

L'AMITIÉ AU PREMIER REGARD.

DOCTORAT SAUVAGE EN NUMÉRIQUE

Échelle . Ventres
inconnue . mous

GOOGLE EST MON AMI



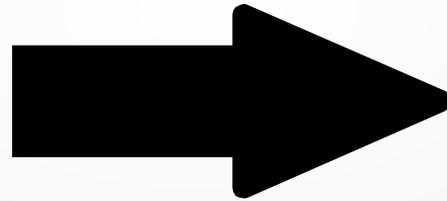
TOUS LES JOURS CHEZ VOUS

Mais aussi au boulot, en vacances, pendant les courses...

Google est mon ami

- Du Minitel à Google
balbutiements de la recherche
- Les débuts de Google
premiers défis
- Google aujourd'hui
Monstre et compagnie
- Les services Google
Nous sommes les frères qui rap'tout
- Dans un futur proche

Du Minitel à Google



Rechercher au temps du Minitel

- Annuaire papier Minitel
- Magazines et hors-séries
- Le Kiosque 3615
- Outils adaptés
 - Minitel pas adapté à une navigation automatisée
 - pas d'hyperlien
 - nombre de services limités
 - indexation → temps → coût

Youpi, le web arrive!

- **1989** création du web
Créé par Tim Berners-Lee au Cern
- **1993** premier web crawler (MIT)
Son créateur, Matthew Gray, travaille depuis plus de 10 ans chez Google
- **1994** Yahoo!
Répertoire de sites catalogués humainement

YAHOO!



La recherche devient essentielle

- **1994 Lycos**

Un des premiers moteurs de recherche

Plus gros index en 1996 avec 60 millions de documents

- **1995 AltaVista**

Démonstration des ordinateurs Dec Alpha 8400 TurboLaser et de leur SGBD

20 serveurs pour un total de 2,5 To RAM, 10 To disque dur

Des défis à relever 1/2

- **Quantité énorme de documents**
à parcourir, à indexer, à stocker, à rechercher, à rafraîchir...
- **Complexité des documents**
Langues, jeux de caractères, homonymes, fautes d'orthographe, formats
- **Pertinence des résultats**
- **Évolution des algorithmes et des normes**

Des défis à relever 2/2

- Pannes et obsolescence du matériel
- Tricheurs, sécurité
- Consommation électrique
- Connexion Internet
- Mais surtout... comment gagner de l'argent?

Les débuts de Google



La rencontre

- **1995** Larry Page et Sergueï Brin se rencontrent
Université de Stanford dans la Silicon Valley
- **1996** ils collaborent sur 3 projets
 - robot d'indexation
 - algorithme PageRank
 - moteur de recherche BackRub

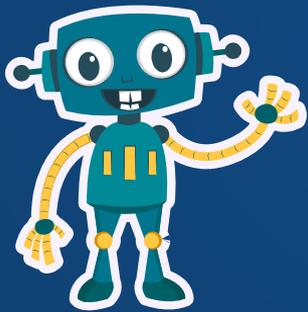


Le lancement

- **1997** google.com est réservé
- **1998** installation dans un garage
- **1999** décollage
 - **février**: 500 000 recherches/jour
 - **mars**: déménagement à Palo Alto
 - **juin**: 25 000 000 \$ de capitaux propre
 - **août**: 3 000 000 recherches/jour
 - **septembre**: déménagement à Mountain View



À l'époque



	A	B	C
hibou	×	×	
joujou			×
caillou		×	×
genou		×	
chou	×		×

méta-données
date de collecte
langue
dernière mise à jour
format
en-têtes du serveur
etc.

Indexation et pertinence

- **Prise en compte du contexte**
ex.: « pied de biche » n'a aucun rapport avec l'animal
- **Importance sémantique des mots**
ex.: le, la, les ont peu d'importance
- **Glissement sémantique**
ex.: évolution du mot « typographie »
- **Fautes d'orthographe**
- **Actualité ou information**

PageRank

- Calcule la popularité d'une page
- Une page gagne en popularité si...
 - elle est pointée par une page populaire
 - la page la pointant n'a pas beaucoup d'autres liens
- Google a la plus grande popularité possible
- Un facteur parmi d'autres

Mais...

