

OPEN DATA

Etat de l'art et perspectives



GISday
2018

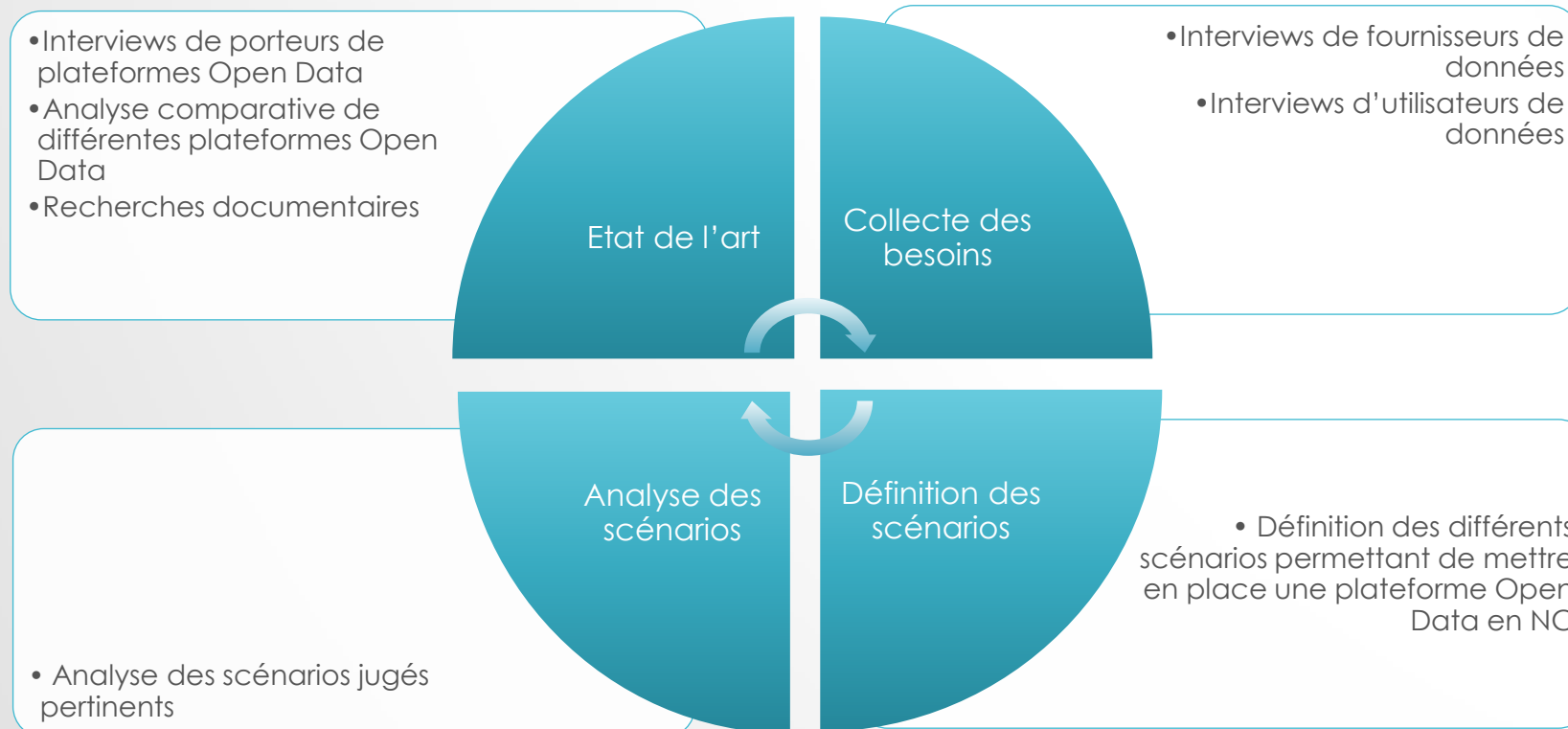


SOMMAIRE

- ***Méthodologie***
- Contexte et enjeux
- Etat de l'art hors Nouvelle-Calédonie
- Etat de l'art en Nouvelle-Calédonie
- Scénarii

ÉLÉMENTS MÉTHODOLOGIQUES

- Ces travaux ont été réalisés de mai à juillet 2018 :



SOMMAIRE

- Méthodologie
- **Contexte et enjeux**
- Etat de l'art hors Nouvelle-Calédonie
- Etat de l'art en Nouvelle-Calédonie
- Scénarii

QU'EST-CE QU'UNE DONNÉE OUVERTE (OPEN DATA) ?

Usages

Selon la philosophie « Open Data », une donnée est ouverte si elle ne souffre d'aucune restriction : tout le monde peut y accéder, la modifier, l'utiliser et la partager, même à des fins commerciales.

