

Données publiques à Météo-France

Frédéric Ameys (Association Infoclimat) - frederic.ameys@infoclimat.fr

Nicolas Baldeck (Météo-Parapente) - n@baldeck.net

Christian Quest (OpenStreetMap France, EtaLab) - cquest@cquest.org

À l'heure de changements sans précédent dans notre climat et de conséquences déjà visibles et coûteuses, la situation des données publiques à Météo-France s'affiche comme une **exception européenne et même mondiale dans la diffusion des données météorologiques** et climatiques, ainsi que la dernière exception au principe général de gratuité d'accès et de réutilisation des informations publiques en France. L'ouverture en Open-Data de ces données a pourtant un impact considérable sur de nombreux secteurs d'activité, en ayant des **bénéfices socio-économiques de l'ordre de 4 à 8 fois l'argent investi** dans Météo-France.

Pourtant, la mise en accès ouvert des données de Météo-France peut se faire à coûts réduits, par des **choix technologiques adaptés**, un accompagnement par la DINUM et/ou les organismes européens tels Copernicus, et la réorientation des emplois actuellement affectés à l'extraction manuelle de données. Au global, il est probable que la mise à disposition des données en open-data soit à terme **génératrice d'économies** pour Météo-France, les redevances actuelles de mise à disposition (1-2M€) ne compensant qu'à peine les coûts pour Météo-France de la mise à disposition payante. Le retour d'expériences de l'Allemagne et du Danemark prouve qu'**un équilibre est possible pour Météo-France**, pour des bénéfices macro-économiques de plusieurs ordres de grandeur supérieurs aux coûts.

Les **bénéfices pour les services publics**, les domaines assurantiels, la protection des personnes et des biens dans les territoires, l'efficacité de l'agriculture, de l'énergie ou de la gestion de l'eau, sont considérables. Le rapport Deloitte les estime atteindre entre **1 à 2 Mds d'euros par an** d'ici à 2028 sur l'Union Européenne, les impacts indirects atteignant 4 à 8 Mds d'euros par an, avec un **accroissement du nombre d'emplois** estimé entre 40.000 et 80.000 ETP sur l'UE.

En outre, positionner Météo-France comme **acteur de référence des données météorologiques et climatiques** placerait le service dans une position centrale, limitant le **risque grandissant de se voir marginalisé** par d'autres acteurs produisant des données de plus faible valeur. Dès lors, **l'excellence scientifique des produits** de Météo-France serait **mieux valorisée**, l'écosystème pourrait difficilement s'en passer, et Météo-France pourrait continuer de fournir des prestations d'expertise et développer significativement celles d'accompagnement dans l'utilisation de ces données.

Néanmoins, une **phase de transition et d'accompagnement** de Météo-France doit être mise en œuvre, afin que les freins psychologiques (perte de souveraineté, risque de "pillage par le privé") techniques (mise en place d'une infrastructure adaptée), et financiers (renouveau du modèle économique) puissent être levés, comme cela a pu être le cas à l'IGN ou l'INSEE.

1. ENVIRONNEMENT JURIDIQUE

Loi Valter (2015) :

- Exception à la gratuité de réutilisation des informations publiques
- Transcription (tardive) de la directive PSI (Public Sector Information)
- La règle: la gratuité (pas de redevance de réutilisation) → [L324-1 du CRPA](#)
- L'exception est justifiée si moins de 75% de l'activité est couverte par des recettes fiscales, des dotations ou subventions (R324-4-1) : en toute logique ce n'est pas le cas de Météo-France
- Les redevances doivent être calculées de façon transparente (L324-3) : absence de transparence à ce niveau
- Les données MF pouvant être soumises à redevance sont listées dans le [tableau du D324-5-1](#) toute autre donnée doit être gratuite (sans redevance de réutilisation)

Loi Lemaire “pour une République numérique” (2016) :

- opendata par défaut : [L321-1-1](#)

Règlement PSI2 “High value datasets” :

- Règlement européen [listant des jeux de données concernés publié en janvier 2023](#) intégrant données climatologiques et les observations météo (entre autres)... archives numériques comprises

Répertoire d'Information Publiques :

- Répertoires des principaux documents administratifs (cela concerne aussi les bases de données) [L322-6](#)
- instaurés de longue date (2005), mais encore très peu respecté, ils n'existent pas ou sont très incomplets

2. SITUATION ACTUELLE DES DONNÉES

Météo-France dispose d'un [catalogue de “données publiques”](#), qui est incomplet, de nombreux produits n'apparaissant pas et ne sont donc pas portés à la connaissance des utilisateurs (voir paragraphe “Quelles données ?”).

Les autres produits sont répartis en trois catégories :

- sous Licence Ouverte Etalab, sans frais de mise à disposition (15% des produits)
- sous Licence Ouverte Etalab, avec frais de mise à disposition (25% des produits)
- sous licences fermées avec redevance (60% des produits)

2.1. Licence Ouverte sans frais de mise à disposition

Il s'agit des produits de modélisation (AROME/ARPEGE), d'observation au sol (stations météo du "réseau principal"), ou d'observation radar. Cependant, il existe une composante commune à tous ces produits :

- ils sont toujours **incomplets** ou n'offrent pas une **donnée suffisamment brute** pour être exploitables informatiquement ou en data-science.
 - les produits d'observation au sol sont fournis toutes les 3 heures pour une soixantaine de stations météorologiques (voir Fig.1, points verts), quand Météo-France dispose des données toutes les 6 minutes pour 1500 stations environ (voir Fig.1, points noirs et rouges). Aucune archive n'est disponible avant 1996 quand Météo-France en dispose pourtant depuis la fin du 19ème siècle.
 - les produits radar sont des images colorisées non-géoréférencées, ne disposant d'aucune échelle de couleurs, ni de données brutes permettant de connaître les valeurs exactes. Elles sont fournies avec 1h30 de retard en résolution très dégradée, et aucune archive n'est mise à disposition.
 - les produits de prévisions (AROME/ARPEGE) sont incomplets, Météo-France ne diffusant qu'un certain nombre de paramètres météorologiques, ce qui est insuffisant pour certaines applications scientifiques avancées. Les données sont également dégradées dans leur résolution spatiale et temporelle. Notamment, les données des modèles ensemblistes ne sont fournies que sous forme de statistiques pré-calculées, et non la mise à disposition de chacun des membres de l'ensemble.
- les **conditions de téléchargement ne sont pas adaptées à une diffusion massive**, car la plateforme est instable techniquement : limitation du débit par adresse IP, **fichiers corrompus** nécessitant de re-télécharger fréquemment et intégralement, inadéquation du modèle de distribution à des données de ce type (voir Annexe A). Un effort considérable d'ingénierie informatique doit être mis en œuvre par les ré-utilisateurs potentiels pour réussir à disposer des fichiers rapidement et de manière intègre, ce qui **limite considérablement leur utilisation pour des ré-utilisateurs ayant peu de moyens**.

Pour pallier ce dernier défaut, Météo-France a mis en place un **système "d'accès Premium"**, offrant un canal dédié entre les serveurs de Météo-France et l'utilisateur des données. Ce service est facturé 5040€ pour chaque tranche de 10Go par jour. Cela signifie que pour disposer avec une bonne **fiabilité des données complètes** des modèles AROME et ARPEGE (hors prévisions d'ensemble et immédiate), il faut compter environ **30.000€ par an**.

D'après une présentation au Conseil Supérieur de la Météorologie le 24 mars 2023, Météo-France compte pérenniser ce mécanisme : **un système gratuit, et un abonnement premium pour contourner l'instabilité du système gratuit**¹.

¹ <https://publicwiki.meteo.fr/download/attachments/23267371/Pr%C3%A9sentation%20Donn%C3%A9es%20Publiques%20au%20CMS%20OBS%2020230324.odp>

Il n'y a cependant que peu d'obstacles techniques ou économiques à la mise en place d'un service fonctionnel. À titre d'exemple, pour pallier aux défaillances des serveurs Météo-France, un particulier — seul — a mis en place un service alternatif de téléchargement de ces fichiers². Ce service distribue aujourd'hui le même volume journalier de données que le service de Météo-France, à des clients tels que EDF, Engie, Allianz, Infoclimat, Windy.com...

