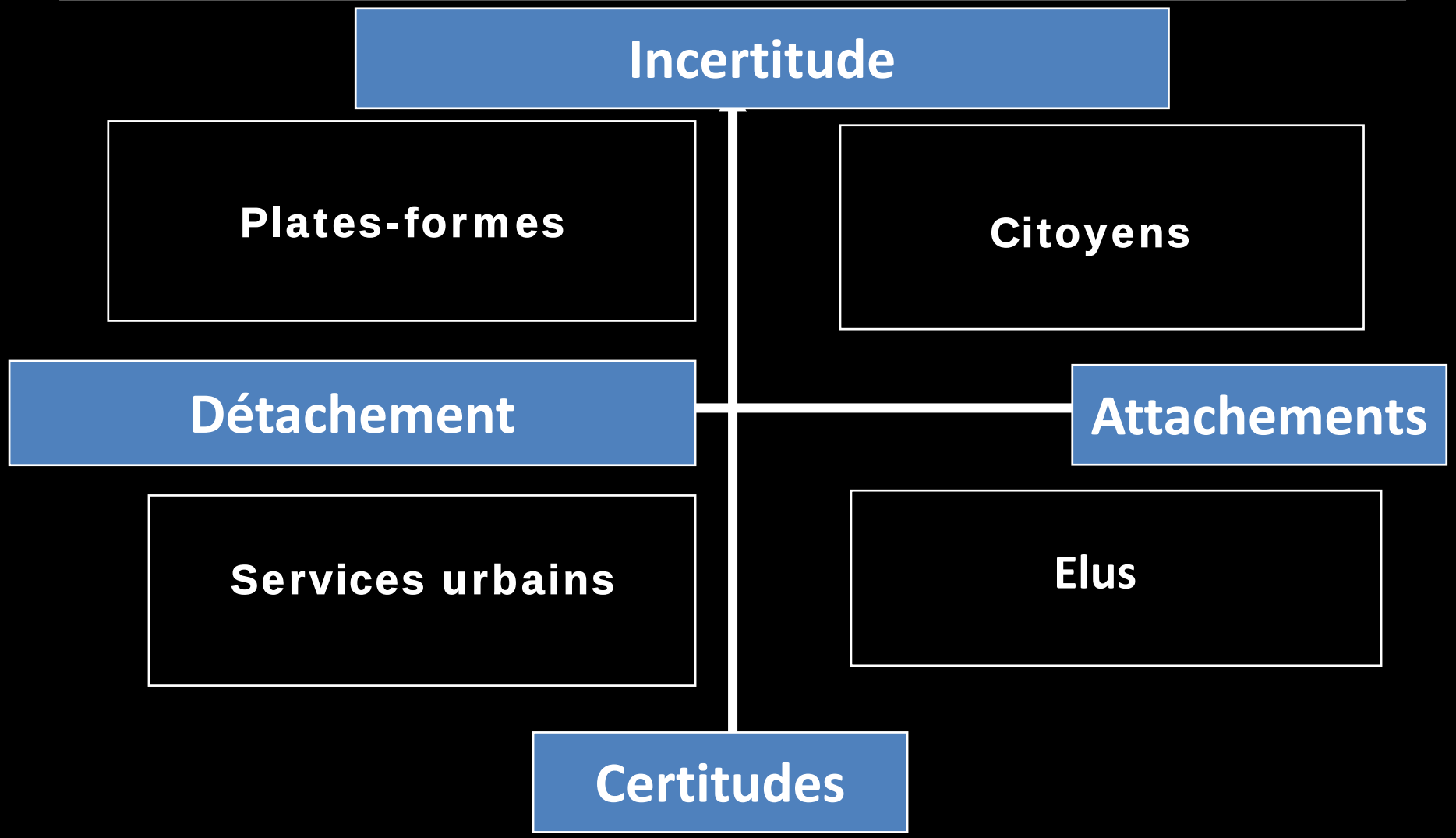
Massachusetts Institute of Technology[*], *University of Washington*[†] *and Columbia University*[‡]

