

## Document d'information 13

Original : anglais

### **Nouvelles tendances en matière de diffusion et d'échange de données**





## CONTEXTE

1. Compte tenu des investissements en temps et en argent qu'exige la collecte de données, il est souhaitable que les différents publics concernés puissent bénéficier au maximum des données recueillies et des produits qui en sont dérivés. Or, les utilisateurs potentiels forment un tout très hétérogène. Il faudra donc adapter le format et le mode de visualisation des données en fonction de la catégorie de consommateurs visée.

### Données brutes

2. Les biologistes comme les halieutes souhaitent pouvoir accéder à des données brutes (désagrégées) afin de repérer les valeurs erronées et aberrantes avant de procéder à des analyses approfondies. Les données brutes sont souvent extraites sous forme tabulaire soit à l'aide de requêtes prédéfinies (telles que celles disponibles pour la base RFID ou encore les enquêtes auprès des pêcheurs et les relevés de prises sur les marchés), soit à l'aide de requêtes personnalisées sous SQL, puis elles sont analysées dans un logiciel statistique, converties en graphiques ou affichées dans un système d'information géographique (SIG).
3. Aujourd'hui, de nombreuses voix s'élèvent pour que les ensembles de données brutes (une fois apurés et analysés) soient diffusés parmi les milieux scientifiques grâce à des portails de données, en particulier lorsque la collecte de données a été financée sur fonds publics. Par exemple, le Système mondial d'information sur la biodiversité (GBIF) est un portail où sont centralisées des données d'occurrence (observations) sur les organismes biologiques, dont les coraux, les graminées marines, les poissons et les invertébrés. Le portail fournit également des références taxonomiques.
4. Dans la plupart des cas, les données brutes sont exportées sous forme de tableaux au format texte plat.

### Données agrégées, tableaux de bord et scores

5. Les responsables de la gestion des pêches souhaitent généralement disposer de données agrégées sur la production, la santé des stocks, l'application de la réglementation et les avis de gestion formulés à partir de l'analyse scientifique.
6. Les données agrégées sont transmises aux gestionnaires sous forme de tableaux de bord et de rapports synthétiques au format prédéfini. Ainsi, le rapport personnalisé établi pour la filière bêche-de-mer de la Polynésie française montre, à l'échelle de chaque île, le quota autorisé, le volume expédié et le pourcentage du quota utilisé. Autre exemple, l'interface en ligne du site web Dorado propose des requêtes prédéfinies permettant aux usagers d'accéder aux données agrégées de TUFMAN 2.
7. Le catalogue des visualisations de données (<http://www.datavizcatalogue.com/index.html>) recense 60 représentations graphiques distinctes pouvant être générées, du classique diagramme en bâtons ou en camembert aux illustrations plus étonnantes, telles que les diagrammes en spirale ou les cartes chloroplèthes. Le choix du graphique dépendra de ce que

l'utilisateur souhaite montrer, par exemple une relation, une comparaison, une distribution ou une composition.

8. Les illustrations de scores, les cartes de données et les indicateurs sont de plus en plus utilisés pour afficher un volume important d'informations sous forme synthétique. Par exemple, le site web « Ocean Health Index » (indice de la santé des océans de Conservation International) représente les scores obtenus pour les différents indicateurs par pays sur une rose de Florence Nightingale, et une liste d'indicateurs potentiels a été proposée au cours de la dernière Conférence des directeurs des pêches.

## Couches SIG et visualisation spatiale

9. Les données qui ont une composante spatiale peuvent s'afficher dans un SIG, se superposer sur des cartes et se combiner à d'autres couches. Pour les fermes aquacoles, par exemple, l'emplacement des concessions maritimes et des bassins d'élevage est souvent ajouté sur des cartes géographiques ou marines. De même, on peut afficher sur une carte marine ou une image satellite le périmètre des aires marines protégées (AMP), les densités d'invertébrés ou de poissons, ou encore les routes empruntées par les bateaux.
10. On compte plusieurs dépôts de données géospatiales dans la région, dont PacGeo ([pacgeo.org](http://pacgeo.org)). L'édition des données géospatiales sur les pêches obéit aux mêmes règles de respect de la vie privée et de sécurité que celles applicables à l'édition de données brutes ou agrégées. Sur certains portails, on peut combiner affichage de couches SIG et visualisation des données, et télécharger l'ensemble de données (<http://geoportail.oeil.nc/galaxia/>).
11. Les bases de données en ligne regroupant les données des inventaires des herbiers, des crabes de cocotier et des mangroves permettent l'exportation de données brutes et agrégées dans des formats pouvant être directement intégrés dans des SIG (KML, ShapeFile), tandis que d'autres données sont générées au format texte avec des coordonnées de longitude et de latitude (RFID).