Donnée payante ou gratuite

Une donnée ouverte peut être payante :

- Le fait de produire, maintenir et publier une donnée a un coût.
- Ce coût peut parfois être élevé, par exemple dans le cas d'une donnée disponible en temps réel.

Ce point peut être sujet à discussion : les données étant produites par des organismes publics, on peut considérer qu'elles ont déjà été financées par le contribuable.

Donnée disponible dans un format libre

Certains formats de données rendent les données plus facilement utilisables. Cela ne joue pas sur le fait qu'une donnée soit ouverte ou non, mais augmente son utilisation.

Exemple : format libre (CSV) vs format sous licence (XLS).

QU'EST-CE QU'UNE PLATEFORME OPEN DATA ?

Définition

Plateforme de téléchargement permettant de récupérer des jeux de données brutes, sans autre filtre et avec en général l'acceptation d'une licence ou de conditions générales d'utilisation - *Source CNIL*.

Fonctionnalités minimums

Une plateforme Open Data regroupe à minima les fonctionnalités suivantes :

- Une administration peut mettre à disposition ses données (chaque donnée contient une licence qui décrit les conditions d'utilisation) ;
- Une administration peut mettre à jour ses données ;
- Un utilisateur (administration, société privée ou particulier) peut rechercher et/ou filtrer les données disponibles dans la plateforme ;
- Un utilisateur (administration, société privée ou particulier) peut accéder aux données, et les utiliser conformément à la licence associée.



Une obligation légale déjà présente

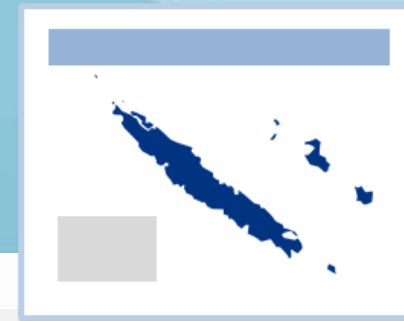
- Déclaration des droits de l'Homme et du Citoyen, article 15
- Code des Relations entre le Public et l'Administration (CRPA)
- Code des Communes

➔ Gouvernement, provinces, établissements publics et établissements chargés d'une mission de service public sont déjà dans l'obligation de communiquer les documents administratifs à la demande de l'utilisateur. Toutefois, ils n'ont pas l'obligation d'ouvrir leurs données publiques (à l'exception des communes).



Un mouvement qui se développe

- De plus en plus d'administrations proposent certaines données via leur site web (Province sud, Ville de Nouméa, etc.)
- Les acteurs locaux (privés et particuliers) commencent à demander l'ouverture des données de la Calédonie