We aim to produce predictive models that are not only accurate, but are also interpretable to human experts. Our models are decision lists, which consist of a series of *if*...*then*... statements (e.g., *if high blood pressure, then stroke*) that discretize a high-dimensional, multivariate feature space into a series of simple, readily interpretable decision statements. We introduce a generative model called Bayesian Rule Lists that yields a posterior distribution over possible decision lists. It employs a novel prior structure to encourage sparsity. Our experiments show that Bayesian Rule Lists has predictive accuracy on par with the current top algorithms for prediction in machine learning. Our method is motivated by recent developments in personalized medicine, and can be used to produce highly accurate and interpretable medical scoring systems. We demonstrate this by producing an alternative to the $CHADS_2$ score, actively used in clinical practice for estimating the risk of stroke in patients that have atrial fibrillation. Our model is as interpretable as $CHADS_2$, but more accurate.

Interprétabilité technique sans interprétabilité sociale?

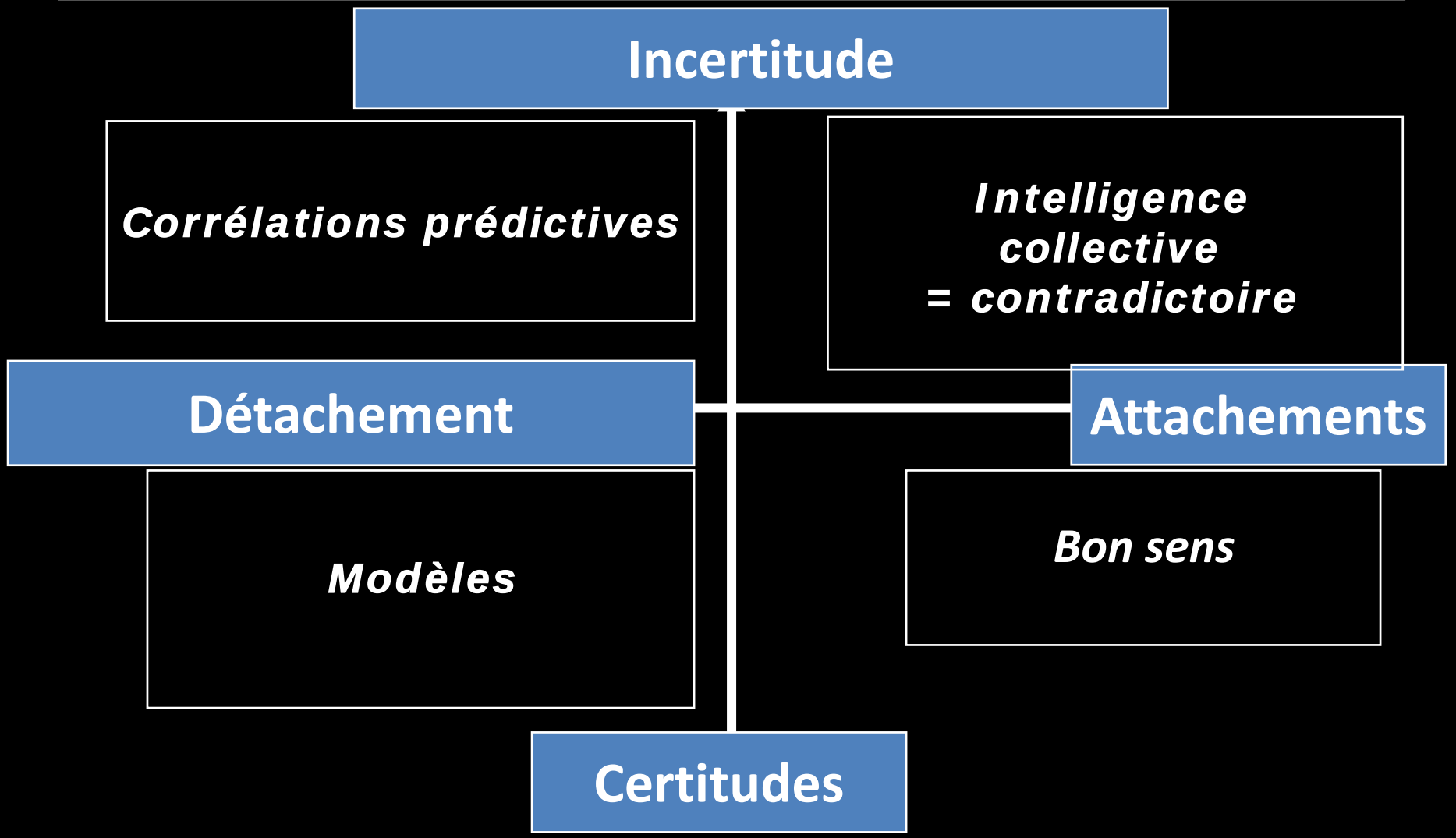


ApprentissageS
« Division of learning » (Zuboff)
Les effets de dépossession (cf. PPP)

