

1 Nouveaux métabolismes, nouveaux outils

Les nouveaux métabolismes urbains de la Smart City étant basés sur des échanges de **données massives**, il serait intéressant que l'architecte soit en capacité de les exploiter. Dans ce cas, la question de la **capacité des outils actuels de l'architecte** à faire face à ce nouveau besoin devient cruciale.

1.1 L'Open Data : vers une profusion d'informations sur l'existant...

L'**Open Data** désigne avant tout une stratégie d'ouverture du plus grand nombre de base de données possibles au public, et ce pour des raisons d'intérêt général. À l'instar des autres mouvements du même type, tel que « l'Open Source », le **traitement et la rediffusion** des données sont autorisés, voire même encouragés comme c'est le cas de la part du gouvernement français² à travers des initiatives telles qu'**Etalab** : « les données partagées trouvent des ré-utilisateurs qui les intègrent dans de nouveaux services à forte valeur ajoutée économique ou sociale. ».

5

Ainsi, il existe aujourd'hui des bases de données en pleine expansion concernant l'**existant** à l'échelle urbaine, tout en commençant à se préciser à l'échelle du **bâtiment**.

Deux plateformes en particulier proposent de tels jeux de données : l'**APUR**³ (Atelier Parisien de l'Urbanisme) et l'**IGN**⁴ (Institut National Géographique).

²L'ouverture des données publiques. *Gouvernement.fr* [en ligne]. [s. d.]. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>.

³Atelier Parisien d'Urbanisme [en ligne]. [s. d.]. [Consulté le 4 février 2021]. Disponible à l'adresse : <https://opendata.apur.org/>.

⁴Géoservices | Accéder au téléchargement des données libres IGN [en ligne]. [s. d.]. [Consulté le 15 septembre 2020]. Disponible à l'adresse : <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html>.