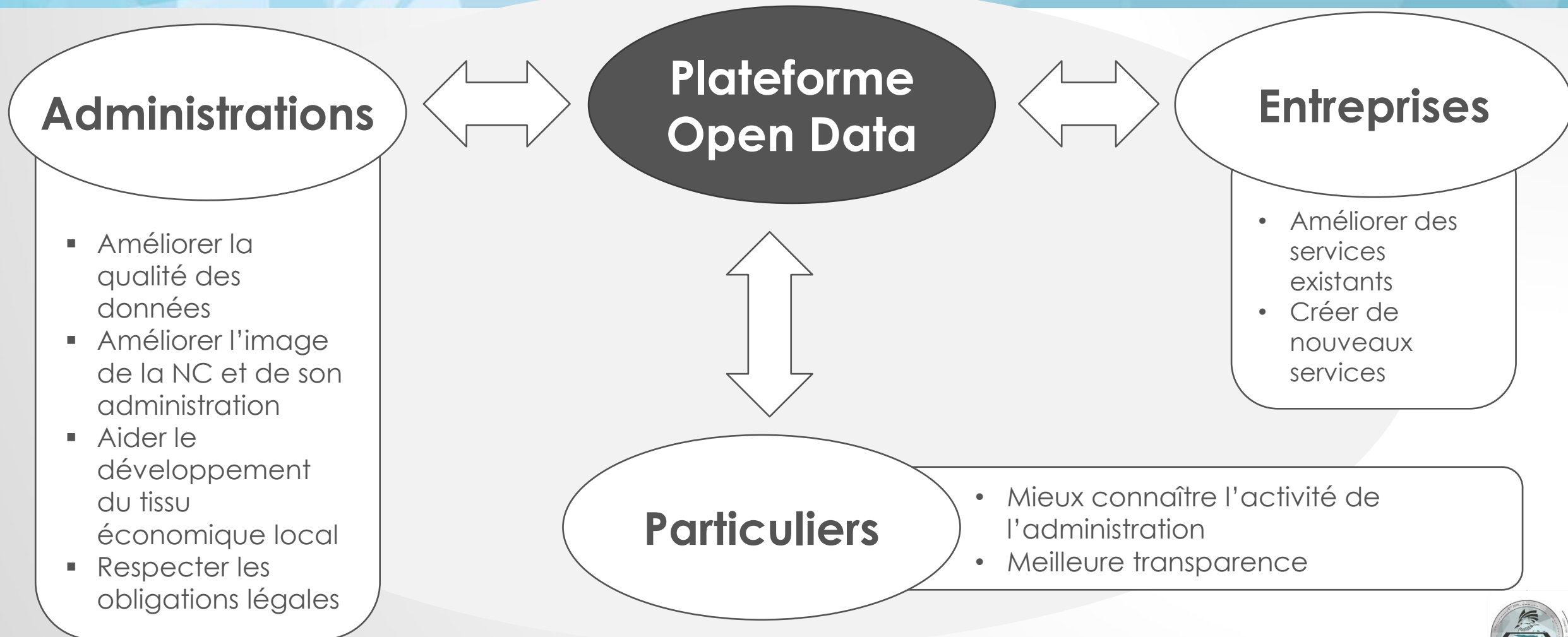


GISday

Au sein du gouvernement, une démarche initiée pour certaines données

- Les données géomatiques sont pour certaines déjà disponibles en Open Data, via le portail GEOREP, etc.
- Le gouvernement doit préparer l'ouverture d'autres types de données

ACTEURS CONCERNÉS ET ENJEUX



SOMMAIRE

- Méthodologie
- Contexte et enjeux
- ***Etat de l'art hors Nouvelle-Calédonie***
- Etat de l'art en Nouvelle-Calédonie
- Scénarii

GOUVERNANCE DE LA PLATEFORME

Financement et
gouvernance centralisée

Le **financement** est **centralisé**, ce qui va de pair avec une **gouvernance**, elle aussi **peu partagée**

Deux fonctions de gestion :
business et technique

Les équipes sont toujours composées d'une **fonction de Business Développement / animation internalisée** et d'une **fonction technique** (développeurs web/informatique) **internalisée ou externalisée**.

Des licences proposées sans
contraintes

Certains pays ont développé leur propre licence (France, Canada), mais **sans** jamais **contraindre** les partenaires fournisseurs de données à utiliser ces licences. Majoritairement, les licences sont ouvertes et gratuites par défaut.

Moissonnage de
métadonnées obligatoire.
Données optionnelles

La fonction de **moissonnage de données** est **obligatoire**. La gestion des **données** est **optionnelle**, mais utile pour les partenaires les moins dotés de moyens techniques et humaines.

Un cadre commun sans
contrainte

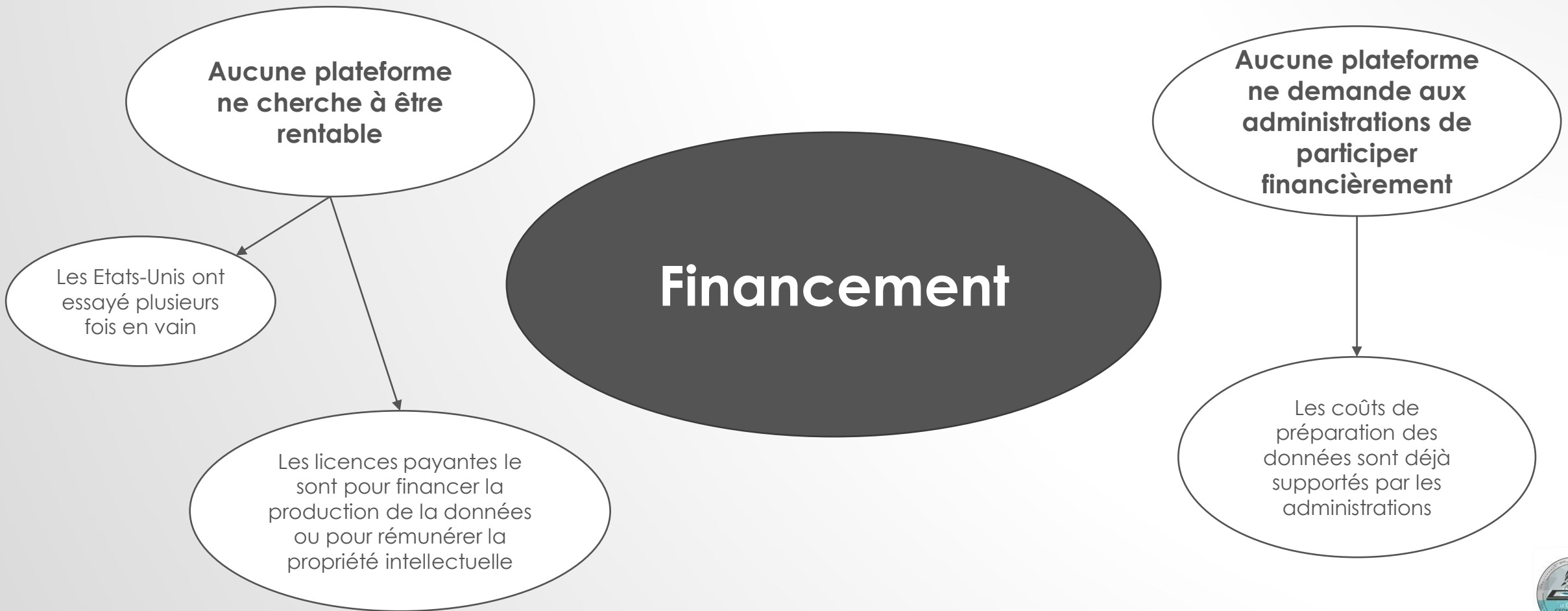
La publication d'un **cadre de travail commun** semble bénéfique (licence, format des données, format des métadonnées, etc.), mais toujours ... non contraignant.

