



Données Météo ouvertes

L'Avis n° 12 du 5 juin 2012 du Conseil national du numérique relatif à l'ouverture des données publiques (http://www.cnnumerique.fr/avis12/#_ftn27), la proposition n°5, dit ceci :

« D'autres établissements producteurs de données peuvent être soumis à une contrainte d'autofinancement, avoir consenti à des investissements exceptionnels ou être en situation de concurrence. L'existence même des données et leur mise à disposition dans des conditions favorables à leur réutilisation doivent alors primer. Dans tous les cas, le calcul des redevances doit être justifié et viser une réutilisation maximale. Pour les données enrichies, des systèmes non discriminants de partage de la valeur créée peuvent éventuellement être imaginés, en particulier pour les EPIC s'ils venaient à être intégrés dans le régime de réutilisation (cf. proposition n°2).»

Trois paragraphes plus haut, on peut lire :

«Le droit de réutilisation de certaines données publiques donne aujourd'hui lieu au paiement d'une redevance. Ces redevances mettent en évidence une volonté de valorisation des informations publiques, mais, si elles sont trop élevées, constituent des barrières d'entrée qui sont à l'avantage des gros acteurs privés et favorisent la formation d'oligopoles, au détriment de la diffusion des données et de l'innovation. Elles freinent en particulier l'expérimentation par des startups. Les expériences passées ont montré que la réduction des redevances à des niveaux faibles, voire nuls, pouvaient augmenter le nombre de réutilisateurs de 1000% à 10000% [24]. Les redevances brident dans ce cas le potentiel économique (et indirectement les recettes fiscales qui en résulteraient) et social de la réutilisation des données publiques, pour des revenus souvent marginaux en regard du budget des organismes producteurs.»

La proposition n°2 a quant à elle pour titre :

«Ne prévoir à la définition du périmètre des informations publiques d'autres exceptions que celles découlant de l'application d'autres lois (données à caractère personnel, secret statistique, secret défense etc.). Etendre le droit de la réutilisation aux données brutes des SPIC.

Directives 2013

<http://www.etalab.gouv.fr/article-la-feuille-de-route-du-gouvernement-en-matiere-d-ouverture-et-de-partage-des-donnees-publiques-115767801.html>

Dans la feuille de route Etalab, le nouveau gouvernement souhaite favoriser l'ouvertures de données stratégiques et de qualité qui permettent de créer des service utiles au quotidien (comme pour le transport, l'environnement,...).

Cette feuille de route dit aussi :

À l'occasion du CIMAP du 18 décembre 2012, le Premier ministre a confié au SGMAP, en lien avec le ministère du Budget, une mission d'évaluation des modèles économiques des redevances existantes, notamment en auditant les coûts et les recettes associés. Ces recommandations seront

remises au premier semestre 2013. Etalab et le SGMAP accompagneront les administrations qui commercialisent de l'information publique dans l'évolution de leurs modèles économiques. En effet, le respect du principe d'accès libre et gratuit du citoyen à des données publiques réutilisables n'empêche pas de concevoir des stratégies de monétisation de services à haute valeur ajoutée ou de création de plateformes.

Application à Météo-France

Météo-France est un Service Public à Caractère Industriel et Commercial (SPIC). Son budget de fonctionnement est lié à la commercialisation des données. Si l'on en croit le rapport de la Cour des Comptes de 2003 (<http://www.senat.fr/rap/r04-264/r04-26413.html>), cette part de revenus (en croissance d'environ 3% par an), serait de 45M€ en 2006 pour 291M€ de produit total, soit 15%. Le rapport Trojette de juillet 2013 sur l'ouverture des données publiques donne (page 50) le chiffre de 1,585 millions d'euros comme recettes des redevances de Météo-France (hors recettes refacturées à sa branche commerciale) dont le budget annuel est de 370 millions d'euros. On peut penser qu'une gratuité des données n'aurait qu'un faible effet sur Météo-France, avec, en revanche, un développement économique national beaucoup plus valorisant, sans compter les économies que Météo-France du fait de la suppression de la structure de vente.