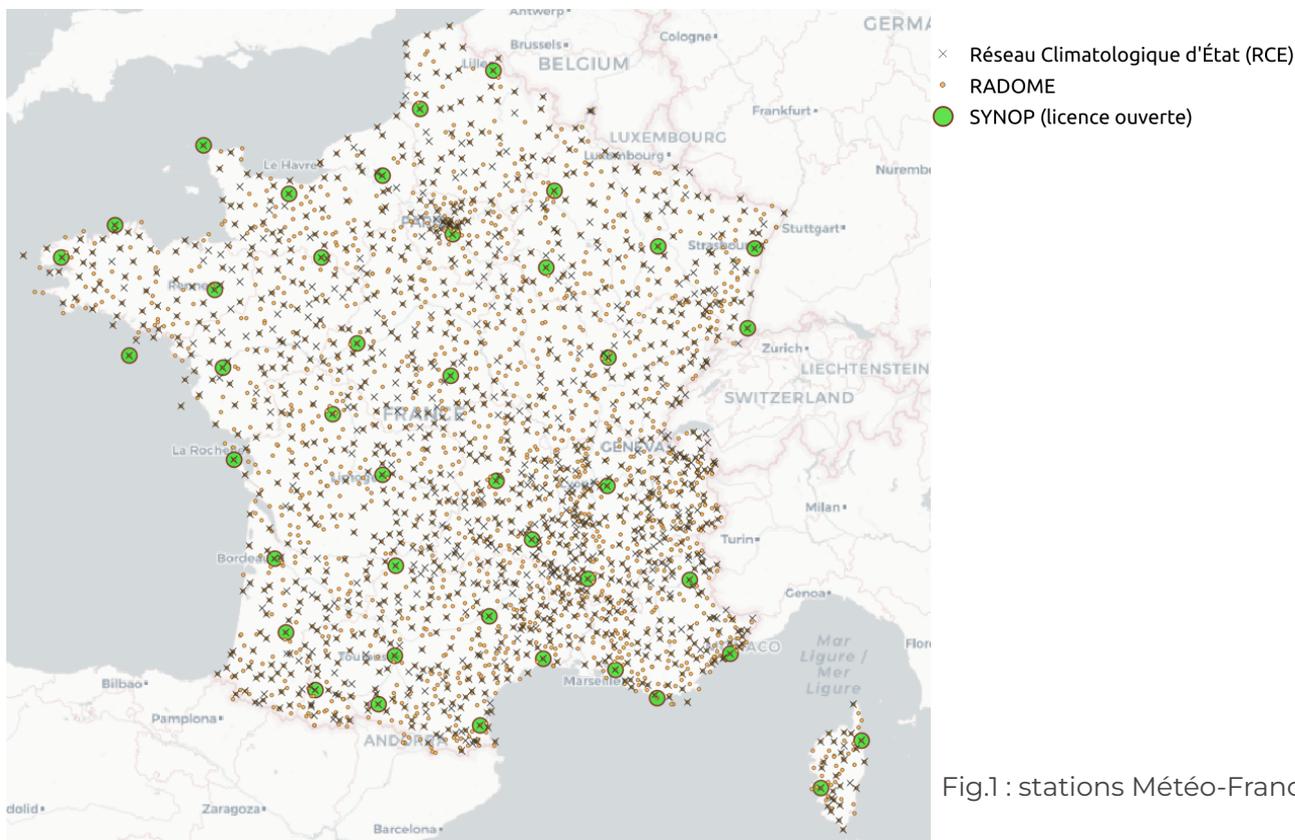


Fig.1 : stations Météo-France

2.2. Licence Ouverte avec frais de mise à disposition

De nombreuses données sont également disponibles sous des licences ouvertes, mais sont soumises à des **frais de mise à disposition**, censés refléter des coûts d'extraction des données depuis le système d'informations de Météo-France, et/ou la bande passante correspondante.

Les frais de mise à disposition sont détaillés en section 4.3 de la grille tarifaire de Météo-France³, et sont très complexes à comprendre, des "frais de transformation" pouvant être facturés en sus. On prend ci-dessous l'exemple des modèles de prévision immédiate d'AROME (prévision à haute fréquence, nécessaires pour la prévision des épisodes de pluies intenses), mais une tarification similaire s'applique à de nombreux autres produits..

Les frais s'estiment à environ 7000€ par an, en plus des frais initiaux de mise en place du service, ce qui limite considérablement l'accès à ces données sous Licence Ouverte.

4.3.7. Modèles de prévision						
Fourniture en temps réel	Frais de mise en place du service	Frais de calculateur	Frais annuels de suivi selon le rythme d'exécution			
			2 à 8 réseaux quotidiens	1 réseau quotidien	Hebdo.	Mens.
	65 €/h avec un minimum de 260 €	0,0278 € par échéance	780 €	390 €	195 €	195 €
Désarchivage	Frais d'extraction	Frais de calculateur	Frais à chaque envoi			
	65 €/h avec un minimum de 260 €	0,0278 € par échéance	6,20 € pour un envoi électronique ; dans les autres cas : coûts de support (cédérom, par exemple) et d'expédition (affranchissement, etc.).			

² <https://mf-models-on-aws.org/>

³ https://donneespubliques.meteofrance.fr/?fond=contenu&id_co...

En considérant le coût pour 100Go de bande passante par jour, cela correspond à **des frais 20 fois plus important que chez les hébergeurs cloud français**⁴, on peut donc questionner la réalité du calcul des frais de mise à disposition, sur lesquels la méthodologie de calcul devrait être clarifiée par Météo-France.

2.3. Données sous redevance

Météo-France dispose également d'un **jeu considérable de données soumises à redevance**, sans que la raison n'en soit précisée. On y trouve en particulier des **archives climatologiques** d'historique des conditions météorologiques du 19^{ème} siècle à nos jours, pour l'ensemble des stations météorologiques françaises, ou encore des données des **conditions de sécheresse des sols**. Il est très surprenant que ces jeux de données, de volumes faibles et de mise à jour peu fréquente par rapport aux données de modèles de prévision opérationnels, ne soient pas en libre accès, d'autant que **des plateformes européennes proposent déjà de les héberger**⁵ et de les remettre à disposition dans des APIs, et que cette proposition a également été faite par d'autres organismes (KNMI⁶, Infoclimat⁷), sans que Météo-France ne saisisse la main tendue.

On peut dès lors questionner la raison de l'absence de mise à disposition de ces données, qui ne repose **ni sur des éléments de coût** (cela représente l'épaisseur du trait par rapport à la mise à disposition des données modèles et radar), **ni sur une impossibilité technique** (les flux de données sortants existent déjà pour les clients et partenaires de Météo-France). Ces données sont pourtant **primordiales dans bien des usages de suivi du changement climatique**, des sécheresses, de la durée de retour des périodes d'inondations, etc. Elles sont d'ailleurs facturées aux collectivités locales qui souhaitent les utiliser, par exemple dans le cas de Syndicats Mixtes de bassin (EPTB) dans leurs travaux de suivi des aménagements hydrauliques ou de préparation aux événements extrêmes.

Par exemple, l'obtention des séries climatologiques sur une seule station météorologique (sur les 1500 que possède Météo-France) coûte 7000 à 8000€. La mise à disposition de ces mêmes données en temps réel coûte 376€ par an par station, ou bien **126.000€ par an pour l'ensemble des stations de mesure** (réseau dit "Étendu"). On notera également que l'interface de mise à disposition des données est extrêmement complexe d'utilisation, et utilise un système de points pré-crédités rendant difficile d'estimer à l'avance le coût des données⁸.

De la même manière, l'obtention des données historiques du réseau de mesure de précipitations par radar coûte environ 6000€ par année d'archive, et l'abonnement aux données temps réel **150.000€ par an, et quasiment 270.000€ par an en incluant l'outre-mer**.