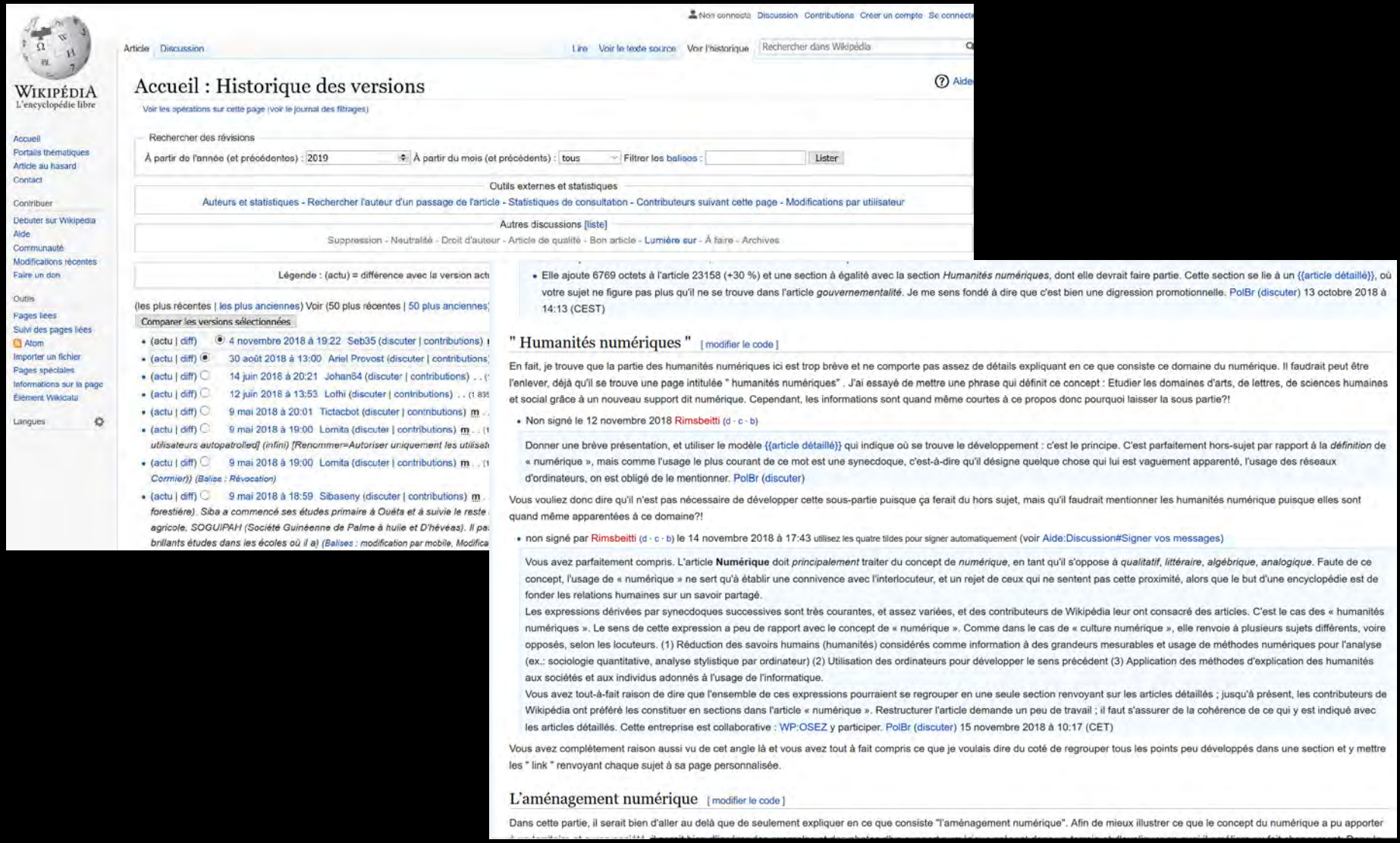


4. WIKI CITY :
L’INTELLIGENCE COLLECTIVE

IntelligenceS



L'importance de l'historique et de la discussion: le contradictoire!



WIKIPÉDIA
L'encyclopédie libre

Article | Discussion | Lire | Voir le texte source | Voir l'historique | Rechercher dans Wikipédia

Accueil : Historique des versions

Voir les opérations sur cette page (voir le journal des filtrages)

Rechercher des révisions

À partir de l'année (et précédentes) : 2019 | À partir du mois (et précédents) : tous | Filtrer les balises : | Lister

Outils externes et statistiques

Auteurs et statistiques - Rechercher l'auteur d'un passage de l'article - Statistiques de consultation - Contributeurs suivant cette page - Modifications par utilisateur

Autres discussions [liste]

Suppression - Neutralité - Droit d'auteur - Article de qualité - Bon article - Lumière sur - À faire - Archives

Légende : (actu) = différence avec la version actu

(les plus récentes | les plus anciennes) Voir (50 plus récentes | 50 plus anciennes)

Comparer les versions sélectionnées

- (actu | diff) 4 novembre 2018 à 19:22 Seb35 (discuter | contributions)
- (actu | diff) 30 août 2018 à 13:00 Ariel Provost (discuter | contributions)
- (actu | diff) 14 juin 2018 à 20:21 Johan84 (discuter | contributions)