- La pertinence peut encore être améliorée
- Cela ne règle pas la question du modèle économique

Google aujourd'hui



Algorithmes, algorithmes

- Analyse fine des pages indexées
- Personnalisation des résultats
- Évolutions régulières

Un modèle économique

- Pour Google
 - ses clients: des annonceurs
 - leur produit: vous!
- Besoin de collecter un maximum de données
 - pour proposer le meilleur ciblage aux annonceurs
 - pour vous fournir le meilleur service
- Maximum de données → maximum de traitements

15 datacenters



Quelques chiffres

- entre 900 000 et 2 000 000 de serveurs (estimations)
- 1 datacenter = plusieurs bâtiments
- 1 bâtiment = 75 000 serveurs
100 To/s de bande passante
- 2,6 giga-watts d'énergie renouvelable
Google consomme notamment la totalité des 114 méga-watts produits par une ferme d'éoliennes dans l'Iowa depuis 2010
- Google reste discret sur ses chiffres

<https://www.youtube.com/watch?v=zDAYZU4A3w0>

<http://www.datacenterknowledge.com/archives/2017/03/16/google-data-center-faq>

Les services Google

Les services Google

- **Gratuité**
- **Simplicité**
Utilisation, installation, mise en œuvre...
- **Performances**
Temps de réponse, bande passante, volumes...
- **Disponibilité**
- **Facilement insérables dans une page web**

Les données collectées

- **Ce que vous faites**
recherches, sites web consultés, vidéos visionnées, annonces visitées, situation géographique, matériel, adresse IP, cookies
- **Ce que vous créez**
e-mail via Gmail, contacts, agenda, photos, vidéos, Google Docs/Sheets/Slides, Drive
- **Ce qui vous rend unique**
nom, adresse mail, mot de passe, date de naissance, sexe, numéro de téléphone, pays

The image features a light gray background with decorative blue geometric shapes in the corners. These shapes consist of several parallel lines of varying lengths, creating a sense of depth and movement. The lines are arranged in a way that suggests a 3D effect, with some lines appearing to recede into the distance.

Le Google visible

Google Search

- Moteur de recherche le plus rapide/pertinent
 - 30 000 milliards de pages indexées
 - 20 milliards de sites visités/jour
 - 3,3 milliards de requêtes/jour
- Monopole des recherches (92%)
- Moteur de recherche par défaut
Firefox, Chrome



Android

- Système d'exploitation pour mobiles et tablettes
- Monopole mondial (85%)
- Connexion maximale avec les services Google agenda, contacts, applications, drive, vidéo, géolocalisation...
- Google avec vous, tout le temps!



Gmail

- Messagerie électronique
- Plusieurs giga-octets d'espace
- Anti-spam performant
- Recherche made in Google
- Toutes vos correspondances lues par Google!
projets d'achat ou d'investissement, mots de passe,
mots plus ou moins doux, photos intimes...



Google Maps

- Service de cartographie
- Rapidité, Street View, info trafic
- Où vous êtes, ce que vous recherchez



Google Chrome

- **Navigateur web**
- **Mode non connecté**
préférences linguistiques, assistance à la navigation, statistiques d'utilisation
- **Mode connecté**
Synchronisation avec son compte Google (sites visités, favoris, mots de passe...)



YouTube

- Plateforme de partage de vidéos
- Les chiffres:
 - 30 millions d'utilisateurs/jour
 - 5 milliards de vidéos visionnées/jour dont 1 milliard sur mobile
 - 300 heures de vidéo ajoutées/minute
- Ce que vous aimez, détestez, partagez



Google Docs/Sheets/Slides

- Suite bureautique en ligne
- Édition simultanée de documents
- Documents soumis à la discrétion de Google...



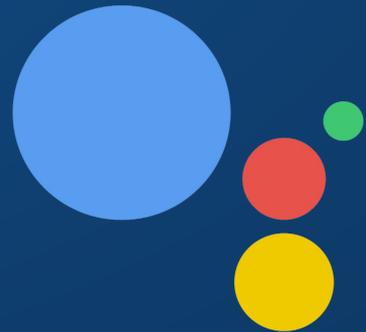
Google Forms

- Formulaires en ligne
- Collecte de données diverses



Google Assistant

- Assistant vocal bidirectionnel « Ok Google! »
- Rendre instinctive l'utilisation de Google
- Écoute en permanence