Il serait intéressant de connaître la part des revenus liés à la prévision météo et celle liée aux données historiques, d'une part agrégées et d'autres part en données brutes. Les données historiques brutes sont probablement faiblement diffusées, car payantes. Le coût des données dissuade demande et la faible demande leur fait perdre Météo-France sur l'attente réelle des usagers. Météo-France ne voit donc pas de nécessité à produire gratuitement les données historiques brutes : le cercle est vicieux.

On conçoit que les données à haute valeur ajoutée produites par Météo-France fassent l'objet d'un partenariat avec leurs utilisateurs. Par contre, dans la dynamique de l'ouverture des données publiques, il serait très dommage que Météo-France ne mette pas en ligne les données brutes historiques de toutes ses stations. Ces données brutes existent et dès lors que les processus d'acquisition sont automatisés, on ne voit pas pourquoi elle ne seraient en ligne que pour les besoins internes de Météo-France.

A noter que les données météo sont publiques aux Etats Unis depuis 100 ans !

CIMAP du 18 décembre 2013 : on botte en touche

Malheureusement, dans la décision n°26 du Comité interministériel pour la modernisation de l'action publique du 18 décembre 2013, le Gouvernement précise sa doctrine en matière d'exceptions au principe de gratuité :

– aucune redevance ne saurait être exigée sur des données résultant des missions de service public des administrations générales. Il est demandé aux administrations d'accroître leurs efforts en matière de publication de données leur permettant de mener à bien leurs missions, de motiver leurs décisions, de mesurer leurs activités ou leur bilan.

Le site data.gouv.fr, est en mesure d'accueillir ces données ;

– le respect du secret statistique et donc l’anonymisation des informations pouvant éventuellement permettre d’identifier des personnes est un principe central du droit français. Certaines redevances ont été instaurées pour financer cette tâche, qui est indispensable avant ouverture de la donnée. Le Gouvernement demande au SGMAP d’apporter tout le soutien possible aux services publics concernés, afin d’identifier les moyens de réduire ces coûts sans dégrader la qualité de cette anonymisation ;

– enfin, le Gouvernement réaffirme que les opérateurs dont la mission même est de produire des données doivent rechercher des modèles économiques leur permettant de faire face à un paysage économique en profonde reconstitution. Conformément aux conclusions du rapport Trojette, il leur demande d’engager, dans les meilleurs délais, avec l’appui du SGMAP et du ministère du Budget, une réflexion sur les évolutions de leurs modèles économiques. Il leur demande de rechercher des modèles stimulant l’innovation autour de leurs données, favorables aux entrepreneurs innovants, et soutenables à l’heure de l’économie numérique, de la production de nombreuses données par les citoyens eux-mêmes, et des stratégies de plateformes. Afin de mener cette transition - qui vise à, maintenir et améliorer la qualité des informations publiques, à garantir la pérennité du service public et à développer des stratégies numériques pérennes, efficaces et innovantes,

le Gouvernement demande au SGMAP d’accompagner au mieux ces opérateurs, tel l’IGN (Institut géographique national), le SHOM (Service hydrographique et océanographique de la marine) ou Météo France, en lien avec les ministères de tutelle pour définir avec eux des trajectoires d’investissement dans l’innovation compatibles avec le redressement des finances publiques.

Pour l’instant, Météo France met en ligne sans redevance un certain nombre d’observations et d’éléments climatologique

<https://donneespubliques.meteofrance.fr/?fond=caracteristique&caracteristique=11&caracdisp=28>

Par exemple, on peut y trouver des données essentielles pour une cinquantaine de stations pour une date-heure à choisir. Il y a de quoi faire pour qui veut suivre l’évolution du temps et tous ses aspects localisés, à condition de s’y investir lourdement.

