

Module 7 : Conquérir et/ou défendre une position stratégique

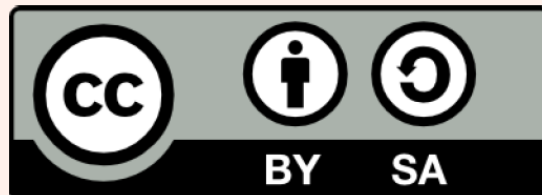
**Modèle économique et
métiers de la donnée**

Dataactivist, 2021-2022

Ces slides en ligne : http://dataactivist.coop/SPoSGL/5A_module7.html

Sources : <https://github.com/dataactivist/SPoSGL/>

Les productions de Dataactivist sont librement réutilisables selon les termes de la licence [Creative Commons 4.0 BY-SA](#).



7. Conquérir et/ou défendre une position stratégique

7.1 Constituer un patrimoine de données

7.2 Cartographier ses données

7.3 Construire un data pipeline performant

Etude de cas : Netflix, de la distribution à la production de contenu

Conquérir et/ou défendre une position stratégique

Les données sont utilisées comme un moyen pour analyser les opportunités et y répondre. À ce titre, elles constituent un actif stratégique de la structure qui doit en prendre soin.

7.1 Constituer un patrimoine de données

Conquérir ou défendre une position stratégique avec les données nécessite dans un premier temps de constituer un véritable **patrimoine de données** pour analyser les opportunités.

Les données dans les entreprises représentent une valeur considérable mais encore trop souvent méconnue. Certaines structures n'ont pas une connaissance exhaustive de toutes les données qu'elles produisent. Pour cela, elles peuvent réaliser une **cartographie des données**.

7.2 Cartographier ses données

Cartographier des données consiste à recenser toutes les données produites par une structure avec :

- leur provenance
- leur point d'entrée logiciel
- les éléments qu'elles contiennent
- leur fréquence de production et de mise à jour
- leur qualité et les différents contrôles auxquelles elles sont soumises
- leurs traitements
- leurs usages
- ...

□	Ensemble de données	Verticale métier	Exemples de jeux de données	Tags	Organisations compét...	Source des données	Standards de données existants	Standards
1	Référentiel des élus locaux	Administration générale		Elus Démocratie	Commune Conseil Départe	système d'informati...	standard(s) de jure	schema.gouv
2	Gestion et suivi des fonds européens (FEDER, FSE, ...)	Administration générale			Conseil Régional		aucun standard, donnée non structurée	
3	Données des assemblées (délibérations, votes)	Administration générale		Elus Démocratie Vote		saisie manuelle (typ...	standard(s) de facto	SCDL schema.gouv
4	Conseils consultatifs et stratégiques (CESER, CRJ, ...)	Administration générale			Conseil Régional	saisie manuelle (typ...	aucun standard, donnée non structurée	
5	Gestion du domaine public, maritime, fluvial, aéro...	Administration générale			Commune EPCI Conseil R	système d'informati...	aucun standard, donnée non structurée	
6	Répertoire des informations publiques	Administration générale			Commune EPCI Conseil D	saisie manuelle (typ...	aucun standard, donnée non structurée	
7	Contractualisation territoriale (Contrat de plan Eta...	Administration générale Aménagement du t					aucun standard, donnée non structurée	
8	Lieux et heures d'ouverture des services publics e...	Administration générale Relation Usagers		Horaires Ouverture	Commune EPCI Conseil D	système d'informati...	standards existants mais utilisation non homogé...	Opening Hours - OSM
9	Absentéisme des élus, rémunération des élus	Administration générale RH	Présences et absences des élus de l	Elus Démocratie Absen	Commune Conseil Départe	système d'informati...	aucun standard, donnée non structurée	
10	Description des opérations	Aménagement du territoire	Localisation des opérations Calend				aucun standard, donnée non structurée	
11	Enquêtes publiques	Aménagement du territoire	Résultats des enquêtes publiques			crowdsourcing	aucun standard, donnée non structurée	
12	PLU/PLUIH	Aménagement du territoire	Secteur à pourcentage de logemen	SIG	EPCI	système d'informati...	standard(s) de jure	

Cartographie des données des collectivités territoriales réalisée par Dataactivist

7.3 Le data pipeline

Un **data pipeline** fait référence aux étapes du **cycle de vie de la donnée** au sein d'une structure :

- **la collecte** : enregistrer et clôturer les données sur une période donnée
- **la sécurisation** : pour éviter les risque de modification ou de suppression
- **l'archivage** : figer et dater avec un dispositif technique garantissant l'intégrité des informations. Sans cela il sera impossible d'analyser les données sur la durée.
- **le partage** : au sein des différents services de la structure et vers l'extérieur
- **l'exploitation** : pour valoriser les données

Une bonne maîtrise de chaque étape du pipeline est indispensable à une réelle valorisation des données d'une structure. Si une étape est négligée, il sera difficile, voir impossible de construire une stratégie de développement basée sur les données.

Etude de cas : Netflix, de la distribution à la production de contenu

Produire les données de sa réussite

Absolument tout ce qui se passe côté utilisateur est loggé, stocké et utilisé pour personnaliser au maximum l'expérience. Les données sur lesquelles s'appuient les services de Netflix sont d'une grande richesse quantitative mais aussi qualitative.

- notations des contenus par les abonnés
- popularité des contenus (évaluée de multiples manières)
- données de lecture des vidéos (durée, heure, matériel...)
- listes d'attente
- interactions des abonnés avec la présentation d'un contenu
- termes de recherche ...