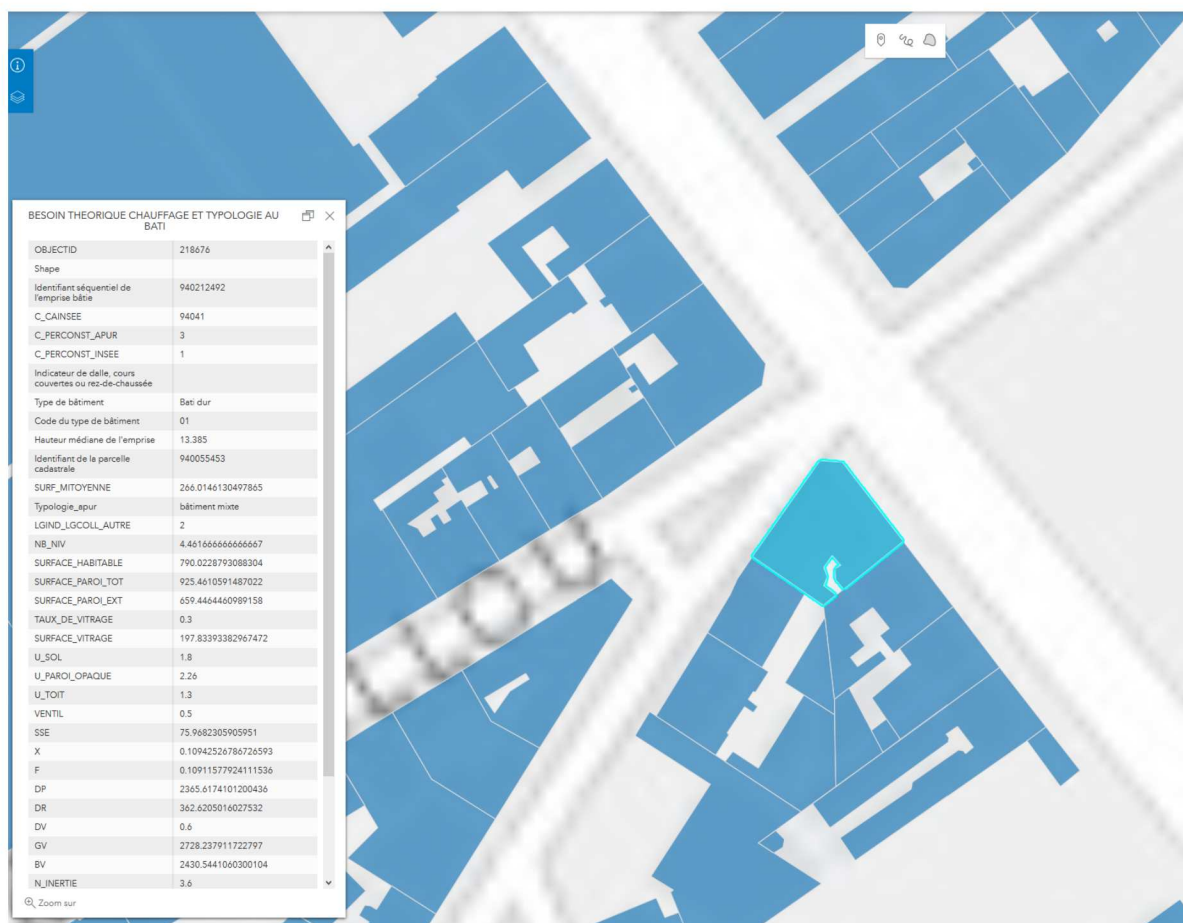


Fig. 1.1 : Prévisualisation du jeu de données de l'APUR

Au sein de l'APUR, le jeu de données intitulé « Besoins théoriques chauffage et typologie au bâti » associe une **cartographie composée de polygones géoréférencés** (représentant les bâtiments existants sur une zone donnée, ici le Grand Paris) à des **données alphanumériques sur chaque bâtiment**. Comme l'illustre l'extrait de la carte interactive mise à disposition sur le site de l'APUR, certains critères comme la **hauteur**, la **destination**, la **période de construction** ou encore le **taux de vitrage** représentent une **source d'informations précieuse pour le domaine du bâtiment**, en particulier lorsque l'on souhaite fonder son travail sur une masse d'informations.

Sur la plateforme d'IGN, le jeu de données nommé **BD TOPO** regroupe une multitude de données constituée de couches d'informations disponibles pour l'ensemble de pays, tel que le réseau routier, les fleuves, ou plus particulièrement une couche sur le **bâti**. A l'instar du jeu de données de l'APUR montré plus haut, il associe des polygones géoréférencés à leurs informations correspondantes. Dès lors, pour un même bâtiment, parmi quelques données déjà présentes sur le jeu de l'APUR figure la **caractérisation des matériaux de toiture et de façade** (indiquée selon un code).

7

Un enjeu majeur est alors celui de l'**agrégation**, de sorte à pouvoir compiler son propre jeu de données regroupant des informations provenant de jeux différents. Ce jeu pourrait ensuite servir de base d'entraînement pour créer des **algorithmes de prédiction** afin de **recréer des données manquantes sur de l'existant** par exemple.

Dès lors, ce type de données représente un **potentiel particulièrement intéressant pour l'architecte**, dont un des coeurs de métier est la capacité d'analyse de l'existant. Cependant, d'un point de vue théorique et pratique, ces derniers manquent en réalité d'outils spécifiques permettant de se saisir de ces ressources.



Fig. 1.2 : Prévisualisation du jeu de données d'IGN

1.2 ..face à des architectes manquant d'outils de traitement de données

Par leur expertise orientée sur la conception réclamant un esprit de synthèse affûté, les architectes ont tendance à préférer des résultats explicites d'analyse de données plutôt que les données en elles-mêmes.

De plus, tel que l'illustre un article de recherche intitulé «Classification and Use of Design Tools : The Roles of Tools in the Architectural Design Process» produit par l'Université d'Has-selt,⁵ les outils numériques sur lesquels les architectes pratiquent leur exercice relèvent très majoritairement des domaines du dessin ou de la modélisation. A l'inverse, le graphique ci-contre exprimant les domaines ayant le plus **besoin d'amélioration au niveau des outils** chez les agences interrogées illustre un manque certain dans ceux de **l'évaluation et de l'analyse**.

9

Cela se constate également au niveau des jeux de données identifiés précédemment. En effet, au delà des données vectorielles représentant l'emprise des bâtiments (qui, de manière isolées sont facilement manipulables au sein de logiciels de dessin comme AutoCAD), les données alphanumériques associées répondent à une **nomenclature précise** constituée de **variables** et/ou de valeurs codifiées afin de faciliter leur compilation (comme des abréviations par exemple). Cet ensemble de règles spécifique à chaque jeu est indiqué au sein d'un document appelé **métadonnées**, et dont le lecteur doit impérativement se saisir pour pouvoir extraire les données qu'il souhaite.