⁴ 100Go par jour de bande passante sortante coûteraient 365€ HT/an à Météo-France chez OVH par exemple : <https://www.ovhcloud.com/fr/public-cloud/prices/#439>

⁵ <https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-surface-land?tab=overview>

⁶ <https://www.ecad.eu/>

⁷ <https://twitter.com/fredericameye/status/1628447646644940802>

⁸ <https://publitheque.meteo.fr/okapi/accueil/okapiWebPubli/index.jsp>

Il est à noter que la “licence spéciale” de Météo-France autorise la rediffusion de ces données. On trouve donc sur internet des sites web commerciaux, réaffichant ces données, généralement en couvrant les frais par la mise en place de publicités ou la revente de données.

3. RÉUTILISATION PAR D'AUTRES SERVICES PUBLICS

Outre les acteurs privés et associatifs, **de nombreux acteurs publics territoriaux ont besoin de données météorologiques et climatiques** pour réaliser leurs missions. C'est par exemple le cas des grandes métropoles et villes (pour l'aménagement des dispositifs liés aux ruissellements urbains et aux inondations, ou les études d'urbanisme liées aux îlots de chaleur urbain), ou des Syndicats Mixtes de gestion des bassins versants (EPTB).

Ces services publics, réalisant une mission de protection des biens et des personnes, doivent pourtant **acheter leurs données à Météo-France**, et ne bénéficient pas d'accord particuliers de mise à disposition ; seuls les organismes de recherche en bénéficient, sur mise en place de conventions particulières et avec des critères stricts d'utilisation des données.

L'association Infoclimat a été confrontée à de nombreux témoignages de **collectivités qui produisent leurs propres données** en installant des stations météorologiques, **réutilisent celles des communautés open-data** (comme Infoclimat), en **achètent au secteur privé** (EDF⁹, HD-RAIN¹⁰), mais... **ne disposent pas des données de Météo-France**, les factures étant considérées comme trop élevées. Dans certains cas, les collectivités font appel à Predict, filiale de Météo-France, qui achète les données à la branche commerciale de Météo-France et les revend sous forme de produits à valeur ajoutée.

Cette situation rend le travail de nombreux agents des collectivités difficile, et **ne leur permet pas de prendre les décisions les plus éclairées**, par manque de données disponibles.

4. LE CLIMAT, UNE SCIENCE DU PARTAGE : comparaison avec les autres pays

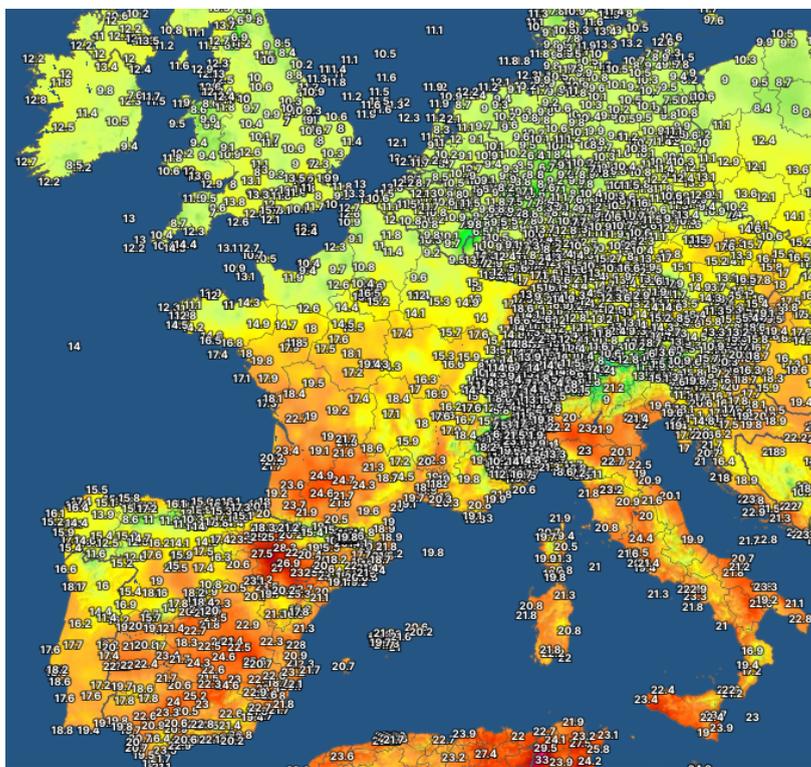
La météorologie a toujours été une **science du partage**, à cause du fameux effet papillon : les conditions météorologiques d'aujourd'hui aux États-Unis ont un impact sur les conditions qui seront observées dans cinq jours en Europe. La coopération internationale a donc toujours été favorisée, notamment par l'Organisation Mondiale de la Météorologie, et par les organismes de sûreté aéronautique.

L'OMM a voté en 1995 la “Résolution n°40”, qui définit un set minimum de données essentielles que les pays membres doivent mettre à disposition gratuitement à tout un chacun.

⁹ EDF dispose historiquement d'un réseau important et ancien de mesures météorologiques, pour la sûreté et l'exploitation de ses ouvrages hydrauliques. Les acteurs de l'éolien et du photovoltaïque développent aussi d'importants réseaux de mesures privée, de même que les acteurs de l'agriculture.

¹⁰ HD-RAIN est une start-up développant un réseau de mesure de précipitations basé sur des antennes satellite et des algorithmes d'intelligence artificielle, complémentaire aux réseaux traditionnels de pluviomètres et de radars.

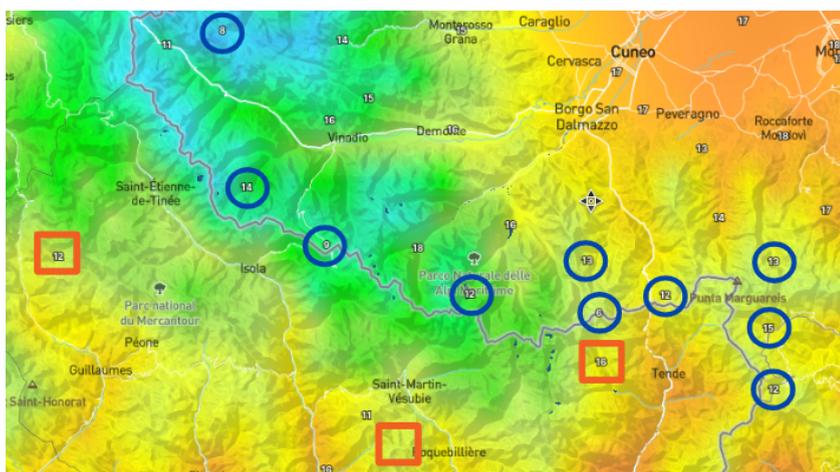
Météo-France ne s'y est plié que 12 années après le vote de cette résolution, notamment par l'insistance des acteurs du Conseil Supérieur de la Météorologie, et a défini une liste de **62 stations** météorologiques¹¹ qui restent aujourd'hui, en 2023, les seules données d'observation de stations météorologiques qui sont placées sous licence ouverte émises par l'organisme. En outre, leurs archives ne sont disponibles que depuis 1996, quand ces données représentent pourtant un levier d'étude du climat, certains de ces points de mesure ayant des relevés depuis la fin des années 1880.



Si Météo-France s'est strictement limité à la Résolution n°40 depuis 1996, **d'autres pays ont décidé d'une plus grande ouverture**, et ont mis tout leur réseau en accès libre depuis de nombreuses années, la comparaison de la densité des points de mesure en libre accès en Europe de l'Ouest étant frappante (ci-contre, *note: depuis la production de cette infographie, l'Italie dispose également de données libres d'accès*). De nombreux autres pays du monde adoptent une stratégie d'ouverture large, comme l'Australie, les pays scandinaves, le Japon, et les États-Unis.

La situation en zone montagneuse est particulièrement frappante.

L'infographie ci-contre présente la situation à la frontière italienne, les cercles bleus étant les stations des services météo officiels, les encadrés rouge les stations associatives. Aucune station de Météo-France ne permet aux usagers de la montagne de disposer des conditions météorologiques !