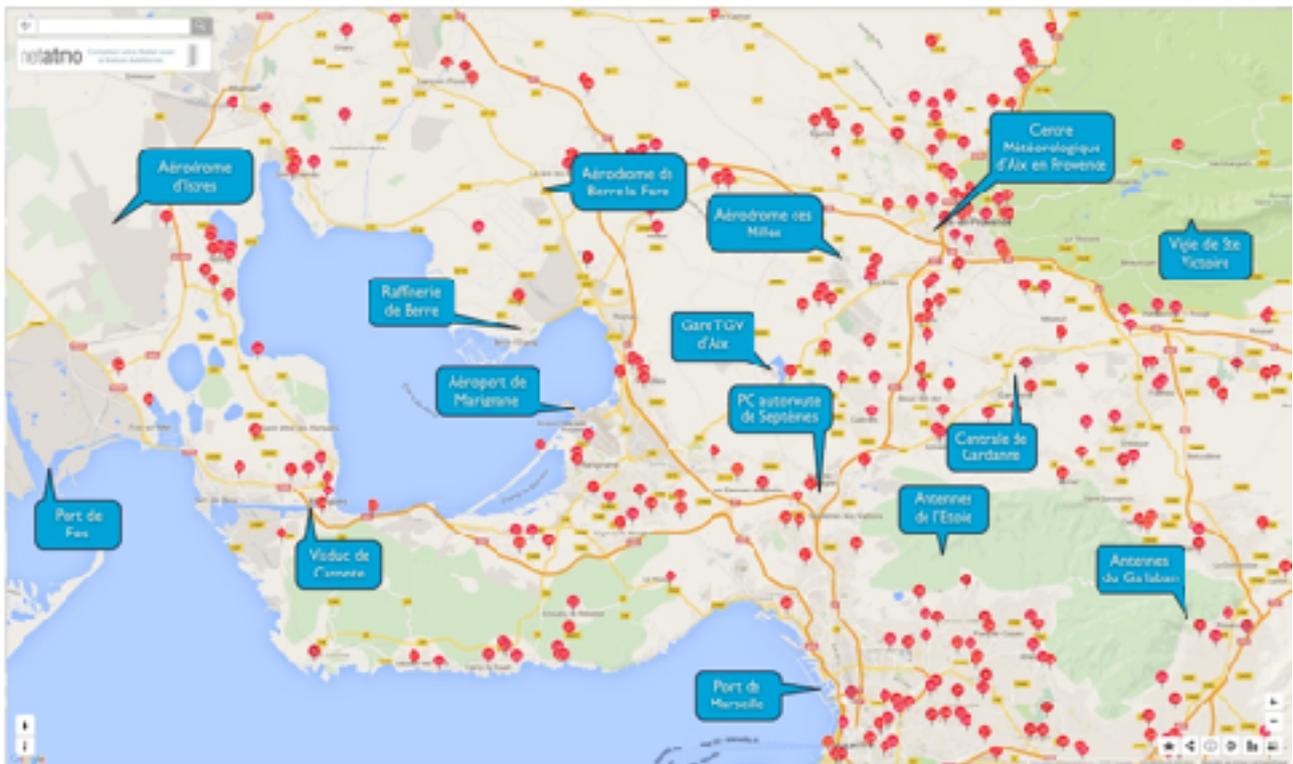
Pour qui s’intéresse à l’historique des températures et autres données météo de base, il ne semble pas qu’il y ait des données facilement accessibles de la part des institutions officielles.