⁵WEYTJENS, Lieve, VERDONCK, Evelien et VERBEECK, Griet. *Classification and Use of Design Tools : The Roles of Tools in the Architectural Design Process*. Janvier 2009, p. 298-300.

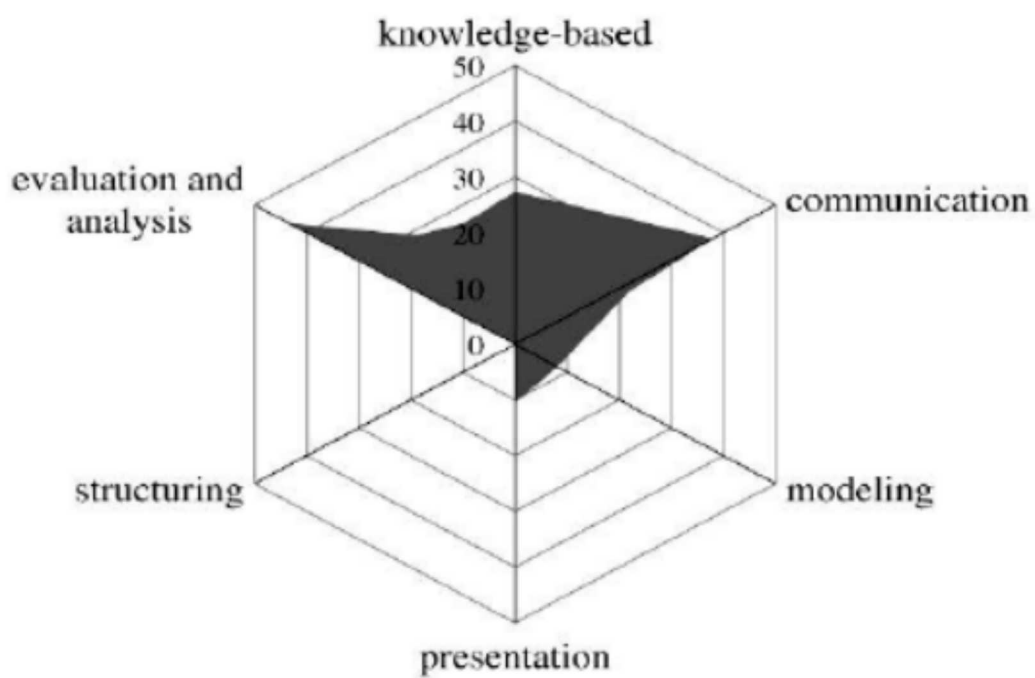


Fig. 1.3 : Secteurs d'amélioration en termes d'outils pour les agences

Par exemple, dans le cas du jeu de données issu de l'APUR, il est nécessaire de consulter ses métadonnées si l'on souhaite obtenir les dates de constructions de différents bâtiments, afin de connaître le **nom de la bonne variable** (*C_PERCONST_APUR*) ainsi que les **époques associées à chaque chiffre**.

A cette difficulté s'ajoute celle du **format de fichier**. En effet, dû au fait que ce type de jeu de données regroupe des types de données différents (alphanumériques et vectorielles), les formats tabulaires répandus comme le format Excel ne peuvent être utilisés. A l'inverse, ce sont des formats de données plus spécifiques comme le **shapefile** ou le **GeoJSON** qui sont ici employés pour échanger ces données. Ces formats privilégient par leur structure une lecture par la machine, plutôt que par l'humain. En conséquence, ils s'éloignent naturellement de la volonté de synthèse que réclame l'architecte.

Ainsi, il existe un véritable enjeu à disposer d'un outil à la fois **puissant** et **souple**, permettant de passer de **données brutes** encapsulées dans des formats orientés pour la machine, vers des formats **synthétiques** plus pertinents pour l'architecte.