Google Play

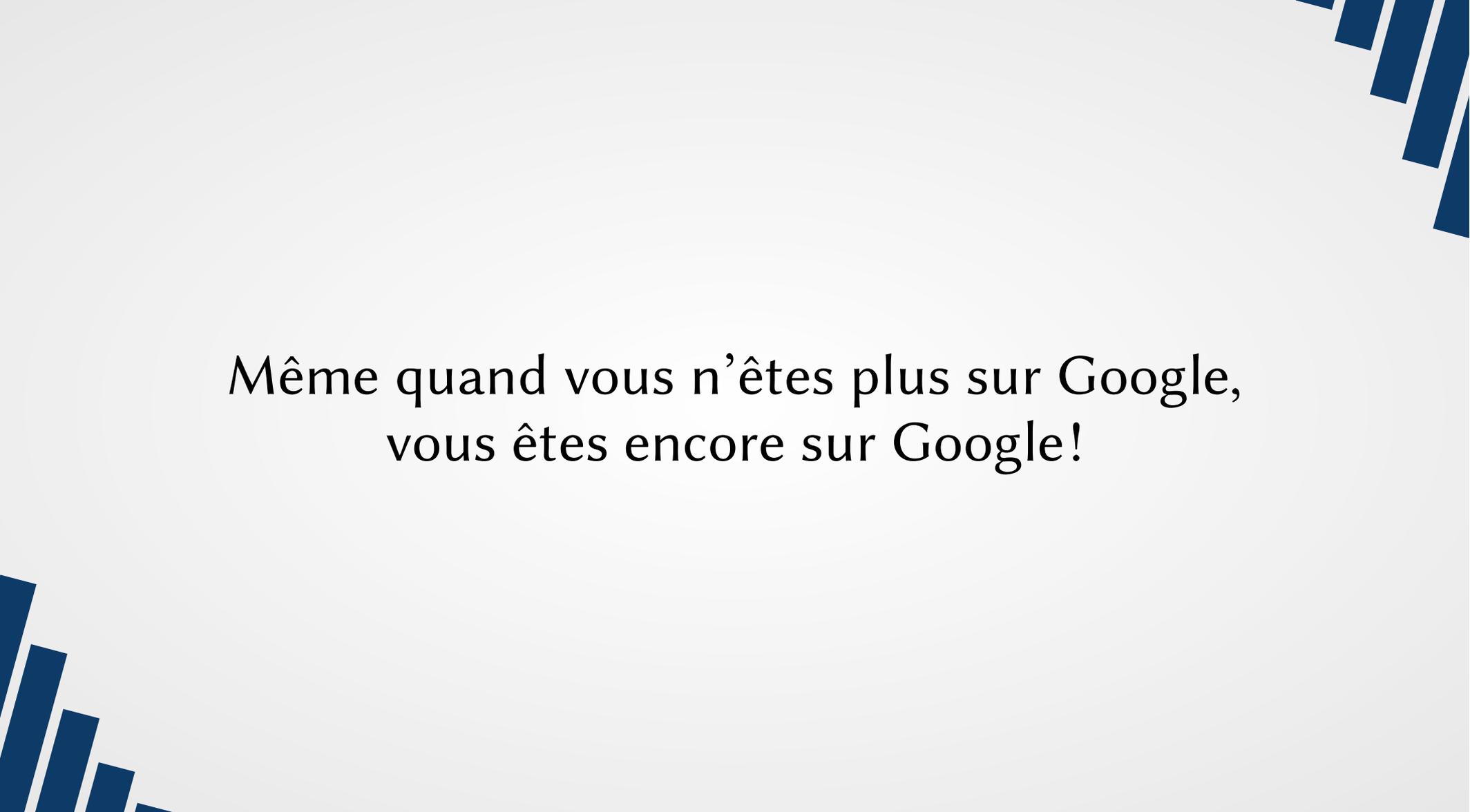
- Magasin d'applications Android
- Google sait quelles applications vous installez



Google Plus

- Réseau social de Google
- Se ~~voulait~~ veut concurrent de Facebook
Facebook limite fortement l'intrusion de Google
- Google ne peut pas toujours réussir



The image features a light gray background with decorative dark blue geometric shapes in the corners. These shapes consist of several parallel, slightly curved lines that create a sense of depth and movement. The text is centered in the middle of the page.

Même quand vous n'êtes plus sur Google,
vous êtes encore sur Google!