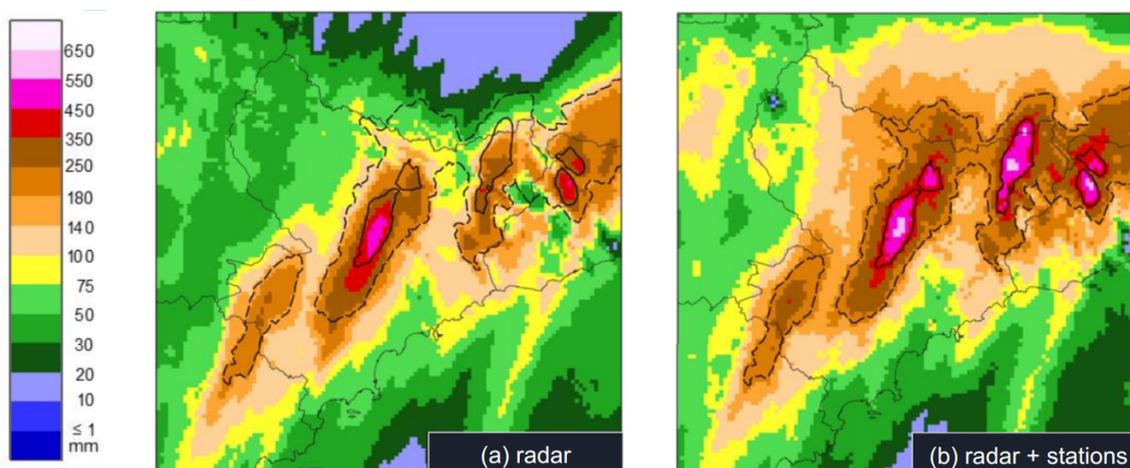
En Allemagne¹², Finlande¹³, Norvège¹⁴, Danemark¹⁵, Suisse¹⁶ pour ne citer que quelques exemples, les **données climatologiques et météorologiques complètes sont en libre accès**. Pour les

¹¹ https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32
¹² <https://opendata.dwd.de/>
¹³ <https://en.ilmatietaenlaitos.fi/open-data-sets-available>
¹⁴ <https://www.met.no/en/free-meteorological-data>
¹⁵ <https://confluence.govcloud.dk/display/FDAPJ>
¹⁶ https://opendata.swiss/fr/dataset?keywords_fr=temps-meteo

pays en disposant, les modèles météorologiques et les radars de précipitations sont également en libre accès, dans des formats exploitables informatiquement. **La France fait figure d'exception** parmi les services météorologiques disposant de capacités avancées.

Pourtant, notre pays est un leader dans la recherche en météorologie et en climatologie, et le modèle de prévisions de Météo-France figure parmi ceux ayant les meilleures performances au monde¹⁷.

Au vu **d'événements climatiques de plus en plus violents** et sortant de l'ordinaire, la **complémentarité des moyens de mesure et de modélisation** relève aujourd'hui de l'impérieuse nécessité. Lors de la tempête Alex en octobre 2020, qui a fait 18 morts, les chercheurs de Météo-France, les hydrologues d'EDF, les acteurs associatifs (Nice-Météo 06, Infoclimat), les services météorologiques italiens, les acteurs privés (NOVIMET) et territoriaux (SMIAGE) ont réuni, gracieusement et en quelques jours, le meilleur de leurs données et algorithmes pour reconstituer les précipitations de l'événement avec précision :



Ces cartes représentent la quantité d'eau tombée durant la tempête Alex sur le département des Alpes-Maritimes. La visualisation de gauche représente la donnée opérationnelle de Météo-France, disponible au moment de l'événement, la visualisation de droite représente la visualisation obtenue par la mutualisation de toutes les sources de données, obtenue quelques jours plus tard¹⁸. Les précipitations, très sous-estimée dans les vallées de la Roya par les seules données de Météo-France, **révèlent des quantités d'eau bien plus importantes lorsque toutes les données sont prises en compte.**

La coopération entre acteurs, et la création d'une **véritable banque de données publiques de la météo**, est primordiale pour l'adaptation au changement climatique. Météo-France doit être un acteur de cette mutualisation, non sous la forme d'un organisme prédateur qui accumule les données météorologiques sans les redistribuer, mais d'une banque ouverte qui **crée un cercle vertueux de production et réutilisation des données.**

¹⁷ http://www.senat.fr/rap/r20-840/r20-840_mono.html#toc149

¹⁸ D'après Chochon et al., 2020 <https://hal.science/hal-03374712/document>

L'association Infoclimat travaille avec les acteurs territoriaux pour les aider dans la production de données, et applique une condition nécessaire à toute aide à l'hébergement et à la valorisation de données : ces dernières doivent être sous licences libres. Météo-France devrait inciter ses partenaires à en faire de même, comme le recommandait en 2018 un rapport non publié de la Cour des Comptes¹⁹ :

Dans un souci de clarification, les trois établissements pourraient sensibiliser leurs clients et partenaires publics et les inciter à ouvrir, sur une base volontaire et hors restrictions prévues aux articles L. 311-5⁶⁸, L. 311-6⁶⁹ et L. 312-1-2⁷⁰ du CRPA, les informations engendrées par les prestations réalisées. Cela pourrait se faire notamment en insérant, sauf demande contraire du client, une clause facultative d'ouverture des données dans les conventions types⁷¹. Cette clause pourrait, par exemple, charger l'établissement de la mise en œuvre des obligations d'ouverture des données générées par les prestations réalisées afin de sécuriser juridiquement son client.

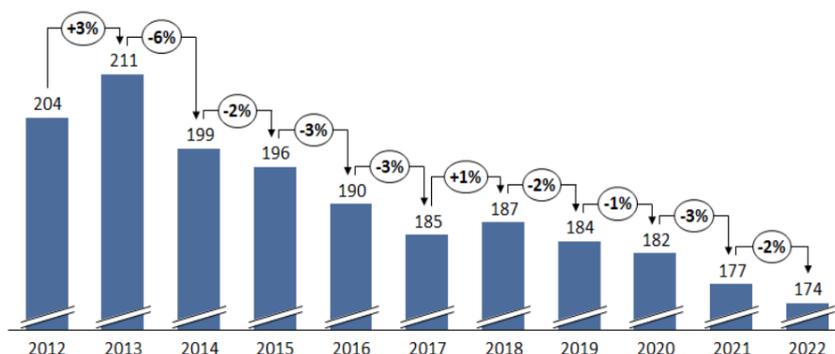
Projet de recommandation n° 2. Sauf demande contraire du client ou du partenaire public, proposer une clause d'ouverture des données dans les conventions types (IGN, Météo France, CEREMA).

5. QUELS IMPACTS ÉCONOMIQUES ?

La production de données météorologiques a un coût non négligeable pour le service public (installation de radars météorologiques, qualification et maintien d'un réseau de stations de mesure météorologique et nivologique, infrastructure informatique). Météo-France a, au cours des dernières années, eu une **politique de rationalisation de ses dépenses**, notamment par la réduction du nombre d'antennes locales et de points de mesure météorologiques. Cela s'explique en partie par la baisse de certaines recettes (des services au grand public "Audiotel" par exemple), ainsi que de la **subvention pour charges de service public** qui baisse en moyenne de 1.6% par an :

Évolution de la subvention pour charges de service public (2012-2022)

(en millions d'euros)



Source : commission des finances du Sénat d'après les documents budgétaires

¹⁹ Rapport S2018-1218, 2ème Chambre, 2ème Section "La valorisation des données de l'IGN, de Météo France et du CEREMA".

En parallèle, Météo-France dispose de **recettes via sa branche commerciale**, sa spin-off Predict, ainsi que d'acheteurs variés qui vont des particuliers aux multinationales en passant par les PME et start-ups. Il est difficile d'estimer l'impact réel de l'ouverture des données sur le budget de Météo-France, car **les comptes publiés ne détaillent pas la différence entre frais de mise à disposition et redevances, ni la ventilation des frais liés à cette mise à disposition**. En mars 2023, Météo-France indiquait²⁰ qu'environ **2M€ de CA** étaient générés par la vente de produits, dont 0.6M€ pour la vente de produits de climatologie archivée (plateforme "Publithèque"). Se pose également la question de la part de ces 2M€ que représente l'achat des données par la branche commerciale de MF à la branche service public de MF. Le CA "hors-MF" en est d'autant réduit.