NOM DES VARIABLES	DESCRIPTIF DES VARIABLES	VALEURS POSSIBLES
N_SQ_EB	Identifiant informatique séquentiel de l'emprise	
N_SQ_EB_ORI	Identifiant informatique séquentiel d'origine de l'emprise	
C_PERCONST_APUR	Période de construction (découpage APUR en 11 classes)	1 : Avant 1800 2 : 1801-1850 3 : 1851-1914 5 : 1915-1939 6 : 1940-1967 7 : 1968-1975 8 : 1976-1981 9 : 1982-1989 10 : 1990-1999 11 : 2000-2007 12 : Après 2008 99 : Non daté
C_PERCONST_INSEE	Période de construction (découpage INSEE en 7 classes)	1 : Avant 1949 2 : 1949-1974 3 : 1975-1981 4 : 1982-1989 5 : 1990-1998 6 : 1999-2006 7 : Après 2006 99 : Non daté
B_DALLE	Indicateur de dalle, cours couvertes ou rez-de-chaussée	O (Oui), N (Non)
DUR	Type de bâtiment	Bati dur Bati léger
DUR_CODE	Code du type de bâtiment	01 : Bati dur 02 : Bati léger
H_MEDIANE	Hauteur médiane de l'emprise bâtie (mètres)	
SURF_MITOYENNE	Surface de mur mitoyenne (m²)	
Typologie_apur	Nature du bâtiment / occupation	Logement individuel Logement collectif Bâtiment mixte Bâtiment tertiaire ou industriel Non déterminée
LGIND_LGCOLL_AUTRE	Nature du bâtiment simplifiée selon 3 classes	1 : logement individuel 2 : logement collectif / bâtiment mixte 3 : bâtiment tertiaire ou industriel
NB_NIV	Nombre de niveaux/ étages du bâtiment	
SURFACE_HABITABLE	Surface habitable estimée du bâtiment (m²)	
SURFACE_PAROI_TOT	Surface totale de murs (m²)	
SURFACE_PAROI_EXT	Surface de murs non mitoyens (m²)	
TAUX_DE_VITRAGE	Taux de vitrage du bâtiment	
SURFACE_VITRAGE	Surface de vitrage du bâtiment (m²)	
U_SOL	coefficient de transmission thermique du plancher (W/m²/K)	
U_PAROI_OPAQUE	coefficient de transmission thermique des murs (W/m²/K)	

Fig. 1.4 : Métadonnées du jeu de données de l'APUR

1.3 L'intérêt du langage de programmation Python

Lorsque l'on évoque la question de la manipulation de données brutes et massives, les **langages de programmation** apparaissent comme une solution pertinente.

En l'occurrence, le langage **Python** est en plein essor depuis les années 2010. La **simplicité de sa syntaxe** ainsi que sa **profusion de bibliothèques spécialisées** (comparables à des «plug-in») en ont fait un outil extrêmement **populaire** parmi les domaines où le traitement automatique de données est aujourd'hui crucial.

C'est le cas au sein des Systèmes d'Informations géographiques (SIG) tels que ArcGIS, où ses caractéristiques mentionnées ci-dessus permettent de manière **accessible** de mener des travaux complexes autour des données géographiques ouvertes, de l'**analyse** et la **visualisation** à l'entraînement de **modèles de prédiction**.⁶

13

Des bibliothèques comme **GeoPandas** permettent notamment de résoudre le problème très pragmatique des formats de fichiers de l'Open Data abordé précédemment. En effet, cette dernière permet de gérer la **majorité des formats complexes** évoqués, tout en **conservant une hiérarchisation sous forme de tableau**, afin de garantir une lisibilité certaine pour l'utilisateur, permettant d'opérer plus facilement les traitements souhaités.

⁶*Automate Road Surface Investigation Using Deep Learning | ArcGIS for Developers* [en ligne]. [s. d.]. [Consulté le 3 février 2021]. Disponible à l'adresse : <https://developers.arcgis.com/python/sample-notebooks/automate-road-surface-investigation-using-deep-learning/>.