Google Analytics

- Outil de statistiques pour site web
- Monopole en France (70% en 2014)
elysee.fr, sncf.com, metropole-rouen-normandie.fr, reseau-astuce.fr, onvasortir.com, rouentourisme.com, musees-rouen-normandie.fr...
- Installé par de nombreux prestataires
- Son omniprésence permet à Google de suivre les internautes à la trace

Hosted libraries

- Scripts utilisables par les développeurs
- Google se charge de l'hébergement
- Nombreux scripts disponibles

AngularJS, Angular Material, Angular UI Router, Dojo, Ext Core, Hammer JS, Indefinite Observable, jQuery, jQuery Mobile, jQuery UI, MooTools, Prototype, script.aculo.us, Shaka Player, SPF, SWFObject, three.js, Web Font Loader

Google AMP

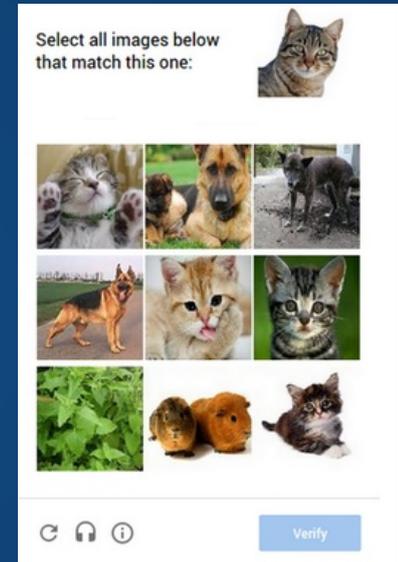
- Pages Mobiles Accélérées
- Accélère le chargement des pages
Document simplifié, mise en cache
- Pages mises en avant par Google
- Pages servies sur les serveurs de Google



<https://80x24.net/post/the-problem-with-amp/>
https://www.ampproject.org/docs/guides/how_cached

reCAPTCHA

- Filtrage des robots
- Analyse votre navigation
- Vous travaillez gratuitement pour Google!
en entraînant leur intelligence artificielle



Google Fonts

- Fontes gratuites pour le web
- Hébergées par Google...

Google Fonts

Contenus embarqués

- Insertion de contenus en provenance de Google
- Porte ouverte à Google



The image features a light gray background with decorative blue geometric shapes in the corners. In the top-right corner, there are several parallel blue lines of varying lengths, creating a fan-like pattern. In the bottom-left corner, there are several parallel blue lines of varying lengths, creating a similar fan-like pattern.

Et encore...

Vous avez échappé à ça

- Google Public DNS
- Google Trends
- Google AdSense
- Google AdWords
- Google APIs
- Google ChromeCast
- Blogger
- Google Wifi
- Google Payments API
- Android Auto
- Pixel
- Google Photo
- Google Local Guide
- Google Earth
- Google Waze
- Google SignIn
- Google Books
- Google Education
- Google Scholar
- ...



Here comes the GDPR!
Et voici le RGPD!

40 ans

- **1978** loi Informatique et Liberté → Cnil
- **1995** directive 95/46/CE
- **2009** Paquet Télécom
- **2016** loi pour une République Numérique
- **2018** entrée en vigueur du RGPD le 25 mai

RGPD

- **R**èglement **G**énéral de **P**rotection des **D**onnées
- Règlement européen relatif à
 - la protection des personnes physiques
 - à l'égard du **traitement des données à caractères personnel**
 - et à leur libre circulation

Traitement

- **Toute opération**
collecte, enregistrement, organisation, structuration, conservation, adaptation, modification, extraction, consultation, utilisation, communication, diffusion, mise à disposition, rapprochement, interconnexion, limitation, effacement, destruction
- **Manuelle ou automatisée**

Données à caractère personnel

- Toute information...
quel que soit le support (électronique, papier...)
- ... se rapportant à une personne physique
identifiée ou identifiable
- **Dès qu'une quantité d'informations permet
d'identifier une personne, ce sont des données
à caractère personnel!**

Principes

- **Consentement**
obligatoire, explicite, facile à comprendre
- **Finalités déterminés**
précises, explicites, légitimes
- **Durée de conservation**
limitée, raisonnable
- **Minimisation**
seules les données nécessaires au traitement sont collectées
- **Protection**
par défaut et par design, technique et organisationnelle

The image features a light gray background with decorative blue geometric shapes in the corners. These shapes consist of several parallel lines of varying lengths, creating a stylized, modern look. The text is centered on the page.