## PROBLEMATIQUES ET PREOCCUPATIONS

### Différence entre données et information

12. Les données forment des éléments de base qui peuvent être traités et analysés, tandis que l'information suppose une explicitation et une mise en contexte. Un rapport livre de l'information et sera généralement accessible au public, tandis qu'un fichier texte plat comprenant des colonnes de chiffres constitue un élément de donnée, dont l'accès est souvent restreint pour des raisons liées à la protection de la vie privée et à la sensibilité des données (données commerciales, lieux de pêche ou lieux où se trouvent les ressources, données personnelles).
13. Pour exploiter correctement les données brutes, il est nécessaire de posséder une bonne connaissance du périmètre de l'étude concernée, des méthodes employées et des limites associées afin d'éviter toute interprétation erronée. Ainsi, les densités d'invertébrés calculées à partir d'inventaires doivent être analysées par type de station et ne peuvent généralement pas à être regroupées ou généralisées à une zone entière par extrapolation. Utiliser des

données sans en saisir le contexte ou la finalité peut amener l'utilisateur à formuler des conclusions fausses ou à prendre de mauvaises décisions.

14. Plusieurs moyens sont à l'étude pour la distribution des données côtières brutes et agrégées, notamment les portails de données existants pour les données tombées dans le domaine public ou « désensibilisées » et le site web de la CPS lorsque des règles d'authentification et de sécurité s'appliquent.

### **Biais dans les indicateurs et la visualisation des données**

15. Le choix du type de diagramme, de l'échelle et de la colorimétrie est très important, dans le sens où un choix inadapté peut induire en erreur l'utilisateur. Par exemple, le rouge est souvent synonyme de problème ou de danger dans notre culture, tandis que le vert évoquera la bonne santé d'un stock ou l'absence de problème. Compte tenu du puissant biais cognitif associé aux couleurs rouge, orange et vert, ces couleurs doivent être utilisées avec prudence ou remplacées par des couleurs neutres (dans les premières versions de l'indice de santé des océans OHI+, le score des pêcheries s'affichait en rouge sur la rose de Nightingale, ce qui pouvait prêter à confusion). Cette précaution est particulièrement importante pour les illustrations de scores et les tableaux de bord, même s'ils sont accompagnés d'explications.

### **Rôles de la CPS, des services des pêches nationaux et des ONG dans la diffusion de l'information**

16. Le Programme pêche côtière de la CPS a essentiellement pour rôle de transmettre des avis scientifiques et techniques aux services des pêches des pays. Certains supports de sensibilisation et d'éducation sont aussi régulièrement élaborés et envoyés aux services nationaux et aux ONG, qui se chargent alors de les distribuer aux populations locales. La distribution de supports imprimés à l'échelle d'un pays pose parfois problème. Compte tenu de la généralisation des téléphones portables, il serait peut-être envisageable de mettre au point et de distribuer des applications mobiles permettant à un plus large public d'accéder à l'information relevant du domaine public.

### **SUJETS DE DISCUSSION POTENTIELS**

- Quels sont les formats d'exportation et de visualisation des données que les services des pêches souhaiteraient voir développés en priorité dans les bases de données en ligne de la CPS ?
- Est-il nécessaire de faire dialoguer les ensembles de données existants avec d'autres systèmes, par exemple via des interfaces de programmation (API), et si oui, lesquels ?
- Comment pouvons-nous améliorer la diffusion de l'information et des données au profit des communautés ? Peut-on voir dans les applications mobiles une solution nouvelle pour diffuser l'information ? Si oui, à qui revient la responsabilité du contenu des données ?
- Quels types de retours d'information ou de récompenses devrions-nous donner aux acteurs impliqués dans la collecte de données (en particulier, non halieutiques) afin de les inciter à maintenir leurs efforts ?