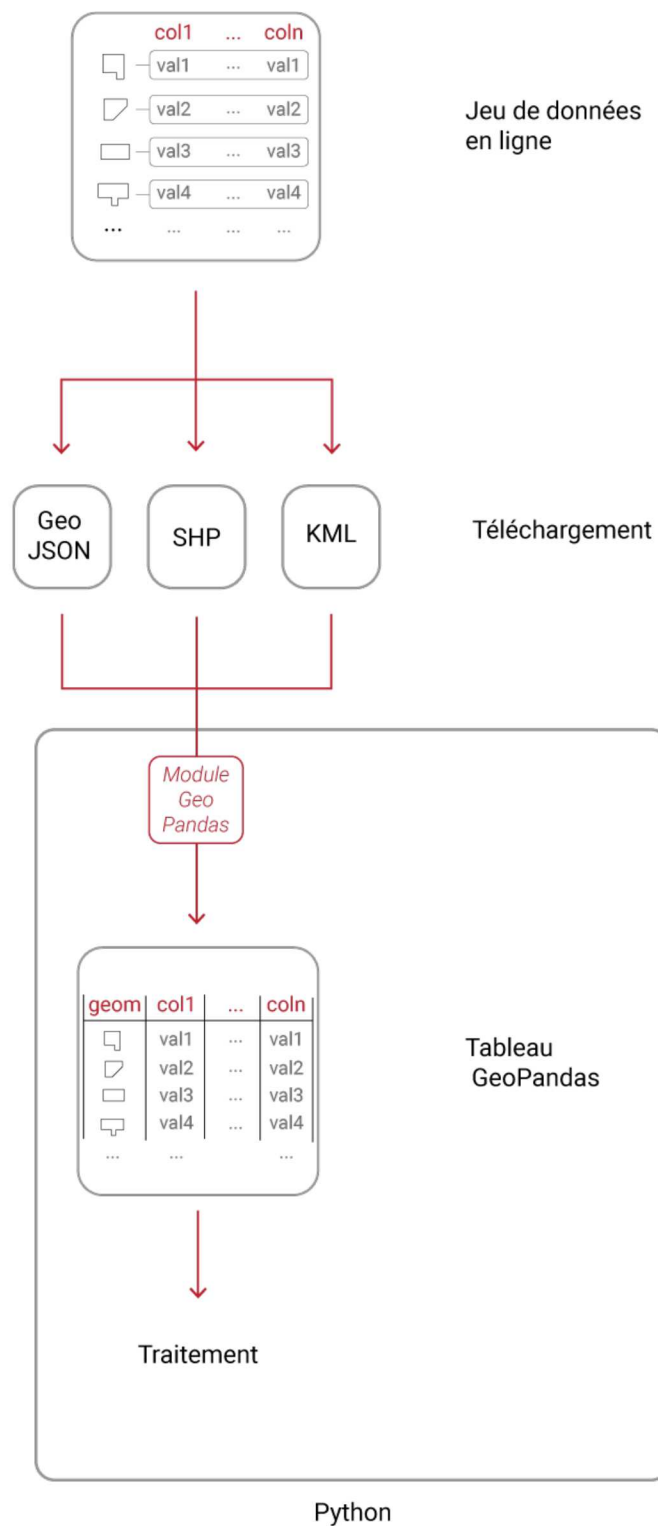


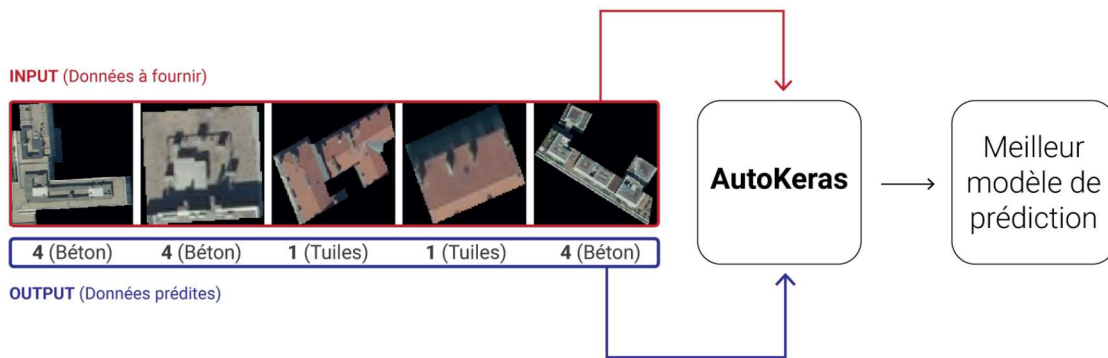
Fig. 1.5 : GéoPandas : module Python facilitant la manipulation des formats géographiques courants

Par conséquent, le Python possède également une très forte affinité avec les domaines de l'**IA** et du **Machine Learning**. De grandes infrastructures logicielles comme **Tensorflow**⁷ sont aujourd'hui majoritairement orientées vers ce langage, permettant ainsi facilement de passer du traitement des données à la création d'**algorithmes** de prédiction.