Aider les « petits »

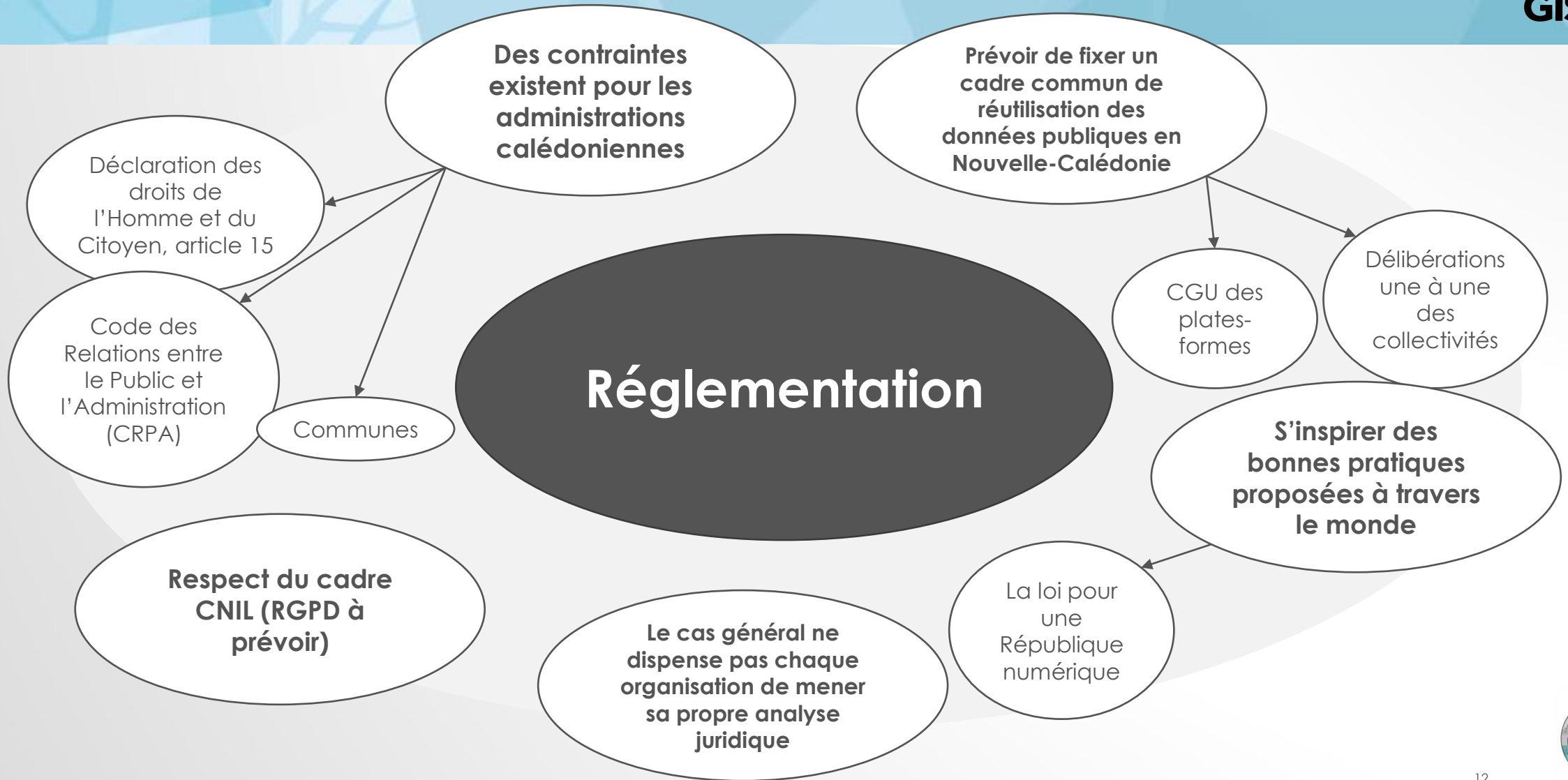
On constate que les pays qui n'**appuient** pas les « petits » contributeurs n'arrivent pas à les faire participer. Il serait donc utile de penser à un moyen de les aider (techniquement, financièrement et/ou humainement).