- (actu | diff) 12 juin 2018 à 13:53 Lofhi (discuter | contributions)
- (actu | diff) 9 mai 2018 à 20:01 Tictacbot (discuter | contributions) m
- (actu | diff) 9 mai 2018 à 19:00 Lomita (discuter | contributions) m
- (actu | diff) 9 mai 2018 à 19:00 Lomita (discuter | contributions) m
- (actu | diff) 9 mai 2018 à 18:59 Sibaseny (discuter | contributions) m

- Elle ajoute 6769 octets à l'article 23158 (+30 %) et une section à égalité avec la section *Humanités numériques*, dont elle devrait faire partie. Cette section se lie à un {{article détaillé}}, où votre sujet ne figure pas plus qu'il ne se trouve dans l'article *gouvernementalité*. Je me sens fondé à dire que c'est bien une digression promotionnelle. PolBr (discuter) 13 octobre 2018 à 14:13 (CEST)

" Humanités numériques " [modifier le code]

En fait, je trouve que la partie des humanités numériques ici est trop brève et ne comporte pas assez de détails expliquant en ce que consiste ce domaine du numérique. Il faudrait peut être l'enlever, déjà qu'il se trouve une page intitulée " humanités numériques" . J'ai essayé de mettre une phrase qui définit ce concept : Etudier les domaines d'arts, de lettres, de sciences humaines et social grâce à un nouveau support dit numérique. Cependant, les informations sont quand même courtes à ce propos donc pourquoi laisser la sous partie?!