- La CPS peut-elle diffuser des données pour le compte des pays ? Quel dispositif faudrait-il privilégier pour obtenir l'autorisation officielle de diffusion ?

## ANNEXE

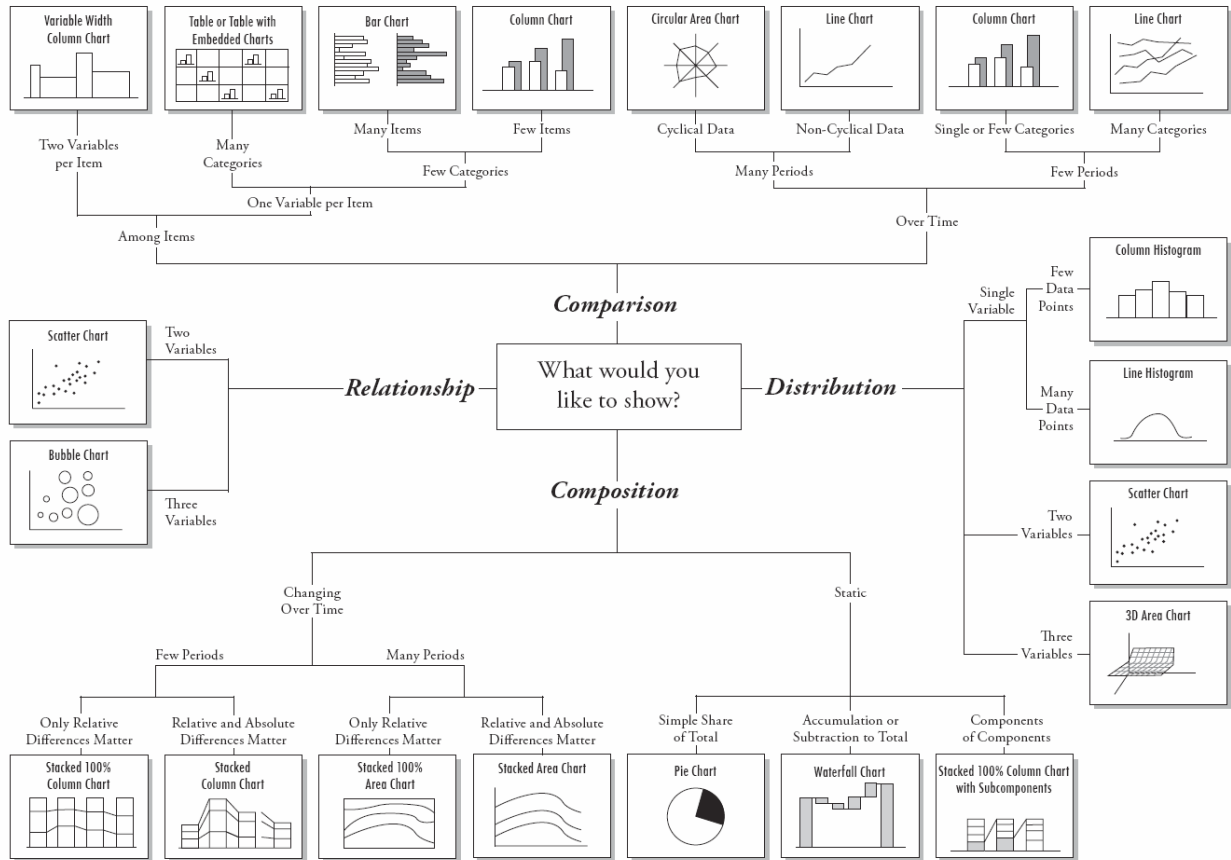


Figure 1. Comment choisir le bon diagramme [Abela 2013]

## BIBLIOGRAPHIE

Abela, A. 2013. Advanced Presentations by Design: Creating Communication that Drives Action. Abela 2013 Advanced Presentations by Design: Creating Communication that Drives Action, 2<sup>nd</sup> ed.