MODÈLE ÉCONOMIQUE DE LA PLATEFORME



RÉGLEMENTATION



PLATEFORMES EXISTANTES

	Etalab	CKAN	DKAN	
 Langage de programmation	Python	Python	PHP	OpenData Soft Junar Socrata Swirr ArcGis Online S A A S
 OS	Windows, Linux	Linux	Linux	
 Serveur d'application	Nginx, Apache, Tomcat	Nginx, Apache	Apache	
 CMS	Drupal, eZpublish	Drupal, Wordpress	Drupal	
 Base de données	MongoDB	PostgresSQL	MySQL	
		Existe une version Windows		

LICENCES D'UTILISATION DES DONNÉES

Restrictions

Par défaut (sauf licence particulière), les porteurs de plateformes étudiés ne mettent aucune restriction sur l'utilisation des données mises à disposition.

Les licences permettent de télécharger, exploiter, modifier et de faire un usage commercial des données.

Parenté

Toutes les licences imposent de mentionner la source des données.

Licences propres

Plusieurs plateformes ont créé leurs propres licences, pour des besoins très spécifiques, mais elles restent toutes compatibles avec la licence internationale **CC-BY**.

La création et l'utilisation d'une licence propre demande un effort supplémentaire de la part des utilisateurs pour vérifier ce qu'ils ont le droit de faire avec cette licence.

Licence payante

Toutes les plateformes utilisent des licences gratuites par défaut (Canada, Japon).

Quand une plateforme applique une licence payante sur un jeu de données, en général, l'utilisateur de données doit contacter le fournisseur de données pour définir une convention pour exploiter ses données. La donnée n'est pas directement accessible sur la plateforme, mais la métadonnée peut l'être.

CAS DU GRAND LYON : 3 LICENCES SONT PROPOSÉES

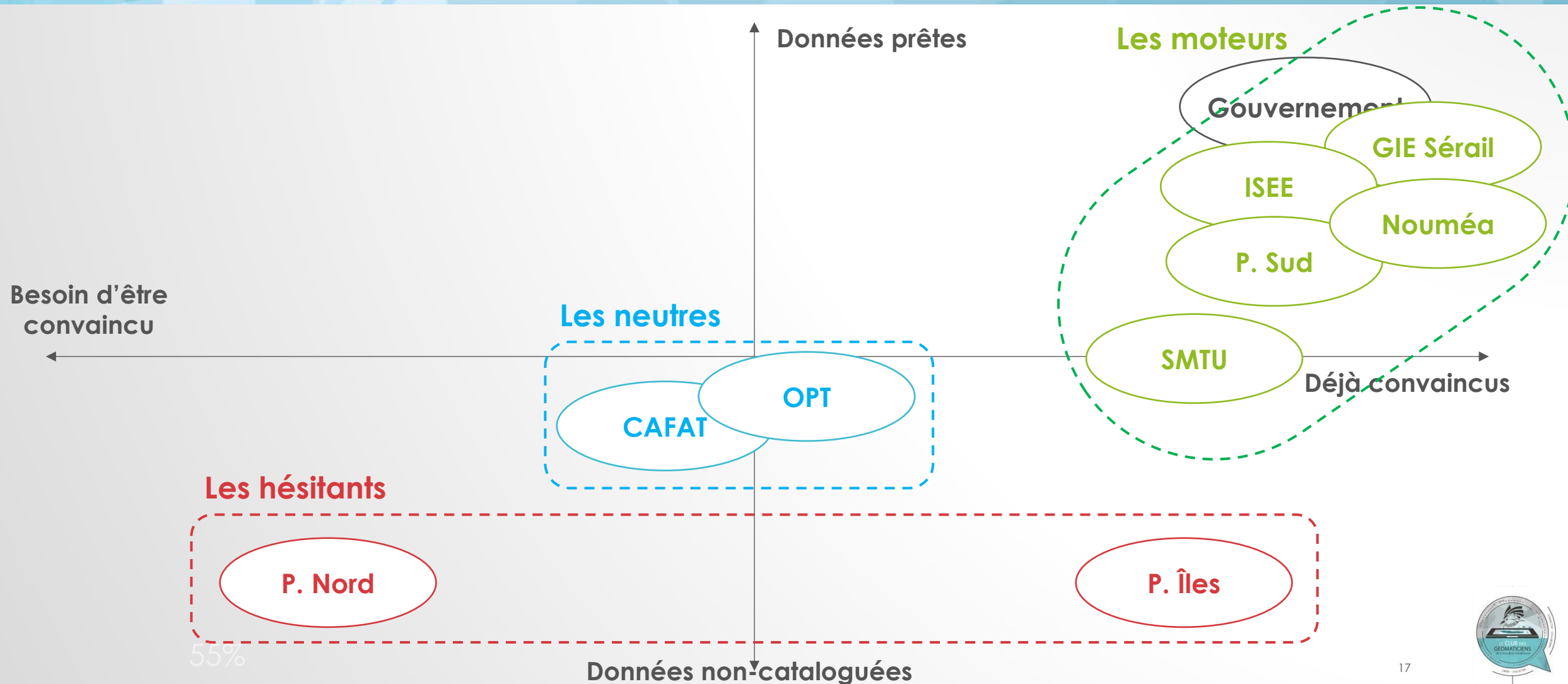
- Une licence **ouverte** , sans condition de réutilisation, licence Etalab, qui concerne 99% des données ;
- Une licence avec **une authentification obligatoire** des ré-utilisateurs, visant à s'assurer que l'utilisation des données soient compatibles avec les politiques publiques. Il s'agit d'éviter, par exemple, que le trafic routier soit envoyé devant des écoles parce que l'axe principal situé dans une rue parallèle est indiqué saturé. Ce type de licence permet des échanges entre le Grand Lyon et les ré-utilisateurs afin de mieux connaître les usages prévus et mieux comprendre les données ;
- Une licence **associée** visant à garantir un écosystème concurrentiel équitable et évitant la formation de monopoles. Pour cela, cette licence outre l'authentification du ré-utilisateur, introduit la possibilité d'une redevance en fonction d'un seuil élevé du taux de pénétration du produit ou du service créé. En cas de position tendant à devenir monopolistique la redevance est élevée. Cette licence n'intègre aucun frais fixes afin d'éviter toute barrière à l'entrée pénalisante pour les petits acteurs. Elle vise à protéger les PME innovantes et acteurs de tailles modestes et favoriser leur développement.