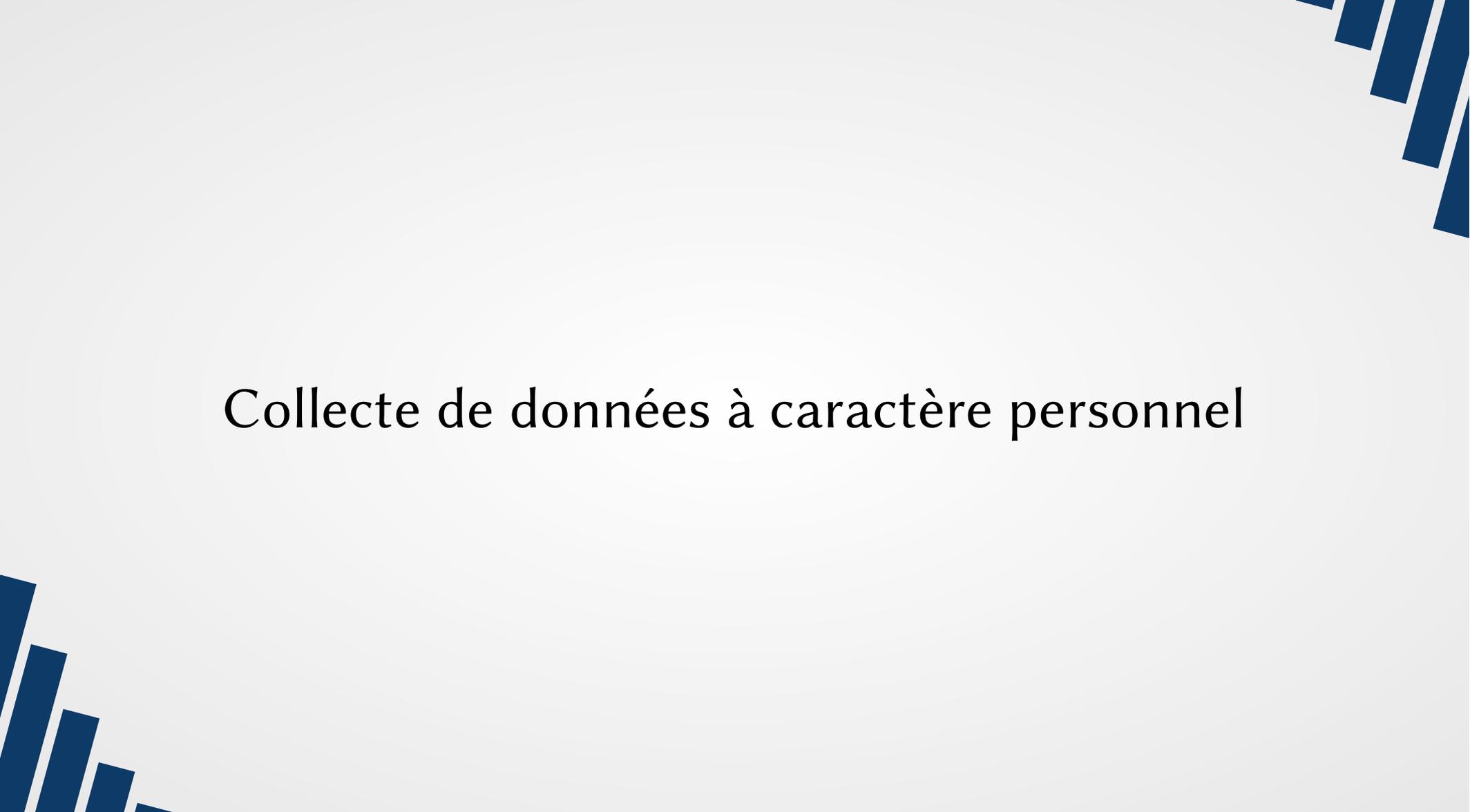
Droits et obligations

Droits des personnes concernées

- Droit à l'information
- Droit d'accès
- Droit de rectification
- Droit à l'effacement
- Droit à la limitation du traitement
- Droit à la portabilité
- Droit d'opposition

Obligations du responsable du traitement

- Délégué à la Protection des Données (DPO)
successeur du correspondant Informatique et Liberté
- Preuve que les traitements respectent le RGPD
- Maintenance du registre
description des traitements, des finalités, des données conservées,
de la durée de conservation...
- Étude d'impact
en cas de risque important
- Certification et code de conduite

The image features a light gray background with decorative dark blue geometric shapes in the corners. These shapes consist of several parallel lines of varying lengths, creating a stylized, abstract pattern. The text is centered in the middle of the page.