Paradoxalement, il existe au moins deux cartographies participatives où l’on peut trouver des données issues de stations météo privées :

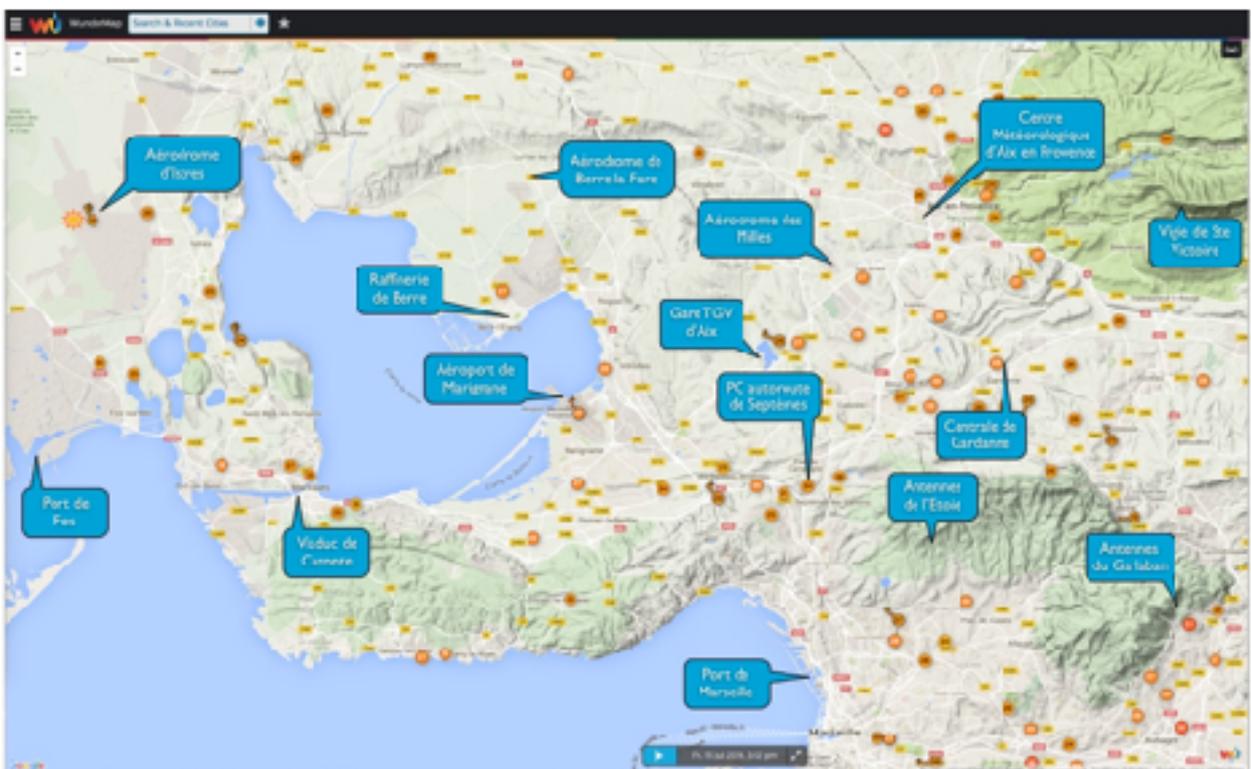
wunderground.com/wundermap, qui propose des icônes affichant directement la température, la force et la direction du vent.

Weathermap, qui propose des icônes affichant la température des stations Netatmo (exclusivement ?)

Les images suivantes montrent que les institutions pour qui la météo locale est un élément important (viaducs, aérodromes, raffineries, services de Météo-France (!), vigies, antennes, ...) ne partagent pas leur données. Les seules données fiables proposées sont celles de Marnay-la-Rochelle et de Istres.



15/7/2016 - Localisations d'institutions officielles, a priori équipées de stations météo, non représentées sur la carte Weathermap. Les icônes roses sont les «données ouvertes» individuelles.



15/7/2016 - Localisations d'institutions officielles, a priori équipées de stations météo, non représentées sur la carte Wundermap. Les icônes orange sont les «données ouvertes» individuelles. Seuls les aéroports de Marignane et de Istres sont partagés.

Ergonomie des données

Au cas heureux où Météo France s'intéresserait à un service de vulgarisation du temps passé, voici un souhait de simple citoyen, qui voudrait par exemple voir l'évolution des pluies ou de la vitesse des vents sur les n dernières années (si Météo France dispose des mesures d'il y a 70 ans comme à Dumont d'Urville depuis 1951... ou du Parc Montsouris depuis 1900...)