	 LICENCE OUVERTE	 LICENCE ENGAGÉE	 LICENCE ASSOCIÉE
Utilisation commerciale de données	✓	✓	✓
Mention de la source et de la date de mise à jour	✓	✓	✓
Déclaration sur l'usage de la donnée	✗	✓	✓
Authentification nécessaire avant téléchargement	✗	✓	✓
Possibilité d'une redevance	✗	✗	✓

SOMMAIRE

- Méthodologie
- Contexte et enjeux
- Etat de l'art hors Nouvelle-Calédonie
- ***Etat de l'art en Nouvelle-Calédonie***
- Scénarii

NIVEAU DE MATURITÉ DES FOURNISSEURS DE DONNÉES



ATTENTE DES UTILISATEURS EN MATIÈRE DE DONNÉES

GISday

Besoins exprimés par les utilisateurs

Données recommandées par la loi pour une république numérique*

Gouvernement

Données relatives à la transparence et la vie démocratique (budget, ...)

Cadastre,
Registre parcellaire

Les moteurs

Données géographiques

Répertoire des entreprises et des associations, Code officiel géographique

Les neutres

Données de santé (propagation des maladies, ...)

Base adresses,
Référentiel à grande échelle

Les hésitants

Données relatives à la transparence et la vie démocratique (budget, ...)