Cependant, un rapport remis par Deloitte dans le cadre du projet européen de directive PSI2 a analysé le retour d'expérience des pays européens ayant basculé vers une ouverture large de leurs données.

En Allemagne, les revenus de la vente de données météorologiques se sont éteints, et représentaient environ **2.3M€ par an**. Le gouvernement fédéral a compensé cette baisse par une allocation supplémentaire de crédits au service météorologique allemand.

Au Danemark, le service météorologique a estimé que la mise en conformité au règlement HVD coûterait 5.4M€ sur 4 ans, la perte de revenus étant du même ordre de grandeur. En Pologne, le service météorologique va créer une nouvelle plateforme dont les coûts sont estimés à 450k€, la Slovaquie estime ces coûts à 2M€, et estime avoir besoin de 2 ETP supplémentaires pour le support aux utilisateurs de la plateforme de données.

En France, il a été estimé que 10 ETP seraient nécessaires pendant 3 à 5 ans pour se mettre en conformité avec le règlement HVD, ce qui représente **moins de 5% du budget annuel de l'établissement**, alors qu'il est évalué qu'1€ de budget alloué à Météo-France équivaut à 4 à 8€ de bénéfices socio-économiques²¹.

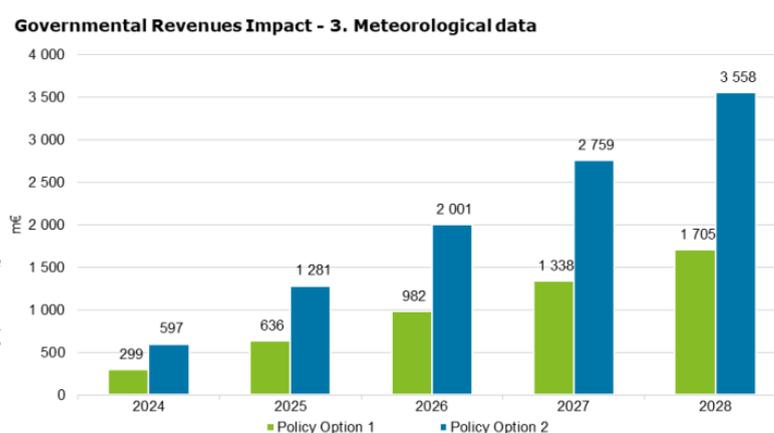
En outre, le rapport Deloitte précise que d'ici 2028, **1 à 2 Mds d'euros de bénéfices économiques directs** seraient générés par le passage à l'open-data complet des données météorologiques en Europe, et **3 à 7 Mds d'euros de bénéfices indirects**. Cela s'accompagne également de la **création de 40.000 à 80.000 emplois** pour accompagner l'utilisation de ces données.

L'accroissement de la création de valeur économique (par la création de nouvelles start-ups par exemple) provoque des

²⁰ Réunion de la Commission Observation du Conse Desponts

²¹ "LES BÉNÉFICES SOCIOÉCONOMIQUES GÉ <https://www.senat.fr/rap/r20-840/r20-8405.htm>

Figure 35 - Governmental revenues impact - Meteorological data



bénéfices supplémentaires pour l'État, par les taxes notamment, dont le montant surpasse aisément les coûts de la mise en accès libre des données²² (ci-contre).

La mise en place d'une politique d'ouverture de ces données ouvre la voie à **l'utilisation "créative" de données précédemment peu valorisées** (par exemple les données de radiations solaires, ou les données climatologiques anciennes aidant à la prise de décision).

Pour Météo-France, cela signifie également une **réorganisation de ses ressources humaines**, les ETP actuellement affectés à la fourniture de données sous licence (extraction de données du SI de Météo-France, bons de commande, maintien de la plateforme de commande payante "Publithèque"...) pouvant être réallouées à des activités de support technique aux utilisateurs des nouvelles API Open-Data par exemple.

6. QUELLES DONNÉES ?

Le règlement européen PSI2/HVD précise un certain nombre de jeux de données qui doivent être mis à disposition :

Datasets	Observations data measured by weather stations	Climate data: validated observations	Weather alerts	Radar data	NWP model data
<i>Granularity</i>	Per weather station, full temporal resolution	Per weather station, full temporal resolution	Alerts, 48 hrs or more ahead	Per radar station in the MS and national composite	Minimum 48 hrs ahead in 1hr steps, national, at 2.5km/best available grid
<i>Key attributes</i>	All observation variables measured	All validated measured observation variables; daily average per variable		Reflectivity, Backscatter, polarization. Precipitation, wind, and echotops	Deterministic and/or ensembles if available, for meteorologically relevant parameters and levels

<i>Format</i>	BUFR, NetCDF, ascii (for real time data), JSON (for hourly data)	NetCDF, JSON	XML (Cap or RSS / Atom)	HDF5, JSON	GRIB (or NetCDF), JSON
<i>Update frequency and timeliness</i>	Every 5-10 minutes in real time for automated stations, hourly unvalidated for all stations, for the last 24 hrs	Daily validated hourly (and better temporal resolution) and daily average observations data; all digitised historical data	As issued or hourly	Near real time in 5 minute intervals (or available shortest interval)	Every 6 hrs, or better temporal resolution, from the last 24 hrs.

À ce jour, ces critères ne sont que très partiellement remplis :

- "Observations data" : **Météo-France ne diffuse que 62 stations sur les 1700 de son réseau²³**, et ne les diffuse qu'à une fréquence de 3 heures, quand elle en dispose au pas de temps 6 minutes voire 1 minute. Aussi, les archives de ces 62 stations ne remontent que jusqu'à 1996 quand Météo-France dispose de bien plus de données.
- "Climate data" : Météo-France ne diffuse actuellement aucune de ces données.
- "Weather alerts" : Météo-France diffuse ces données de manière correcte depuis 2021.

²² Source : Deloitte 2020, §3.3.2.1

²³

https://publicwiki.meteo.fr/download/attachments/23267371/20201126_CSM_OBS_Observer-Surface.pdf?version=1&modificationDate=1606378039695&api=v2

- “Radar data” : Météo-France ne diffuse actuellement aucune de ces données.
- “NWP model data” : Météo-France diffuse très partiellement ces données, dans des conditions de distribution faiblement performantes.

Une inquiétude importante est présente dans la communauté des utilisateurs de données météorologiques, notamment liées au fait que de nombreux produits ne sont pas listés, et pourraient être délibérément exclus de l'ouverture par Météo-France :

- les **produits de fusion de données**²⁴ notamment radar (ANTILOPE, PANTHERE), primordiaux pour l'analyse des événements de précipitations extrêmes, ne sont pas listés. Ces produits incluent parfois des données d'acteurs privés avec lesquels Météo-France est partenaire (EDF, NOVIMET), ce qui rend leur situation juridique ambiguë. Les algorithmes utilisés pour générer ces produits de fusion ne sont en outre pas publics.
- les **produits climatologiques spatialisés**²⁵, comme les jeux de données COMEPHORE (réanalyses pluviométriques) ou les analyses de surface (normales climatologiques spatialisées).
- les **indicateurs de suivi climatique**, comme l'indicateur thermique national ou l'indicateur de précipitations et de sécheresse, dont ni les données ni la méthodologie de calcul ou les algorithmes ne sont diffusés.
- les **modèles de surface ou d'enneigement** (humidité des sols, SAFRAN), qui devraient être considérés comme “NWP model data”, ne doivent pas être oubliés et représentent une valeur considérable pour nombre de réutilisateurs (agriculture, tourisme/montagne, risques inondation et aléa retrait-gonflement d'argiles). Il en est de même pour les modèles des **conditions routières** (OPTIMA) ou de risques de feux de forêt , dont ni les codes ni les données ne sont diffusés.
- les **données de foudroiement** sont particulières, car elles sont produites par une filiale de Météo-France et de la société privée finlandaise Vaisala : Météorage S.A.²⁶. La situation légale de ces données est peu claire.

