

Séance 1

Des données ouvertes

Des données (aisément) disponibles

Voir site api.nasa.gov

Une première adresse

- Cette Uri renvoie des données, dites ouvertes, de coordonnées et météorologiques de Mars.
- Le site ***mars.nasa.org*** propose ces informations (l'hôte des données est visible dans l'URI)
- Cette information est structurée dans un fichier sous format JSON, sous forme d'objet. Différents types d'informations s'y trouvent, comme des objets, des entiers, des flottants, des dates, des chaînes de caractères.
- Par exemple, la donnée nommée « **First.UTC** » représente le premier fuseau horaire UTC parmi les différents existants. Les données commençant par « **compass** » représentent des coordonnées représenté sur une boussole.

Une deuxième adresse

- Cette Uri renvoie des données concernant des photographies provenant de robots Rover ayant voyagé sur Mars.
- Ces informations proviennent d'une api de la Nasa, comme indiqué dans l'URI : ***api.nasa.gov***
- Cette information est structurée dans un fichier sous format JSON, sous forme d'objet. On peut y trouver différentes type d'informations, comme des noms (chaîne de caractères), des id (entiers), des liens, ou url, d'images, des dates.
- La donnée « **img_src** » représente un lien vers une image. Pour chaque photo, on y trouve un objet camera présentant différents attributs, comme son nom « **name** » ou « **full_name** »
- En comparant cette URI avec une variante de celle-ci, la différence se situe au niveau de la partie *Query* de l'URI : dans la variante, on y retrouve une requête supplémentaire concernant l'un des objets :
camera=fhaz
- Ces résultats différents sont obtenus lors de l'exécution de leur URI respective

Plus près de nous et des préoccupations contemporaines

Ces 2 adresses renvoient des informations sur des zones géographiques situés dans le Grand Est, et leur qualité de l'air.

Dans le tableau ***fileds***, on retrouve des informations uniques concernant chaque zone, comme le nom, l'alias, le type de données SQL.

Dans le tableau ***features***, on retrouve des données concernant la qualité de l'air, donc

multiples par zone.

Quelques utilisations

- Le lien précédemment vu qui affiche des données de qualités de l'air dans le Grand Est serait utile à la mairie de Nancy, afin d'étudier ces données et de lancer une étude sur la qualité de l'air dans l'agglomération de Nancy.
- La grande différence se fait au niveau de la présentation des informations : l'une en données structurées dans un fichier JSON, et l'autre un site web avec une structure HTML /CSS, plus agréable pour un utilisateur lambda.

Des formats variés

- Cette URI renvoie des données sur des prévisions météorologiques concernant une position géographique (coordonnées de celle-ci en requête dans l'URI)
 - Description des champs :
 - échéance : données enregistrées à un instant T
 - température : les différentes températures en Kelvin
 - pression : pression atmosphérique ramenée au niveau de la mer
 - pluie : précipitation tombée en 1h
 - humidité : taux d'humidité %
 - vent moyen : vitesse du vent en moyenne (km/h)
 - vent rafales : vitesse des rafales de vent
 - vent direction : direction du vent en cardinalités (degrés)
 - iso zéro : l'altitude minimale à laquelle la température atteint la valeur de **0** en m
 - risque neige : risque de neige actuel
 - cape : énergie potentielle représentant le niveau de possibilités d'orages en J/kg
 - nébulosité : part de ciel couvert
 - Ce document est structuré sous forme de balise, un format retrouvé dans différents langages comme le XML, HTML, ...
 - Auparavant, les données étaient structurées sous forme d'objets, d'attributs, de tableaux. Nous retrouvons ici un format XML, composé de balises.
-
- Cette URI renvoie des informations concernant le suivi de l'épidémie Covid-19 sur le territoire Français.
 - Description des champs :
 - 'date' = Date
 - 'dep'= Département
 - 'reg'= Région
 - 'lib_dep'= libellé département
 - 'lib_reg'= libellé région
 - 'hosp'= Nombre de patients actuellement hospitalisés pour COVID-19.
 - 'incid_hosp'= Nombre de nouveaux patients hospitalisés au cours des dernières 24h.