- Non signé le 12 novembre 2018 Rimsbeitti (d · c · b)

  Donner une brève présentation, et utiliser le modèle {{article détaillé}} qui indique où se trouve le développement : c'est le principe. C'est parfaitement hors-sujet par rapport à la *définition* de « numérique », mais comme l'usage le plus courant de ce mot est une synecdoque, c'est-à-dire qu'il désigne quelque chose qui lui est vaguement apparenté, l'usage des réseaux d'ordinateurs, on est obligé de le mentionner. PolBr (discuter)

Vous vouliez donc dire qu'il n'est pas nécessaire de développer cette sous-partie puisque ça ferait du hors sujet, mais qu'il faudrait mentionner les humanités numérique puisque elles sont quand même apparentées à ce domaine?!

- non signé par Rimsbeitti (d · c · b) le 14 novembre 2018 à 17:43 utilisez les quatre tildes pour signer automatiquement (voir Aide:Discussion#Signer vos messages)

  Vous avez parfaitement compris. L'article **Numérique** doit *principalement* traiter du concept de *numérique*, en tant qu'il s'oppose à *qualitatif*, *littéraire*, *algébrique*, *analogique*. Faute de ce concept, l'usage de « numérique » ne sert qu'à établir une connivence avec l'interlocuteur, et un rejet de ceux qui ne sentent pas cette proximité, alors que le but d'une encyclopédie est de fonder les relations humaines sur un savoir partagé.
  Les expressions dérivées par synecdoques successives sont très courantes, et assez variées, et des contributeurs de Wikipédia leur ont consacré des articles. C'est le cas des « humanités numériques ». Le sens de cette expression a peu de rapport avec le concept de « numérique ». Comme dans le cas de « culture numérique », elle renvoie à plusieurs sujets différents, voire opposés, selon les locuteurs. (1) Réduction des savoirs humains (humanités) considérés comme information à des grandeurs mesurables et usage de méthodes numériques pour l'analyse (ex.: sociologie quantitative, analyse stylistique par ordinateur) (2) Utilisation des ordinateurs pour développer le sens précédent (3) Application des méthodes d'explication des humanités aux sociétés et aux individus adonnés à l'usage de l'informatique.
  Vous avez tout-à-fait raison de dire que l'ensemble de ces expressions pourraient se regrouper en une seule section renvoyant sur les articles détaillés ; jusqu'à présent, les contributeurs de Wikipédia ont préféré les constituer en sections dans l'article « numérique ». Restructurer l'article demande un peu de travail ; il faut s'assurer de la cohérence de ce qui y est indiqué avec les articles détaillés. Cette entreprise est collaborative : WP:OSEZ y participer. PolBr (discuter) 15 novembre 2018 à 10:17 (CET)

Vous avez complètement raison aussi vu de cet angle là et vous avez tout à fait compris ce que je voulais dire du coté de regrouper tous les points peu développés dans une section et y mettre les " link " renvoyant chaque sujet à sa page personnalisée.

L'aménagement numérique [modifier le code]

Dans cette partie, il serait bien d'aller au delà que de seulement expliquer en ce que consiste "l'aménagement numérique". Afin de mieux illustrer ce que le concept du numérique a pu apporter

Statut du public: utilisateurs-consommateurs

SeeClickFix
power to the community

Tools to help Communities help Themselves!