Météo-France indique préférer s'intégrer dans le **projet européen EUMETNET E-SOH**²⁷, qui est une plateforme de partage de données d'observation météorologique, plutôt que de réaliser ses propres développements. Si cette volonté est louable, cette plateforme ne couvre cependant qu'une partie des données citées ci-dessus, et l'exécution des Work Packages futurs du système ne sont pas assurés, qui laisse craindre une plateforme qui ne sera jamais accessible aux citoyens, mais seulement aux services météorologiques nationaux européens.

²⁴

https://publicwiki.meteo.fr/download/attachments/23267371/20201126_CSM_OBS_PPC.pdf?version=1&modificationDate=1606289814978&pi=v2

²⁵ https://publicwiki.meteo.fr/download/attachments/23267371/LeMoigne_CSMOBS.pdf

²⁶ <https://meteofrance.fr/etablissement/filiales/les-filiales-de-meteo-france>

²⁷

https://publicwiki.meteo.fr/download/attachments/23267371/CSM_20220325_ESOH_v2.pdf?version=1&modificationDate=1648195056647&pi=v2

7. QUELS SERVICES ET QUELS CHOIX TECHNIQUES ?

Les jeux de données météorologiques représentent des **typologies de données qui peuvent être difficiles à distribuer massivement** :

- les modèles numériques de prévision sont **mis à jour 4 à 24 fois par jour**, et les **volumes de données sont importants** (100 Go à 1 To par jour). En outre, chaque ré-utilisateur souhaite télécharger, à chaque actualisation, le même jeu de données, en même temps et le plus rapidement possible, ce qui sollicite considérablement les infrastructures informatiques, qui sont alors généralement surdimensionnées.
Des solutions techniques existent pourtant (partage en pair-à-pair), et des stratégies de lissage des pics existent (souscription à des APIs “push/sub”, API permettant l'accès à des sous-ensembles, étalement dans le temps des pics), outre les technologies classiques du Cloud Computing (“Object Storage”).
- les données météorologiques d'observation (stations météorologiques) représentent des **mises à jour très fréquentes de petits volumes de données**, auxquels de nombreux utilisateurs souhaitent accéder simultanément, ce qui peut également occasionner une surcharge des infrastructures. Là aussi, des solutions techniques modernes existent pour limiter la charge des infrastructures et la consommation de bande passante (publish/subscribe), et sont déjà mis en œuvre par le service météorologique américain par l'entremise d'Amazon Web Services ; des équivalents français existent et sont à privilégier (OVH, Scaleway).
- les autres données ne représentent aucun challenge de diffusion, s'agissant de volumes de données faibles et moins souvent mis à jour.

Au global, si l'on estime la diffusion de 500 Téraoctets par mois (hypothèse très haute, la diffusion des données publiques Météo-France étant estimée à 60 To/mois en 2021²⁸), les coûts de bande passante s'élèvent à moins de 5000€ par mois chez des fournisseurs tel qu'OVH ou Scaleway (tarif avant négociation).

Des solutions techniques modernes permettraient de réduire considérablement les coûts d'infrastructure et de bande passante, en adoptant la bonne stratégie de diffusion.

Météo-France, qu'elle réalise en interne la plateforme technique de diffusion de données publiques ou qu'elle la sous-traite, devrait s'entourer d'acteurs utilisateurs de données météorologiques et d'experts du Cloud Computing, afin d'adopter des techniques économiquement et techniquement viables. Les auteurs proposent en Annexe A une analyse technique de ces problématiques, et des préconisations de solutions. La DINUM a également proposé ses services afin de prendre en charge la distribution de ces données par le biais de data.gouv.fr.

²⁸ communication personnelle des auteurs avec M. Jean-Marie Carrière et M. Christophe Morel

8. Synthèse des recommandations

8.1. Disposer d'une plateforme technique stable

Les utilisateurs doivent bénéficier d'une plateforme leur permettant d'avoir confiance dans la donnée diffusée par Météo-France.

8.2. Être accompagné dans les choix de la plateforme technique

Météo-France doit développer sa plateforme Données Publiques en concertation avec les utilisateurs et en suivant les standards de l'industrie, afin de satisfaire les cas d'utilisation les plus fréquents, et rendre plus efficace sa plateforme, limitant ainsi les coûts d'opération du service.

8.3. Coopérer avec les acteurs (européens ou français)

Météo-France doit coopérer autant que possible avec les acteurs publics, associatifs et privés lui permettant d'obtenir une plateforme technique robuste et adaptée aux cas d'usage et aux standards de l'industrie (DINUM, Infoclimat, OVHcloud, Scaleway,...). Météo-France doit également s'appuyer sur les plateformes et modalités de diffusion existantes, lorsqu'elles sont disponibles (exemple : Copernicus).

8.4. Éliminer les systèmes "Premium" pénalisant les petits utilisateurs

Les utilisateurs doivent être sur un pied d'égalité, quels que soient leurs moyens ; il n'est pas acceptable que les utilisateurs gratuits aient des fichiers corrompus, incomplets, ou n'aient accès aux données qu'après de longues heures, rendant l'information déjà obsolète avant d'avoir été reçue. La solution du système Premium payant doit être supprimée, en faveur d'un système gratuit plus optimisé.

8.5. Compléter le catalogue existant avec les produits manquants

Météo-France doit continuer le recensement de ses produits, permettant aux utilisateurs de les connaître et d'envisager des utilisations novatrices. De nombreux produits (notamment de fusion de données ou de réanalyses) sont aujourd'hui non-listés, et leur mise en open-data doit être évaluée.

8.6. Engager une politique d'incitation des partenaires

Par son statut d'acteur de référence de la météorologie, Météo-France doit inciter ses partenaires à rediffuser leurs données sous licences ouvertes, en leur proposant en échange leur hébergement et qualification sur sa plateforme de données ouvertes, créant ainsi un véritable hub de données climatiques publiques. Cela peut être le cas des collectivités territoriales, ou des partenaires privés.

9. Références

- Rapport de la Cour des Comptes N°S2018-1218 (non publié, communication personnelle)
- Réponse de M. Edouard PHILIPPE, Premier Ministre, au référé de la Cour des Comptes <https://www.ccomptes.fr/sites/default/files/2019-03/20190311-refere-S2018-3287-valorisation-donnees-IGN-Meteo-France-Cerema-rep-PM.pdf>
- Analyse d'impact de la directive PSI2 "High Value Datasets", Deloitte, 2020 <https://www.access-info.org/wp-content/uploads/Deloitte-Study-2020.pdf>
- Rapport d'information n°840 du Sénat, 2020 <https://www.senat.fr/rap/r20-840/r20-840.html>
- Contrat d'Objectifs et de Performance de Météo-France, 2020 https://meteofrance.fr/sites/meteofrance.fr/files/files/editorial/Rapport_COP_V14.pdf

Annexe A

Recommandations techniques pour la diffusion des données ouvertes Météo-France

Recommandations techniques pour la diffusion des données ouvertes Météo-France

1. Introduction	2
2. Données d'observation des stations météorologiques	3
2.1. Cas d'usage	3
2.2. Proposition technique	3
2.2.1. Cas temps réel	3
2.2.2. Cas archives récentes (24h)	4
3. Données climatiques passées ponctuelles (stations météorologiques)	5
4. Vigilances météorologiques	6
5. Données radar ou assimilées (fusion de données)	6
6. Données climatiques spatialisées ou réanalyses	6
7. Données de prévisions	7
7.1. Cas général des données de modèles numériques	7
7.2. Prévision ensembliste	8
8. Considérations génériques	9
8.1. API et librairies	9
8.2. Authentification des clients	9
8.3. Auto-hébergement, cloud, scalabilité	9

1. Introduction

Le projet de la Commission Européenne de règlement "*High Value Datasets*" spécifie un ensemble de données à diffuser largement auprès du public. Parmi divers secteurs, les données météorologiques, correspondant à une partie des données produites par Météo-France, sont spécifiées dans l'annexe du règlement comme suit :

Datasets	Observations data measured by weather stations	Climate data: validated observations	Weather alerts	Radar data	NWP model data
<i>Granularity</i>	Per weather station, full temporal resolution	Per weather station, full temporal resolution	Alerts, 48 hrs or more ahead	Per radar station in the MS and national composite	Minimum 48 hrs ahead in 1hr steps, national, at 2.5km/best available grid
<i>Key attributes</i>	All observation variables measured	All validated measured observation variables; daily average per variable		Reflectivity, Backscatter, polarization. Precipitation, wind, and echotops	Deterministic and/or ensembles if available, for meteorologically relevant parameters and levels

<i>Format</i>	BUFR, NetCDF, ascii (for real time data), JSON (for hourly data)	NetCDF, JSON	XML (Cap or RSS / Atom)	HDF5, JSON	GRIB (or NetCDF), JSON
<i>Update frequency and timeliness</i>	Every 5-10 minutes in real time for automated stations, hourly unvalidated for all stations, for the last 24 hrs	Daily validated hourly (and better temporal resolution) and daily average observations data; all digitised historical data	As issued or hourly	Near real time in 5 minute intervals (or available shortest interval)	Every 6 hrs, or better temporal resolution, from the last 24 hrs.