- 'rea'= Nombre de patients actuellement en réanimation ou en soins intensifs.
 - 'incid_rea'= Nombre de nouveaux patients admis en réanimation au cours des dernières 24h.
 - 'rad'= Nombre cumulé de patients ayant été hospitalisés pour COVID-19 et de retour à domicile en raison de l'amélioration de leur état de santé.
 - 'incid_rad'= Nouveaux retours à domicile au cours des dernières 24h.
 - 'dchosp'= Décès à l'hôpital
 - 'incid_dchosp'= Nouveaux patients décédés à l'hôpital au cours des dernières 24h.
 - 'esms_dc'= Décès en ESMS (Etablissements et Services Médico-Sociaux)
 - 'dc_tot'= Cumul des décès (cumul des décès constatés à l'hôpital et en ESMS)
 - 'conf'= Nombre de cas confirmés
 - 'conf_j1'= Nombre de nouveaux cas confirmés (J-1 date de résultats)
 - 'pos'= Nombre de personnes déclarées positives (J-3 date de prélèvement)
 - 'pos_7j' = Nombre de personnes déclarées positives sur une semaine (J-3 date de prélèvement)
 - 'esms_cas' = Cas confirmés en ESMS
 - 'tx_pos'= **Taux de positivité** des tests virologiques sur le nombre total de personnes testées positives ou négatives sur une période donnée
 - 'tx_incid'= **Taux d'incidence** : nombre de personnes testées positives (pour la première fois depuis plus de 60 jours pour 100000 habitants.
 - 'TO'= **Taux d'occupation** : tension hospitalière sur la capacité en réanimation
 - 'R'= **Facteur de reproduction du virus** : nombre moyen de personnes qu'une personne infectée peut contaminer
- Cette information est structurée dans un tableau CSV.

Construire ses propres requêtes

Détailler une requête en croisant des informations

Lien liste des stations : www.velostanlib.fr/service/carto

Lien disponibilité station n°27 : www.velostanlib.fr/service/stationdetails/nancy/27

- Les données du lien des listes de stations sont structurées en balises XML, chacune représenté par la balise, ou champ, **marker** (signifie le point géographique) et différents attributs comme l'id, l'adresse, les coordonnées (latitude, longitude), ...

- Les données du lien de la disponibilité d'une station sont structurées en XML, avec différents champs tel que **available** (nombre de vélos disponibles), **free** (nombre d'attaches disponibles), **total** (nombre de points d'attache au total), **open** (statut ouverture de la station), ...
- Dans les données XML de la liste des stations, je recherche une adresse : en face du 1 Boulevard Charlemagne Nancy, je récupère l'id, et l'applique dans ma deuxième requête : www.velostanlib.fr/service/stationdetails/nancy/32
 - available : 5
 - free : 10

Utiliser les requêtes comme outils de recherches ...

1. Open Movie Database

- Cette requête <http://www.omdbapi.com/?apikey=7403b6e6&s=The%20Social%20Network&r=XML> renvoie une structure de données en XML d'une liste de film comportant la chaîne de caractères « The Social Network »
- Le format de retour est en XML (précisé dans l'URI avec l'attribut **r**)
- URI pour trouver l'affiche du film The Social Network par David Fincher :
 - www.omdbapi.com/?apikey=7403b6e6&i=tt1285016&r=XML
 - www.omdbapi.com/?apikey=7403b6e6&r=XML&t=Fight%20Club
 - www.omdbapi.com/?apikey=7403b6e6&r=XML&s=Avengers
- www.omdbapi.com/?apikey=7403b6e6&r=XML&s=The%20Wolf%20of%20Wall
- Il est possible de changer le format des données en modifiant le paramètre **r** dans l'URI : XML->JSON

2. Wikipédia

- En modifiant le paramètre **srsearch** de la requête, l'attribut **totalhits** de la balise **searchinfo** change, il s'agit du nombre de résultat trouvé pour la valeur du paramètre **srsearch**
- Il existe 8 options de format : json, jsonfm, none, php, phpfm, rawfm, wml, xmlfm
 - **fm** est une option de représentation en HTML
 - fr.wikipedia.org/w/api.php?format=xmlfm&action=query&list=search&srsearch=iut%20charlemagne
- fr.wikipedia.org/w/api.php?format=json&action=query&list=search&srsearch=html5
 - fr.wikipedia.org/w/api.php?format=phpfm&action=query&list=search&srsearch=interopérabilité

D'autres exemples ?

- Curiosity
 - api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?sol=1000&camera=fhaz&api_key=kdllh4yvDwNdv8ag5AYZpCrIYWU8dfU4V1fACMc0
 - api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?sol=1200&camera=mast&api_key=kdllh4yvDwNdv8ag5AYZpCrIYWU8dfU4V1fACMc0
 - api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?

sol=2000&camera=rhaz&api_key=kdllh4yvdWnDv8ag5AYZpCrIYWU8dfU4V1fACMc0

- `api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?`
`sol=2000&camera=chemcam&api_key=kdllh4yvdWnDv8ag5AYZpCrIYWU8dfU4V1fACMc0`

- `api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?`
`sol=2000&camera=mahli&api_key=kdllh4yvdWnDv8ag5AYZpCrIYWU8dfU4V1fACMc0`

- `api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?`
`sol=1100&camera=mardi&api_key=kdllh4yvdWnDv8ag5AYZpCrIYWU8dfU4V1fACMc0`

- `api.nasa.gov/mars-photos/api/v1/rovers/curiosity/photos?`
`sol=100&camera=navcam&api_key=kdllh4yvdWnDv8ag5AYZpCrIYWU8dfU4V1fACMc0`

- Caméras de Opportunity : FHAZ, RHAZ, NAVCAM, PANCAM, MINITES
- Caméras de Spirit: FHAZ, RHAZ, NAVCAM, PANCAM, MINITES

Ces formats peuvent aussi permettre des échanges

1. Sur Google Maps, lors de l'exportation d'une carte précédemment créée, on retrouve 1 format de fichier : **KML** (**KMZ** est la version compressé de **KML**). En étudiant ce fichier, on retrouve une structure de données sous forme de balise

Mais ce n'est qu'un début

Les sites de ce type apportent une mise à jour régulière des données, et donc facilite l'utilisation de celle-ci dans des projets concernant une ville de développement. Les open data permettent d'utiliser des ressources déjà existante et prête à l'emploi, ce qui facilite son utilisation et l'intégration dans une application.