Reste à faire

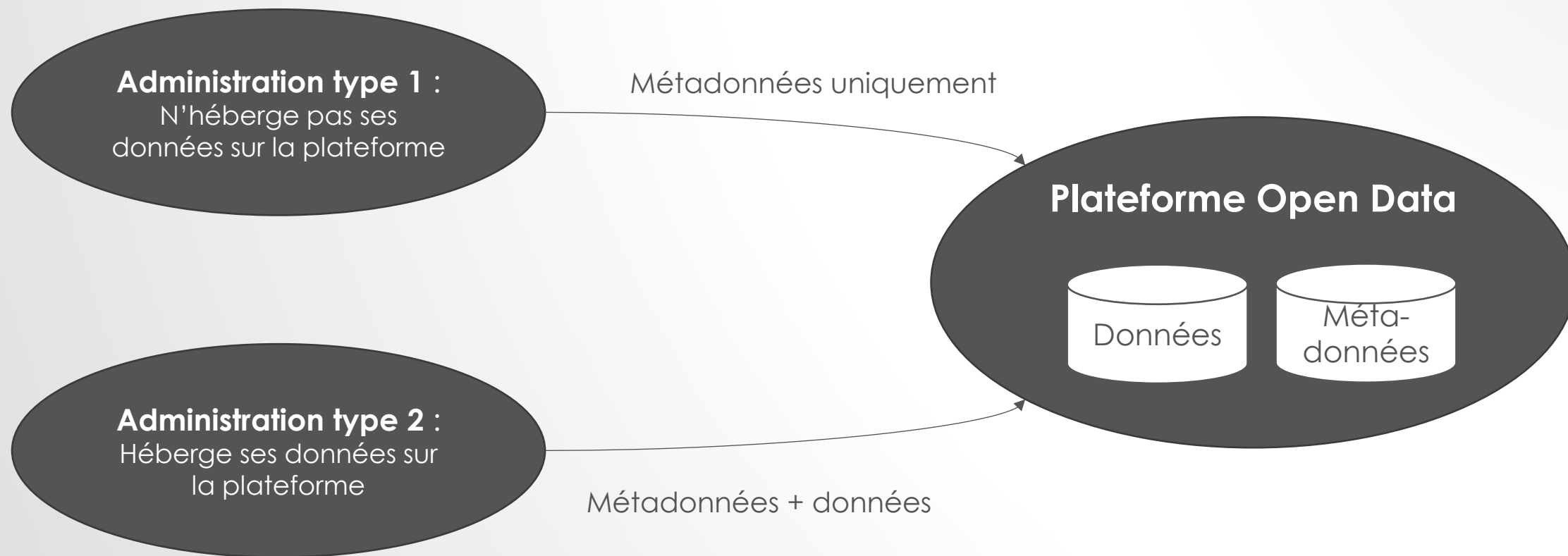
Données des observatoires (prix, immobilier...),
Données juridiques (décisions de justice, ...)

Répertoire opérationnel des métiers et des emplois (ROME),
Organisation administrative de l'état

*Plusieurs jeux de données « prioritaires » y sont préconisés. Loi non applicable en Nouvelle-Calédonie.