Collecte de données à caractère personnel

Concrètement

- Le responsable vous informe du traitement dont vous allez être l'objet
- Il doit fournir les informations
 - d'une façon concise,
 - transparente,
 - compréhensible,
 - aisément accessible,
 - en des termes clairs et simples

Informations à fournir 1/2

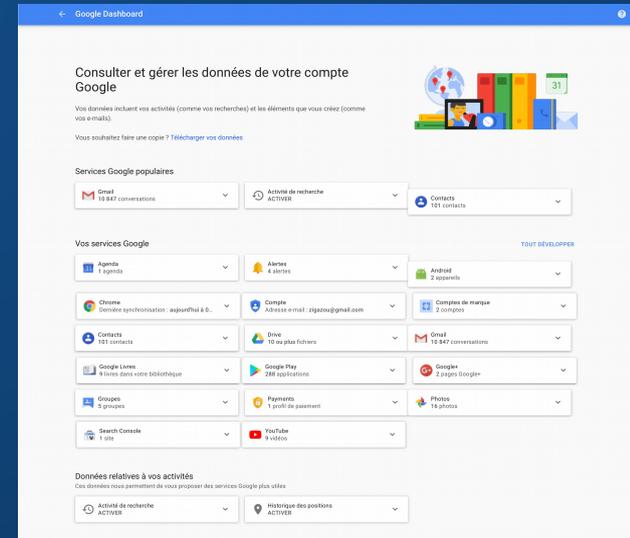
- Identité, coordonnées du responsable
- Finalités et base juridique du traitement
le cas échéant, les intérêts légitimes poursuivis
- Destinataires
- Durée de conservation
- Droit d'accès à vos données
ainsi que la rectification, l'effacement ou la limitation du traitement

Informations à fournir 2/2

- Droit de retirer son consentement
- Droit d'introduire une réclamation
- Caractère de la collecte
règlementaire ou contractuel? Conditionne-t-il la conclusion d'un contrat?
Quelles sont les conséquences en cas de non-fourniture des données?
- Existence d'une prise de décision automatisée
Y compris un profilage (ex.: Google Analytics)
- Les autres finalités

Google Dashboard

- Google a pris de l'avance avec son Dashboard
- <https://myaccount.google.com/dashboard>



Merci d'avoir tenu jusque là!

Quelques liens 1

- Données collectées par les apps – Exodus
<https://reports.exodus-privacy.eu.org/reports/apps/>
- Dégooglisons Internet – Framasoft
<https://degooglisons-internet.org/>
- Voir l’empreinte de son navigateur – Am I unique
<https://amiunique.org/>
- Qui surveille votre navigation? – Greffon Ghostery
<https://addons.mozilla.org/fr/firefox/addon/ghostery/>

Quelques liens 2

- Rien à cacher? – Nothing to hide
<https://nothingtohidedoc.wordpress.com/nothing-to-hide-french-high-quality/>
- Faire une demande Cada – Xavier Berne
<https://www.nextinpact.com/news/105587-comment-faire-demande-cada-daces-aux-documents-administratifs.htm>
- Google utilise un mouchard pour géolocaliser les utilisateurs Android – Anaïs Cherif
<http://www.latribune.fr/technos-medias/google-utilise-un-mouchard-pour-geolocaliser-les-utilisateurs-android-759095.html>
- Le RGPD – EUR-Lex
<http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679>

Quelques liens 3

- **Pourquoi vous ne devriez pas utiliser Google Chrome**
<http://www.revoltenumerique.herbesfolles.org/2014/04/23/pourquoi-vous-ne-devriez-pas-utiliser-google-chrome/>
- **5 tests faciles pour voir tout ce que Google sait sur vous**
<http://www.archimag.com/vie-numerique/2015/11/10/5-tests-faciles-verifier-google-sait-vous-donnees-personnelles>
- **Liste des 178 acquisitions de Google – Olivier Duffez**
<http://www.webrankinfo.com/dossiers/google/acquisitions>
- **Chiffres-clés de Google 2017 – Olivier Duffez**
<http://www.webrankinfo.com/dossiers/google/chiffres-cles>

Quelques liens 4

- Moteur de recherche Qwant
<https://www.qwant.com/>
- Moteur de recherche DuckDuckGo
<https://duckduckgo.com/>