Ces données sont de types, profondeurs temporelles, fréquences et volumes variés, et nécessitent des technologies de dissémination au public différentes, afin d'optimiser les coûts machine et les coûts de bande passante, et optimiser l'expérience d'utilisation par le public de ces données.

Infoclimat, en tant qu'association représentante d'une très grande communauté d'utilisateurs de données météorologiques et climatiques (2200 adhérents), et pionnière dans la mise à disposition d'un commun numérique météorologique (par des APIs) et par l'utilisation d'APIs externes des services météo nationaux étrangers, tente ici de proposer quelques solutions techniques adaptées aux besoins, et propose de mettre sa connaissance de ces techniques et de leur scalabilité au service de Météo-France. Cela permet au global une meilleure efficacité de la distribution de ces données, à coûts (OPEX) très réduits.

2. Données d'observation des stations météorologiques

Ces données représentent l'ensemble du réseau de surface de Météo-France. Toutes les variables doivent être fournies :

- en temps réel à la fréquence la plus élevée possible (1mn, 6mn)
- sur les dernières 24h à la fréquence horaire ou meilleure

2.1. Cas d'usage

Dans le cas "temps réel", les utilisateurs cherchent à obtenir, tous en même temps, le plus rapidement possible, la dernière donnée pour une ou toutes les stations. En utilisation classique d'une API web (type REST), cela représente des milliers de requêtes HTTP à des intervalles très réguliers, ce qui embouteille les serveurs et provoque des pics de charge difficiles à supporter. Les clients réalisent fréquemment, tous en même temps, des requêtes (par exemple toutes les 5 minutes) pour vérifier la fraîcheur des données, et ce comportement doit être évité. Cela est typiquement résolu en architecture logicielle par un fonctionnement de type "*publish-subscribe*", proposé ci-dessous.

Dans le cas "archives récentes" (moins de 24h), il s'agit de requêtes plus éparées, sur une ou plusieurs stations en même temps. Si le cas "temps réel" est bien implémenté (et donc plébiscité par les utilisateurs), ce mode ne doit permettre que de récupérer par exemple des données a posteriori (récupérées après une coupure de courant ou de réseau) ou de récupérer des correctifs liés à des éditions humaines de valeurs erronées.

2.2. Proposition technique

2.2.1. Cas temps réel

Mise en place d'un serveur de souscription "*publish-subscribe*", où un client peut "souscrire" à un ou plusieurs canaux de données (1 canal = 1 station météo), et recevoir les données ("notification") dès qu'elles sont poussées ("*publish*") par le SI de MF. Cette approche est très scalable et nécessite très peu de ressources : typiquement plusieurs millions de clients peuvent être adressés en même temps, avec des milliers de canaux, avec une seule machine. La bande passante est également réduite, car les clients évitent de réaliser de trop nombreuses requêtes de vérification de présence de nouvelles données.

Un client qui souhaite souscrire à plusieurs stations météo peut souscrire à plusieurs canaux. La notification envoyée à tous les clients contient une charge utile qui est la donnée météo la plus récente (par exemple sérialisée au format JSON). Le serveur ne conserve pas d'historique, et se charge de disséminer les données temps réel ; charge au client (utilisateur) d'historiser ces données. Ce type de serveurs permet, avec une unique machine, d'adresser des millions de clients qui souscrivent à plusieurs milliers de canaux.

Exemple de logiciels (« brokers ») : Mosquitto (MQTT), Apache Kafka, HiveMQ (MQTT), socket.io, ZeroMQ, RabbitMQ (recommandé)...

Exemple de services cloud : AWS SNS, Scaleway IoT Hub, Azure PubSub...

Le service météorologique américain, la [NOAA](#), utilise déjà [Amazon SNS](#) pour notifier de nouvelles images GOES disponibles. Dans le cas de données d'observation (stations météorologiques, vigilances), les données sont suffisamment petites pour être incluses dans la notification.

Infoclimat utilise déjà avec succès un modeste serveur websocket (nginx) pour transmettre des millions d'impacts de foudre en direct vers des milliers de clients en simultané.

2.2.2. Cas archives récentes (24h)

Un client peut souhaiter accéder aux archives d'une seule station météo, ou de plusieurs. La stratégie adoptée par la DWD est satisfaisante : mise à disposition d'un fichier pour chaque station (contenant les 24 dernières heures), éventuellement compressé, référencé par l'identifiant unique de la station météo. La mise à disposition se fait préférentiellement par HTTP(S), qui permet des performances optimales. Sur ce type de fichiers statiques, une infrastructure bien conçue est capable de supporter les appels de charge sans problème.

La date de dernière mise à jour des fichiers doit être répertoriée dans un fichier d'index (index.json, index.xml), permettant aux clients d'éviter de re-télécharger des fichiers qui n'ont pas évolué (il est également possible de le gérer avec le mécanisme *HTTP HEAD / HTTP NOT MODIFIED*, mais cela est généralement méconnu par les développeurs).

Mieux encore, un mécanisme *publish/subscribe* (voir 2.2.1) peut être mise en place pour qu'un client soit notifié de mises à jour sur les fichiers, et les télécharger seulement au bon moment. L'envoi de correctifs (valeurs erronées corrigées, ou valeurs absentes récupérées a posteriori) doit pouvoir être tracé, afin que les clients puissent ne télécharger que le nécessaire, sans surcharger inutilement l'infrastructure.

3. Données climatiques passées ponctuelles (stations météorologiques)

Ces données représentent un volume important, que les utilisateurs voudront récupérer :

- en un seul lot (toutes les stations, toutes les archives), afin d'alimenter une base de données locale globale
- les différences depuis leur dernier téléchargement du lot complet (toutes les stations, toutes les archives)
- toutes les données d'une station météo particulière
- toutes les données d'une période particulière, pour toutes les stations

Les données sont à la fois des données horaires validées, et des données journalières validées. Pour une station ayant débuté ses archives en 1900 (hypothèse large), cela représente 1 million de lignes données horaires et 50.000 lignes de données journalières.

La stratégie adoptée par la DWD et la NOAA est satisfaisante à bien des égards, et permet de répondre à beaucoup d'usage en conservant des fichiers modestes :

- mise à disposition de fichiers journaliers pour chaque station et toute la période
- mise à disposition de fichiers horaires pour chaque station, par année

En complément, il est conseillé ici de mettre à disposition pour chaque station un fichier ou une API de listing de chaque date de mise à jour des fichiers de données, pour informer les utilisateurs de données ayant été modifiées (correction d'erreur par un humain par exemple). Cela permet au client de ne retélécharger que le nécessaire (que l'année 1976 pour la station N°07015 par exemple).

4. Vigilances météorologiques

La fourniture actuelle sous forme de fichiers XML est satisfaisante, mais mériterait un système de notification de dernière mise à jour plus poussé (*publish/subscribe*, voir section 2.2.1) afin d'éviter que les clients ne téléchargent régulièrement un fichier pour vérifier la dernière mise à jour du fichier XML. Cela économise quelques ressources machine.

5. Données radar ou assimilées (fusion de données)

Les données radar représentent des données spatialisées de type raster, que les utilisateurs veulent obtenir de deux manières principales :

- télécharger le plus rapidement possible les nouvelles images;
- télécharger des archives d'images par lots de grands volumes.