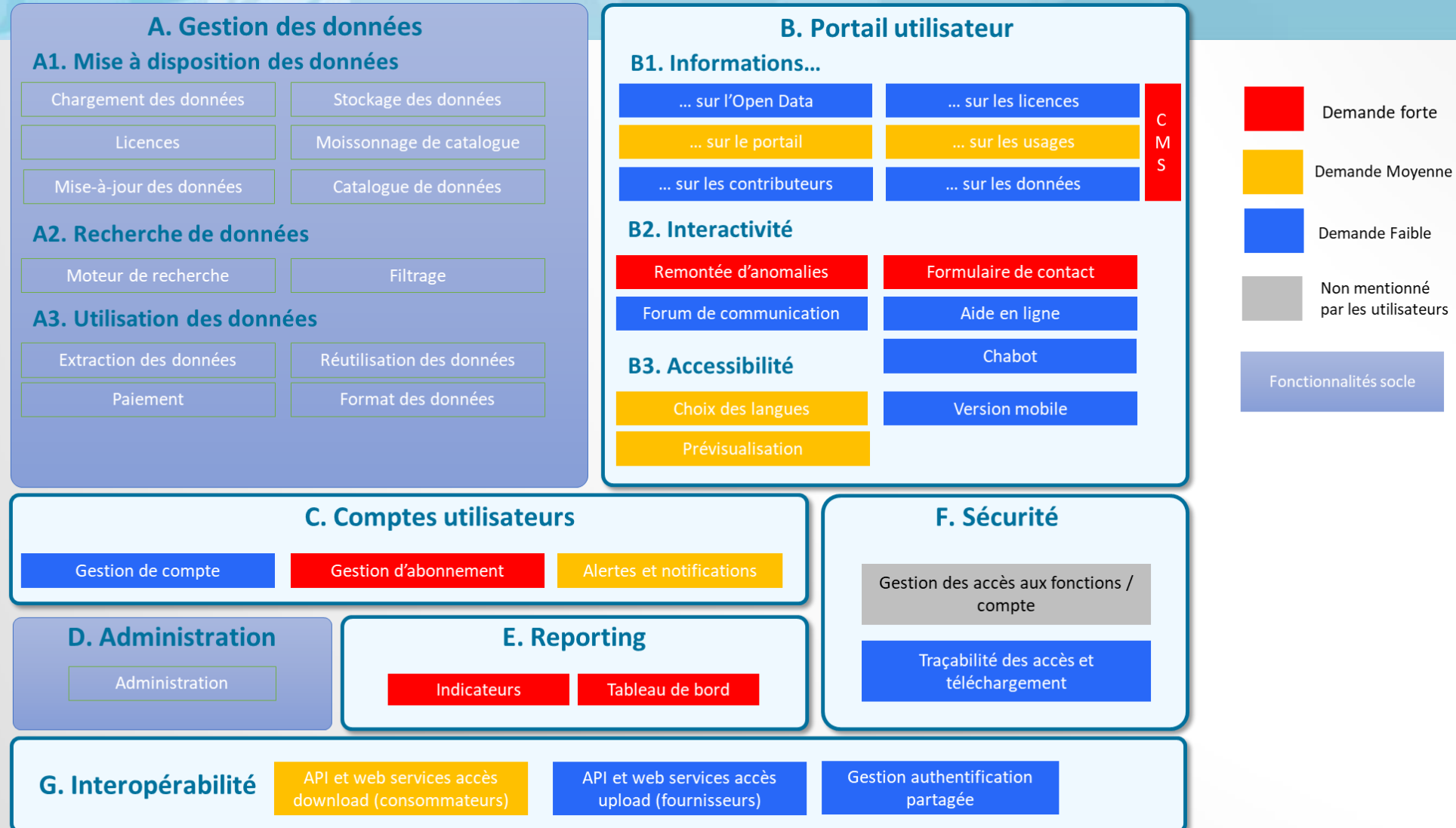
POSITIONNEMENT AU SEIN DU SI



Au préalable, chaque administration doit nettoyer & préparer ses données & métadonnées afin de les rendre utilisables par la plateforme

Attentes des utilisateurs et des fournisseurs de données

Cartographie fonctionnelle



SOMMAIRE

- Méthodologie
- Contexte et enjeux
- Etat de l'art hors Nouvelle-Calédonie
- Etat de l'art en Nouvelle-Calédonie
- ***Scénarii***

SCÉNARIOS ÉTUDIÉS

	SC 1	SC 2	SC 2bis	SC 3	SC 4	SC 5
Description	Produit internalisé au GNC	Produit externalisé en NC, hébergement au GNC	Produit externalisé en NC, hébergement hors GNC	Produit externalisé hors NC	Produit de la CPS	Structure externe
Solution technique	Progiciel du marché	Progiciel du marché	Progiciel du marché	Progiciel du marché	CKAN	Progiciel du marché
Hébergement de la plateforme	DTSI	DTSI	Datacenter en NC	Datacenter en AUS	Datacenter en AUS + une sous-instance à la DTSI	A définir par la structure
Stockage des données (en dehors du stockage chez le fournisseur de données)	DTSI	Sur la plateforme	Sur la plateforme	Data Repository à créer	Sur la sous-instance	A définir par la structure
MOA	ECONUM	ECONUM	ECONUM	ECONUM	ECONUM	A définir par la structure

BILAN DES SCÉNARIOS

	Scénario 1	Scénario 2	Scénario 2bis	Scénario 3	Scénario 4	Scénario 5
Coûts d'investissement	11 900 000 XPF	11 900 000 XPF	11 900 000 XPF	17 420 000 XPF	9 020 000 XPF	12 900 000 XPF
Coûts de fonctionnement	7 700 000 XPF (+0,5 ETP)	7 700 000 XPF (+0,5 ETP)	10 800 000 XPF (+0,5 ETP)	10 800 000 XPF (+0,5 ETP)	7 700 000 XPF (+0,5 ETP)	31 300 000 XPF (hors bureau)
Délais	Délais un peu plus courts car tout est fait en interne	Le fait d'héberger à la DTSl réduit légèrement les délais		Le développement du Data Repository entraine un délai supérieur	Il n'y a pas de phase d'appel d'offres/consultation	La création de la structure & de ses membres peut prendre du temps
Besoin	Les contraintes techniques DTSl peuvent limiter le choix de la solution	Les scénarios couvrent les besoins exactement de la même manière			La couverture fonctionnelle risque de ne pas être totale	Les partenaires pilotent directement la plateforme
Risque	Scénario où il y a le moins d'intervenants	Les VM sont à la DTSl	Les VM sont hors de la DTSl	Développement du Data Repository à faire + plateforme stockée hors NC	La solution utilisera une plateforme déjà existante, les risques techniques sont donc faibles	Risque additionnel concernant la création d'une structure externe avec les bonnes compétences

PLANNING ENVISAGÉ

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	Année +1
Mise en œuvre	C	D	D	D	D F								
Données du gouvernement		P	P	P	D F								
Animation de la plateforme		F	F	F	F	F	F	F	F	F	F	F	F
Données des « moteurs »					P	P	P	P	P	D F			
Données des « suiveurs » & « hésitants »										P	P	P	P D F



Consultation des entreprises



Préparation des données



Installation, Para, Dev, Tests & Dép.



Accompagnement, Formation

Les autres administrations pourront préparer et mettre à disposition leurs données sur la plateforme au fil de l'eau

25