To properly learn how to personalize artwork we need to collect a lot of data to find signals that indicate when one piece of artwork is significantly better for a member. (source)

Un algorithme de recommandation

La fonction principale de la page d'accueil est d'aider chaque membre à trouver facilement quelque chose à regarder. Le défi consiste à **sélectionner** les contenus les plus pertinents pour l'abonné et de les **organiser** tout aussi pertinemment sur la page d'accueil.

L'algorithme de Netflix compose la page d'accueil de chaque abonné :

- il sélectionne les contenus adaptés parmi des milliers de vidéos
- il sélectionne les groupes thématiques adaptés à afficher (il existe 10000 groupes "rows" possibles)
- il organise l'affichage des groupes thématiques pour optimiser leur agencement dans la page
- il organise les vidéos dans un ordre pertinent au sein des groupes



Faire des choix

Chaque algorithme comporte un **scoring** qui lui permet de faire des choix. Celui de Netflix doit trouver le juste équilibre entre :

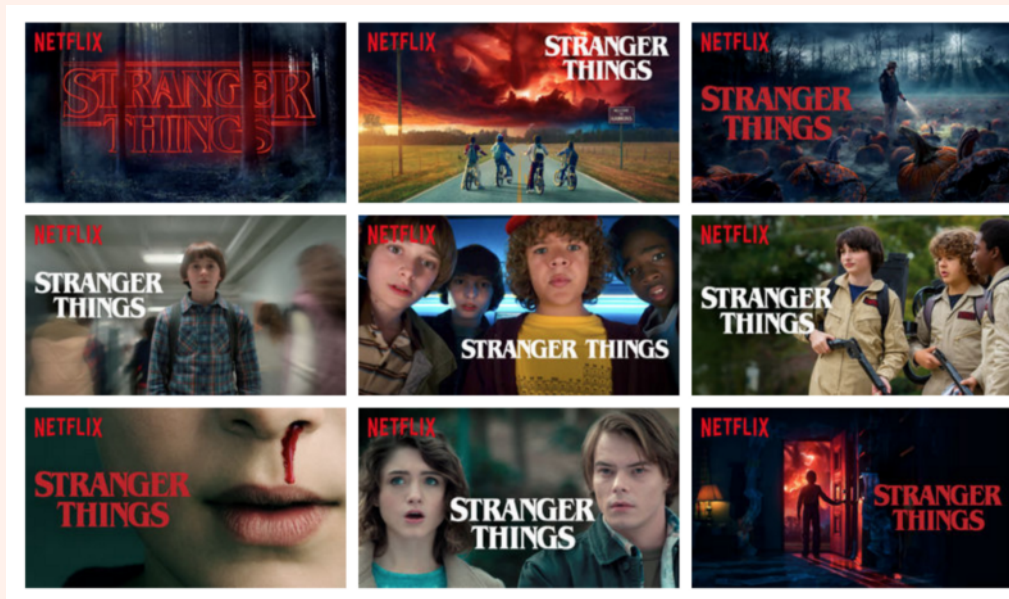
- répondre aux goûts personnels de chaque abonné et veillez à couvrir tous ses centres d'intérêts
- aider chaque abonné à trouver facilement le contenu qui va lui plaire mais aussi l'aider à explorer la diversité du catalogue et peut-être trouver de nouveaux centres d'intérêt
- montrer les nouveautés et la richesse du catalogue tout en maintenant une certaine continuité afin que l'abonné garde des points de repère et puisse retrouver un contenu qu'il se souvient avoir vu sur sa page d'accueil.

Article

Pour en savoir plus : Blog, Netflix Technology. « [Learning a Personalized Homepage](#) ». Medium, 19 avril 2017.

Une personnalisation très fine

Le saviez-vous ? Netflix va jusqu'à personnaliser les visuels qu'il présente à ses abonnés ! Si auparavant la firme recherchait le meilleur visuel d'illustration pour tous ses membres, elle adapte maintenant le visuel à chaque abonné en fonction de ses goûts.



Les illustrations personnalisées de Stranger Things couvrent toutes les facettes de la série (fantastique, vintage, angoisse, amitié, adolescence...) pour attirer un public plus varié vers ce programme.

Une personnalisation très fine

Chloé, abonnée Netflix : "J'avais remarqué que sur le profil de mon copain, le visuel pour Lucifer était une image de combat alors que moi j'avais la version Lucifer torse nu ! Je rigole en imaginant comment l'algorithme m'a catégorisée !"

Cette personnalisation peut également se faire en fonction des acteurs et actrices que les abonnés apprécient d'après leurs données :



Article

Pour en savoir plus : Blog, Netflix Technology. « [Artwork Personalization at Netflix](#) ». Medium, 7 décembre 2017.

Une production de contenus basée sur les données

Dans une industrie historiquement influencée par l'intuition, Netflix revendique de faire du **"data driven programming"**.

La position de distributeur donne ainsi un avantage concurrentiel unique à Netflix en tant que producteur : Netflix se base sur ses données d'activité de distribution pour prendre des décisions dans ses activités de production.

Netflix à Kevin Spacey : "We believe in you. We've run our data and it tells us that our audience would watch this series. We don't need you to do a pilot. How many do you wanna do?" ([source](#))

(Mais le facteur humain reste significatif, même chez Netflix, d'après [ce meme article](#))

Video

Merci !

Contact : julia@dataactivist.coop