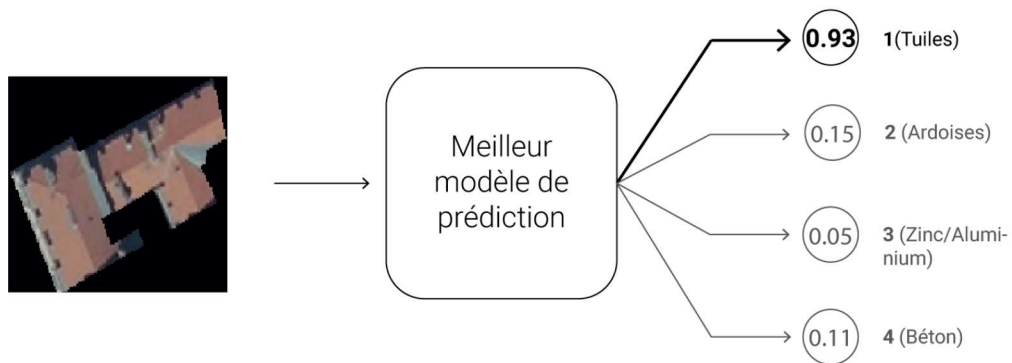
Le langage Python y conserve tout de même sa philosophie, prodiguant là aussi des outils comme **AutoKeras**, destiné à simplifier la création de **réseaux de neurones** en gérant lui-même les paramètres internes, permettant ainsi même aux plus novices de s'approprier ce domaine à travers une **approche pratique** (en se concentrant uniquement sur les données en entrée et sortie du modèle souhaité).

Dès lors, tous ces outils sont désormais à la portée des architectes, formant un atout à caractère **universel** permettant de les guider à travers la **manipulation des structures de données courantes en Open Data**, et ce jusqu'à l'**entraînement d'algorithmes personnalisés**.

⁷TensorFlow [en ligne]. [s. d.]. [Consulté le 20 janvier 2021]. Disponible à l'adresse : <https://www.tensorflow.org/>.



ENTRAINEMENT



UTILISATION

Fig. 1.6 : Autokeras : module rendant accessible la création de modèles de prédiction

1.4 Un algorithme pour exploiter le potentiel de l'existant

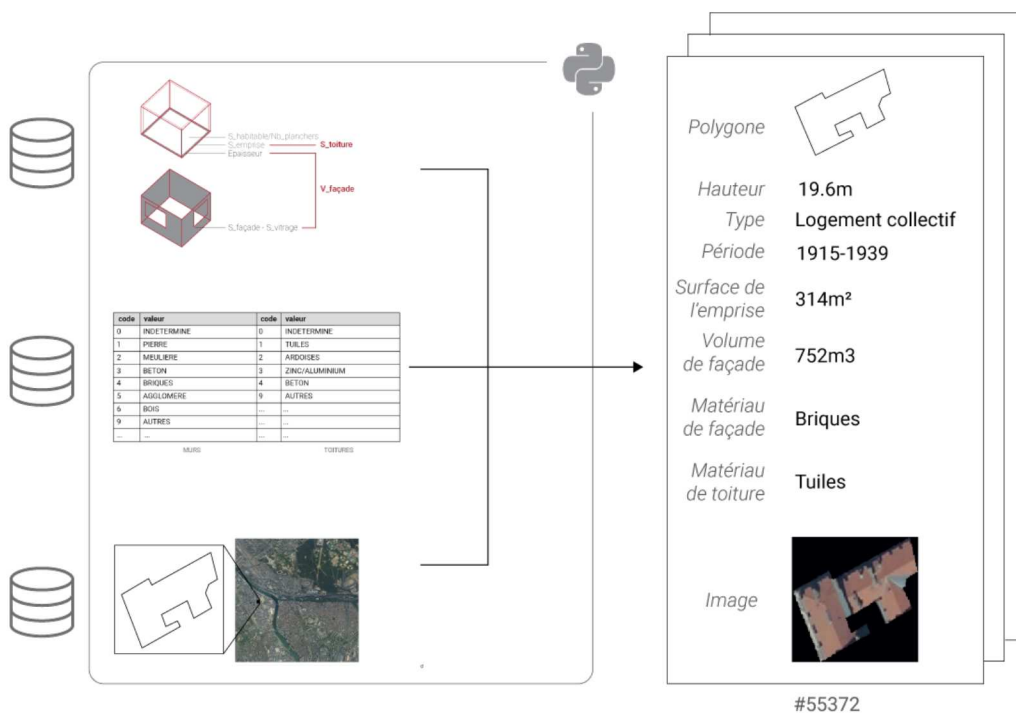
Grâce à la puissance des outils modernes présentés précédemment, les jeux de données disponibles représentent donc le **réactif limitant** de ce qu'il est possible de faire avec des algorithmes. La question principale est en effet de déterminer **à partir des données disponibles, quelles informations sont capables d'être prédites. Qu'en est-il alors de leur intérêt dans le cadre de la conception ?**

L'exemple montré précédemment, bien qu'illustrant parfaitement cette démarche (association relativement «évidente» de la photographie aérienne et du matériau de toiture sur la base de la corrélation forte entre la première et la seconde), ne présente que très peu d'intérêt dans le cadre de la conception architecturale.