Login | Sign up

Citizens | community groups | government | media

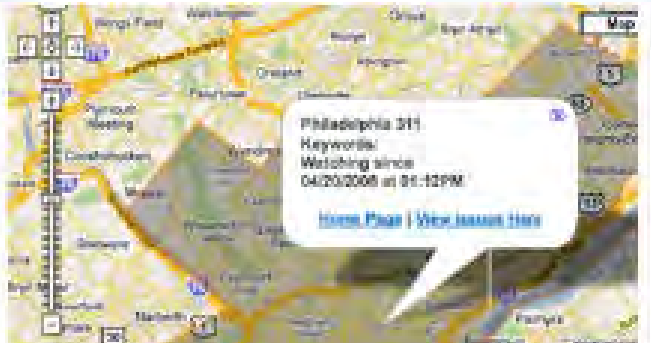
Free Tools for Governing

More Tools for Governing

- Transparent communication with citizens
- Speak to many once
- Free smartphone applications
- Open Data resources hosted by third party
- Deputize Citizen Inspectors
- Prioritize concerns by collective votes
- Facebook, Twitter and Blog Integration

Create a Watch Area

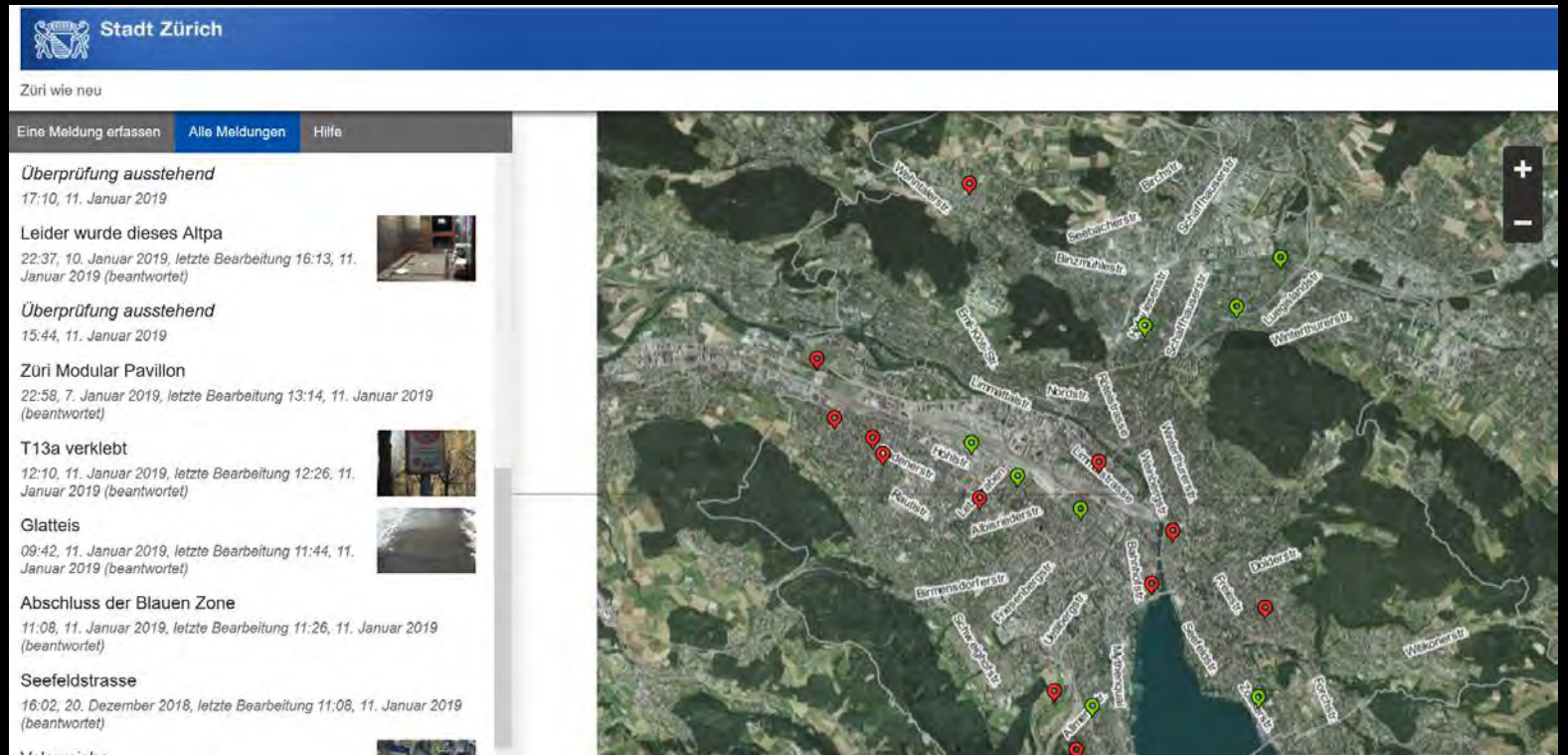
- Sign-up to receive alerts based on geography and keyword.
- Citizens can report to you by mobile web, iPhone app, SeeClickFix.com and from anywhere else SeeClickFix is embedded - like your website!
- Receive alerts by email or standard data resources such as GeoRSS, KML and JSON.