L'ergonomie des données est essentielle :

- simplicité d'accès : stations localisées sur une carte, clic pour accéder aux pages web de la station (chaque station devrait depuis longtemps être autonome avec alimentation solaire et réseau sans fil, et disposer de son propre serveur, à l'instar des stations privées accessibles sur Internet).
- simplicité des formats : pas d'outil logiciel spécifique
- données explicites : voir immédiatement les données et leur signification

Faute de ces exigences, le service n'atteindra que ceux qui manipulent les données mais ne les utilisent pas. Matlab vs Excel : laissons Matlab aux scientifiques et Excel à tous, avec une taille de fichier raisonnable.

On peut espérer par exemple que via un moteur de requête simple (compatible avec tous les navigateurs Internet) chacun puisse télécharger un tableau de données 6h, horaires ou 6 mn selon disponibilité pour une station donnée et pour une période choisie.

Chaque tableau de données doit être «auto-suffisant» : l'utilisateur ne doit pas avoir besoin d'autres outils qu'un tableur (.csv ou .xls) et un navigateur géographique pour identifier chaque mesure. Pour simplifier la vie des utilisateurs, chaque tableau téléchargé comporte une ligne d'en-têtes avec le nom des mesures ou des paramètres de la colonne.

Par exemple :

Données multiples brutes

Sur une sélection simple dans un calendrier, la station propose, pour la période sélectionnée, l'ensemble de ses données disponibles à raison d'une ligne par tranche de temps, sur une période choisie, exportable en .csv (;)

Les localisants / horodates sont dans les premières colonnes.

La première ligne définit les données fournies (selon le choix de l'utilisateur) :

Dépt; Commune; kmz; Horodate; Température (°C); Pression; Humidité (%); Vitesse (km/h); Rafale (km/hmax); Direction (°); Pluie (mm/tranche);

Selon les possibilités de la station, on peut y ajouter la Visibilité (m), l'ensoleillement, la température du sol, l'indice ATMO, le débit de la rivière (m3/sec)

Le localisant doit être lisible sur les cartes IGN au 1/25000 (département, axe routier, lieu-dit ou adresse urbaine)

GPS : coordonnées sous forme d'un lien .kmz (Latitude, Longitude, Altitude - tels que disponibles dans Google Earth). Un simple clic sur le localisant GPS suffit à ouvrir Géoportail (ou GE ou GM, selon les préférences de l'utilisateur), centré sur la station, accompagné de méta-données utiles (altitude, contact courriel, périodes historiques de disponibilité des mesures,... modèle et précision des capteurs,... descriptif des mesures...)

jj/mm/aa hh:mm:ss : horodate de la mesure si la mesure est ponctuelle (ou du début de la séquence de mesure lorsque la mesure est une moyenne).

La page Web de la station comporte une mini-carte de situation. Cette carte permet aussi d'accéder aux autres stations du réseau.

Données uniques brutes

Considérant que les données météo obéissent à un double cycle, diurne et saisonnier, il est intéressant que la station propose pour chaque type de données sur un an par tranche de temps de 10mn, soit un tableau de 144 lignes par 366.

Un simple sur le type de données choisi et sur l'année choisie déclenche l'affichage et la possibilité de téléchargement (ou de copie/coller)

Il est toujours intéressant d'observer les corrélations des différents types de données météo, au moyen d'une visualisation graphique à n dimensions.

Par exemple : http://ertia2.free.fr/Niveau2/Projets/Graphe_n_dimensions.pdf

Données agrégées

La station engrange avec le temps des tableaux prêts à l'affichage (même format que pour les données multiples brutes) tels que les moyennes, les maxima/minima, l'écart type, quotidien sur les 30 derniers mois ou mensuels sur les années renseignées,