Un atout qui se révélerait particulièrement utile serait de pouvoir également **quantifier** cette ressource matérielle, dans l'optique de déterminer le **gisement potentiel brut de matériau d'un ensemble de bâtiments** afin de le réemployer dans le projet par exemple. Disposant déjà de matériaux caractérisés dans le jeu de données issu d'IGN (cf. chapitre 1), il serait alors nécessaire de recouper cette donnée avec une mesure du **volume** de matière suffisamment précise.

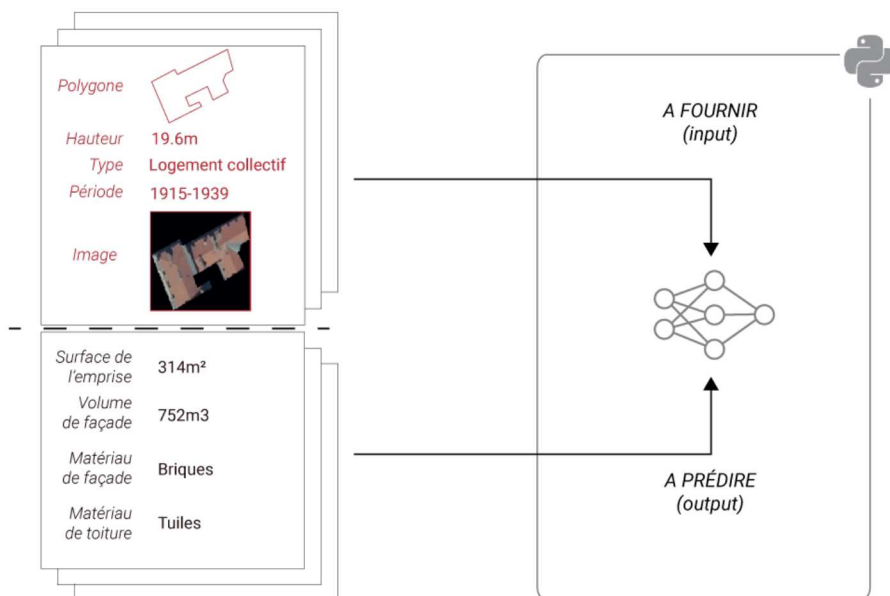
C'est ici que le principe de l'**agrégation** présenté lors du chapitre 1 est crucial. En effet, le jeu de données issu de l'APUR, dans le cadre de ses estimations thermiques, inclut justement un ensemble de mesures volumiques et surfaciques pour chaque bâtiment. Les deux jeux de données (IGN et APUR) doivent donc être agrégés grâce à un script Python (avec *Geopandas*), permettant d'**identifier le même bâtiment sur deux jeux différents pour en associer les paramètres**.

Cette **caractérisation plus complète** des matériaux de l'existant forme donc **la donnée à prédire** (ce que l'on souhaiterait obtenir pour un ensemble de bâtiments). La dernière étape est donc de déterminer **sur quelles autres données sera basée cette prédiction** (la principale contrainte étant d'être capable de le faire à partir d'**informations facilement récupérables** dans un contexte dépourvu de données).



AGRÉGATION

18



ENTRAINEMENT

FIG. 1.7 : l'agrégation : principe fondamental pour exploiter l'Open Data

Or, ce n'est pas aussi trivial que dans l'exemple des matériaux de toiture. En effet, une photographie aérienne ne permettrait ici que de renseigner sur des associations typiques de matériaux de toiture et de façade. Se limiter à cette donnée risque donc de n'aboutir qu'à une prédiction qualitative limitée, ainsi qu'à une prédiction quantitative encore plus approximative.