La proposition technique pour répondre à ces besoins est la suivante :

- mettre à disposition un canal auquel les clients peuvent souscrire ("*subscribe*", cf 2.2.1) pour être avertis de la disponibilité d'une nouvelle image/d'un nouveau fichier ; a minima si un système *publish/subscribe* n'est pas possible, mettre à disposition une API légère ou un fichier texte permettant d'obtenir les informations (*timestamp*) de dernière mise à jour
- mettre à disposition les images archivées sous la forme de fichiers groupés par année ou par mois, contenant la totalité des images sous la format de bandes ou de *subdatasets* (NetCDF, GRIB).

Afin de réaliser des économies de bande passante, la mise à disposition des images (en temps réel ou archivées) peut également être envisagée sous un modèle *peer-to-peer*, bien que le volume des données soit faible par rapport aux données modèles.

6. Données climatiques spatialisées ou réanalyses

Considérations identiques à la section 5.

7. Données de prévisions

7.1. Cas général des données de modèles numériques

Le cas général consiste en de très nombreux clients qui souhaitent récupérer le plus rapidement possible un "run" de modèle, 4 fois par jour, pour un sous-ensemble de la zone géographique ou bien la totalité, pour un sous-ensemble de paramètres ou tous les paramètres.

Le besoin de récupérer rapidement le début d'un *run* oblige généralement les clients à réaliser très fréquemment des requêtes aux serveurs de Météo-France afin de détecter la présence de nouveaux fichiers. Cela occasionne une consommation de bande-passante et de ressources serveurs qui est inutile. De plus, la récupération simultanée des mêmes fichiers par tous les clients occasionne des pointes de charge au moment de la mise à disposition des runs, alors que les systèmes sont en sous-charge le reste du temps.

Un certain nombre de contre-mesures peuvent être mises en place :

- les clients doivent être avertis de la mise à disposition de nouveaux fichiers (groupes d'échéances ou échéances pour un run par exemple) sans avoir à solliciter inutilement le serveur de manière fréquente. Pour cela, la mise en place d'un canal de souscription (*publish/subscribe*) ou d'une API légère de métadonnées (correctement mise en cache) permet de limiter la charge de l'infrastructure.
- afin de lisser les pics de trafic lors de la sortie de nouveaux groupes de fichiers, il serait, dans le pire des cas et bien que cette technique ne soit pas recommandée, possible de recourir à de l'étalement dans le temps de la charge, par exemple :
 - Météo-France propose une librairie (ex: Python) réalisant de manière autonome le téléchargement, que le client n'a qu'à utiliser dans son application. En interne, la librairie introduit un délai aléatoire de quelques secondes à avant de lancer le téléchargement d'un nouveau groupe de fichiers disponibles. Si ce comportement diminue la rapidité de mise à disposition des fichiers, elle permet d'éviter de surdimensionner l'infrastructure et d'éviter son effondrement aux minutes où les clients tentent tous de télécharger les fichiers simultanément. Cette technique dégrade tout de même l'expérience utilisateur, et les technologies d'Object Storage y sont préférables, permettant de ne pas se préoccuper du *scaling* de l'infrastructure, et donc plus indiquées pour cet usage.

-
- Une solution préférentielle serait que Météo-France diffuse les données par le biais d'un mécanisme de *peer-to-peer* (eg. BitTorrent), et s'associe à la communauté de réutilisateurs et de l'open-data (Infoclimat, OpenStreetMap, OpenMeteoData,...) pour mettre en place des relais de diffusion des données. L'intégrité des données et l'authenticité peut être assurée par la diffusion par Météo-France de fichiers contenant les hashes cryptographiques des données diffusées, s'assurant ainsi de la non-modification des données par des acteurs malveillants. Cette solution est la plus adaptée à ce type de mise à disposition de fichiers, où tous les acteurs souhaitent récupérer le même fichier au même moment.

Le volume de données diffusées peut également être réduit en mettant en place des APIs répondant sur des coordonnées géographiques (points ou polygones), permettant de récupérer les données sur un sous-ensemble géographique souhaité par l'utilisateur. Cela nécessite pour Météo-France de développer une indexation correcte des données brutes des modèles météorologiques, afin de limiter la charge processeur et des I/O disque d'une telle opération.

7.2. Prévision ensembliste

Un cas d'usage particulier de la prévision ensembliste consiste à obtenir les données de tous les membres de l'ensemble, pour une zone géographique déterminée (point/polygone). La conception par Météo-France d'une API permettant cet usage éviterait à de nombreux utilisateurs de télécharger l'ensemble des ensembles pour la totalité d'une zone géographique, économisant ainsi une quantité importante de bande passante et de ressources serveur.

Ainsi, seuls les clients souhaitant effectuer des calculs sur des zones plus larges (exemple : cartes de probabilité ou cartes "spaghetti", fusion multi-modèles) auraient besoin de télécharger les fichiers bruts très volumineux.

8. Considérations génériques

8.1. API et librairies

Des **exemples clairs et des codes sources** (Python notamment), voire des librairies, doivent être fournis pour que les **utilisateurs adoptent les bonnes pratiques** dès le début, et évitent de **surcharger les serveurs de MF**. Cela est coûteux, mais bénéfique pour éviter les usages qui effondreront l'infrastructure (requêtes trop fréquentes). La fourniture sous forme de librairie simple (module Python) permet une réutilisation directe, et dirige les utilisateurs vers le bon usage de l'infrastructure MF, qui optimise les coûts d'opération.

8.2. Authentification des clients

Du **point de vue de la sécurité**, les "APIs" proposées ici sont identiques pour tous les clients, ce qui ne nécessite **aucune authentification** (par un token ou OAuth par exemple). C'est la méthodologie utilisée par nombre de serveurs OpenData dans le monde. Les abus peuvent être adressés par l'adresse IP, mais les solutions proposées ici ne sont pas incompatibles avec la gestion d'authentification, qui peut se faire par exemple sur des services type S3 avec des liens temporaires authentifiés. Il est à noter que les apports d'une authentification de chaque client doit bien être mise en balance avec les importants coûts (de développement, de complexité pour l'utilisateur, de CPU) induits.

8.3. Auto-hébergement, cloud, scalabilité

Météo-France héberge actuellement les services Données Publiques dans sa propre infrastructure. Néanmoins, au vu de la quantité et des bande-passante futures à diffuser, Météo-France peut être tenté de basculer vers un **modèle externalisé**, en utilisant par exemple les services d'Amazon (AWS), Microsoft (Azure), ou encore OVHcloud, Scaleway, sous la forme de services clé-en-mains (*object storage, notification services, compute instances,...*). Cette solution a l'avantage considérable d'une simplification à l'extrême des développements côté Météo-France, d'une standardisation des outils, et bénéficie de l'écosystème existant autour de ces technologies. **Les coûts sont à évaluer avec attention**, notamment du point de vue de la bande-passante entrante et sortante, ou des éventuelles limitations (*throttling*) que les opérateurs peuvent mettre en place pour les clients, pénalisant ainsi les performances.

Une autre alternative est d'**héberger soi-même des services "cloud"**, dans sa propre infrastructure (datacenter Météo-France), ou bien dans l'infrastructure d'un autre opérateur (OVHcloud, Integra,...), et de développer toute la couche applicative permettant de proposer les fonctionnalités d'Object Storage par exemple, grâce à des technologies comme OpenStack Swift, Riak S2 ou Minio. Cette solution a l'avantage d'une maîtrise des coûts plus prévisible, mais nécessite une très forte compétence en interne pour la développer et la maintenir.

Une solution alternative pourrait être de démarrer une plateforme sur la base de services "clés-en-mains" standards (S3-like), chez un hébergeur tel que Scaleway ou OVHcloud, tout en **préparant à terme la bascule vers une solution internalisée**, l'utilisation de services standardisés permettant une migration avec le moins d'efforts possible à terme. La bascule entre ces approches est à évaluer en fonction de la maturité technologique des équipes Météo-France, le besoin de sous-traitance, et les coûts d'opération des différentes solutions au fil des années. L'expertise de la DINUM/Etalab devra être mise en complémentarité avec l'expertise de Météo-France pour aboutir à la solution la plus équilibrée possible.