Des données disponibles mais aux constructions parfois opaques...

Url documentation : <https://adresse.data.gouv.fr/api-doc/adresse>

Préoccupation contemporaines, la nouvelle requête :

https://services3.arcgis.com/Is0UwT37raQYl9Jj/arcgis/rest/services/ind_grandest/FeatureServer/0/query?where=lib_zone%3D%27Ligny-en-Barrois%27&objectIds=&time=&geometry=&geometryType=esriGeometryEnvelope&inSR=&spatialRel=esriSpatialRelIntersects&resultType=none&distance=0.0&units=esriSRUnit_Meter&returnGeodetic=false&outFields=*&returnGeometry=true&featureEncoding=esriDefault&multiPatchOption=xyFootprint&maxAllowableOffset=&geometryPrecision=&outSR=&datumTransformation=&applyVCSProjection=false&returnIdsOnly=false&returnUniqueIdsOnly=false&returnCountOnly=false&returnExtentOnly=false&returnQueryGeometry=false&returnDistinctValues=false&cacheHint=false&orderByFields=&groupByFieldsForStatistics=&outStatistics=&having=&resultOffset=&resultRecordCount=&returnZ=false&returnM=false&returnExceededLimitFeatures=true&quantizationParameters=&sqlFormat=none&f=html&token=

Changements observés : lieu indiqué modifié (Nancy -> Ligny-en-Barrois), ainsi que le format rendu (json -> html)

URL pour API Parcoursup :

https://data.enseignementsup-recherche.gouv.fr/api/records/1.0/search/?dataset=fr-esr-parcoursup&q=bts+systemes+numeriques+-+option+informatique+et+reseaux&facet=session&facet=contrat_etab&facet=cod_uai&facet=g_ea_lib_vx&facet=dep_lib&facet=region_etab_af

Séance 2

Interlude... où en sommes-nous ?

Données ouvertes : des données brutes disponibles

La principale motivation de l'open data est la transparence des données accessibles à tout le monde. L'une des motivations est d'également permettre aux développeurs d'utiliser ses données pour les intégrer dans des applications et agrandir la visibilité des open-data

Données ouvertes, des histoires de

formats...

- En naviguant sur les différentes API, plusieurs formats sont proposés régulièrement (json, xml, csv), mais il existe d'autres formats moins répandus pour ce type de ressource : php, xmlfm & jsonfm (version HTML de XML ou JSON), KML (format XML pour du contenu géographique)
- Les données XML sont étiquetées dans des balises
- Les données JSON sont étiquetées sous forme de booléens, entiers, chaîne de caractères compris dans des objets ou de tableaux
- Les données CSV sont étiquetées sous forme de texte séparé par des virgules, qui forment un tableau traité par un logiciel tel que Excel

- Exemple de données en XML avec l'API OMBD :

- **URI** : www.omdbapi.com/?apikey=7403b6e6&r=XML&t=Fight%20Club
- **Paramètres** : **r** = format de données, **t** = titre exact du film
- **Autres paramètres possible** : **i** = id IMBD du film, **s** = chaîne de caractères comprise dans le titre, **y** = année de sortie, ...
- **Réponse** :

```
<root response="True">
  <movie title="Fight Club" year="1999" rated="R" released="15 Oct 1999"
runtime="139 min" genre="Drama" director="David Fincher" writer="Chuck
Palahniuk, Jim Uhls" actors="Brad Pitt, Edward Norton, Meat Loaf" plot="An
insomniac office worker and a devil-may-care soap maker form an underground fight
club that evolves into much more." language="English" country="Germany, United
States" awards="Nominated for 1 Oscar. 11 wins & 38 nominations total"
poster="https://m.media-amazon.com/images/M/MV5BMmEzNTkxYjQtZTc0MC00
YTVjLTg5ZTEtZWwOWVIYzY0NWlwXkEyXkFqcGdeQXVyNzkwMjQ5NzZM
@._V1_SX300.jpg" metascore="66" imdbRating="8.8" imdbVotes="1,958,640"
imdbID="tt0137523" type="movie" />
</root>
```

Et l'interopérabilité dans tout ça ?

Les données récupérées peuvent avoir plusieurs utilisations, comme pour les données de suivi de l'épidémie Covid-19, trié selon l'hôpital ou le département. Les données proposés par l'API OMBD peuvent servir à obtenir des informations sur nos films préférés.