Statut du public: lanceur d'alerte permanent

FixMyStreet, FixMaVille, Zueri wie neu, Signalez-nous Yverdon-les-Bains, ActivezMontreal, etc.

Format Open 311



Statut du public: contributeur à la connaissance

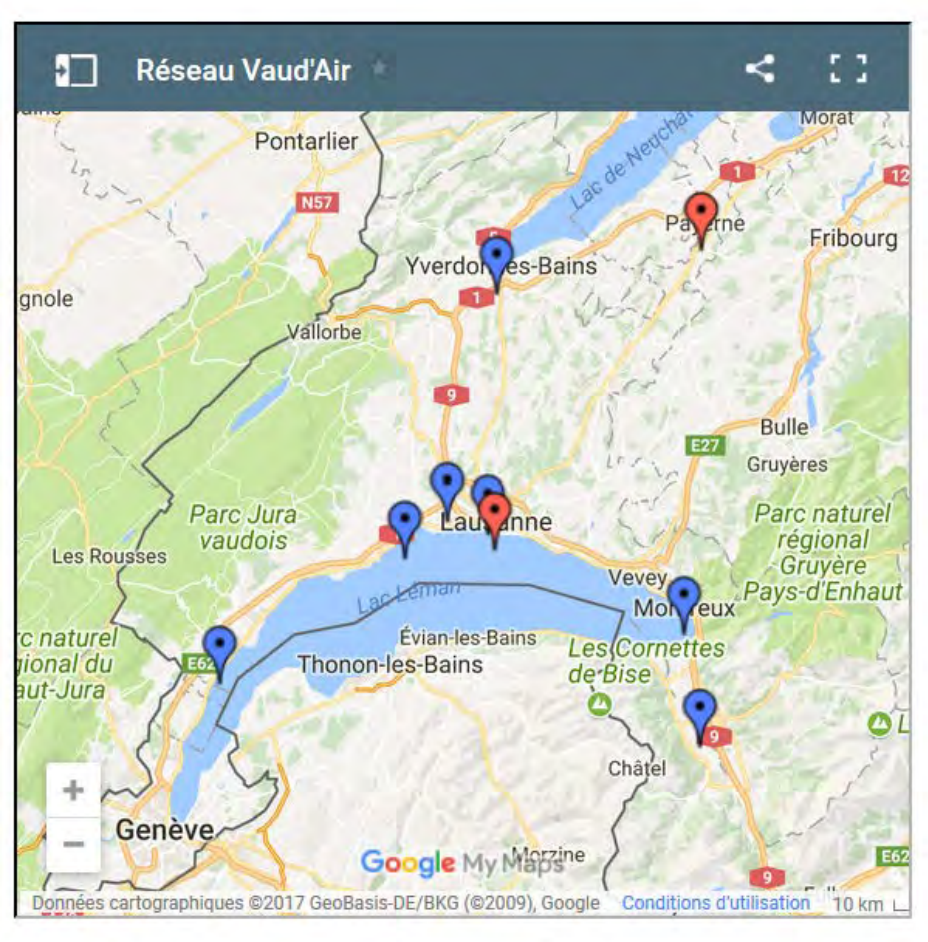
Un ancrage territorial défini par des pratiques