Il est donc nécessaire d'y associer **suffisamment de facteurs** pour obtenir une prédiction satisfaisante, sans toutefois compromettre l'approche *minimale* des données servant à effectuer cette prédiction. Ainsi, pour chaque bâtiment, les informations suivantes ont été retenues (cf. illustration précédente) :

- Photographie aérienne
- Emprise (polygonale)
- Hauteur
- Période de construction
- Destination


19 Ces données ont pu alors être fournies à un **second script Python** (grâce à *AutoKeras*) dans le but d'obtenir le meilleur algorithme de prédiction possible.

Enfin, il est plus que nécessaire de le **mettre à l'épreuve face à un cas d'utilisation type**. L'image ci-contre montre le résultat d'un tel essai, sur un ensemble urbain donné, où l'on constate la **précision satisfaisante** de l'algorithme, le rendant pleinement exploitable. Ainsi, bien que les données disponibles actuellement restent limitées dans leur précision, cet algorithme confirme bel et bien un **intérêt certain dans le cadre de la conception architecturale**.

Épaulé par Python, l'architecte possède donc la capacité de se saisir des données de la ville connectée, et de les mettre au service de son activité de conception, afin de révéler des ressources précieuses jusqu'alors inexploitable. À l'avenir, plus les données sur l'existant deviendront précises et diversifiées, plus les algorithmes entraînés à partir de ces dernières le seront également.

INPUTS :

Usage :
Logement collectif
Hauteur max :
6.9m
Période de construction :
1968-1975

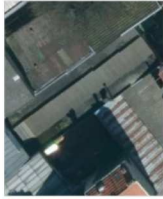


PREDICTIONS :

Matériau de façade :
Béton
Matériau de toiture :
Tuiles
Volume de façade :
50.46 m³
Surface de toiture :
190.44 m²

INPUTS :

Usage :
Tertiaire/industriel
Hauteur max :
4.3m
Période de construction :
Après 2008




PREDICTIONS :

Matériau de façade :
Bloc d'aggloméré
Matériau de toiture :
Tuiles (bitume)
Volume de façade :
26.8 m³
Surface de toiture :
130.46 m²

INPUTS :

Usage :
Tertiaire/industriel
Hauteur max :
7.8m
Période de construction :
1940-1967




PREDICTIONS :

Matériau de façade :
Briques
Matériau de toiture :
Tuiles
Volume de façade :
58 m³
Surface de toiture :
48.3 m²


INPUTS :

Usage :
Tertiaire/industriel
Hauteur max :
7.7m
Période de construction :
1940-1967



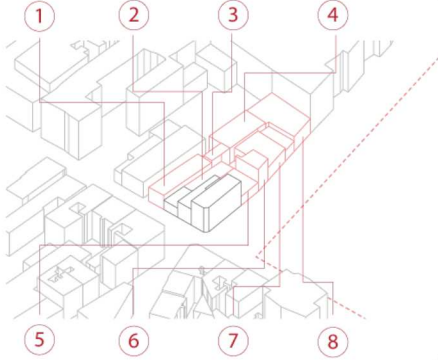
PREDICTIONS :

Matériau de façade :
Briques
Matériau de toiture :
Tuiles
Volume de façade :
170 m³
Surface de toiture :
182 m²




PREDICTIONS:

- Valide
- Erronée



INPUTS :

Usage :
Tertiaire/industriel
Hauteur max :
4.4m
Période de construction :
1915-1939




PREDICTIONS :

Matériau de façade :
Briques
Matériau de toiture :
Tuiles (tôle)
Volume de façade :
14.1 m³
Surface de toiture :
106.6 m²

INPUTS :

Usage :
Tertiaire/industriel
Hauteur max :
7.16m
Période de construction :
1851-1914




PREDICTIONS :

Matériau de façade :
Briques
Matériau de toiture :
Tuiles
Volume de façade :
170.7 m³
Surface de toiture :
256.8 m²

INPUTS :

Usage :
Tertiaire/industriel
Hauteur max :
6.3m
Période de construction :
1982-1989




PREDICTIONS :

Matériau de façade :
Briques (Béton)
Matériau de toiture :
Tuiles
Volume de façade :
162.8 m³
Surface de toiture :
171.3 m²

INPUTS :

Usage :
Logement collectif
Hauteur max :
10.8m
Période de construction :
1851-1914



PREDICTIONS :

Matériau de façade :
Meulières (Autre pierre)
Matériau de toiture :
Tuiles
Volume de façade :
34.6 m³
Surface de toiture :
61.9 m²

FIG. 1.8 : Mise à l'épreuve de l'algorithme