Quelques initiatives

- Relevé de pollutions urbaines

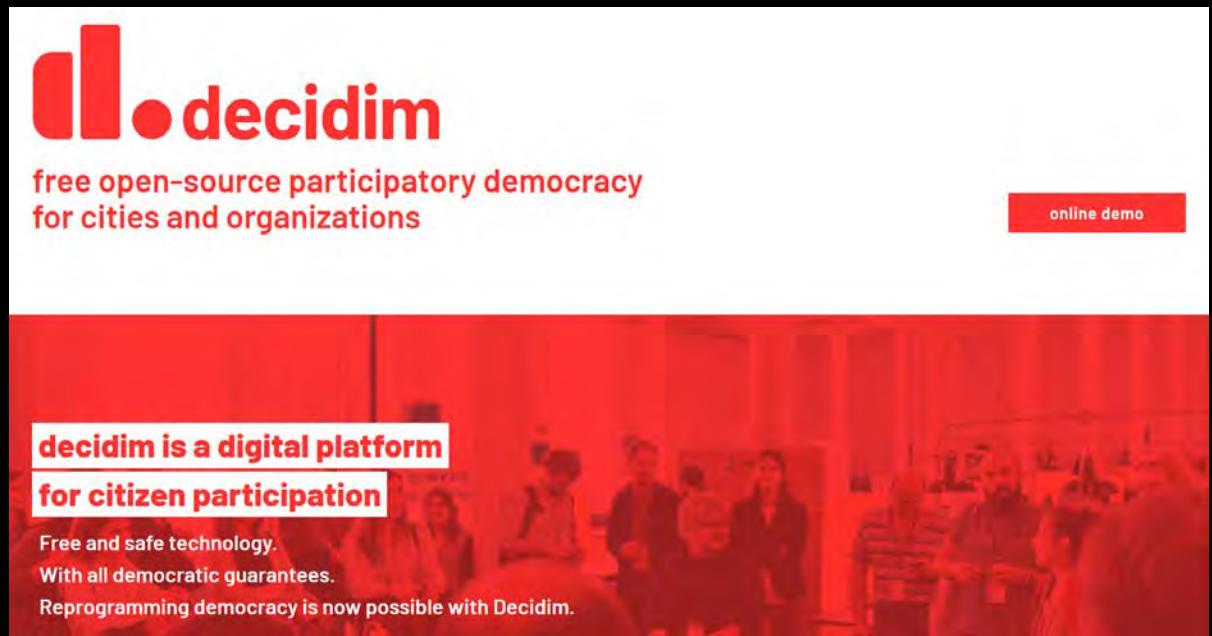
Rencontres SIG la lettre

Au-delà du règne des capteurs et des modèles/ systèmes de la démocratie des experts



Statut du public: les citoyens débattent

- Decidim Barcelona, MeuRio, Decide Madrid,
- Neighborland, Stadt Nürtringen, Fluicity,
- EngagingPlan/ EngagingApps (Urban Interactive Studio)



Civic Tech open source ou propriétaire = des apprentissages distribués différemment

- Open source

97Network, CaptainFact, Demodyne, Bluenove, Framaligue, Eurofordocs, Mieux voter, Collectif Tiriad, Tribute, Voxe (pour son comparateur), 7 Milliards d'Urbanistes.

Propriétaires

- Cap Collectif, Citizaction, CitizenLab, Civocracy, D21, Datagora, Fluicity, ID City, Vooter, Zemus

PROJETER UNE VILLE NUMÉRIQUE =>

PENSER:

1/ LE DESIGN ORGANISATIONNEL,

2/ LA DISTRIBUTION DES APPRENTISSAGES

3/ LES TYPES D'INTELLIGENCE MIS EN OEUVRE