



UNIVERSITÉ DE
BORDEAUX

Université Bordeaux 1

351 cours de la Libération
33405 TALENCE Cedex

Mathias Asseray

NEWSgraphe

NEWSgraphe

SARL Adrénaline
34 bis, avenue de Mirande
33200 BORDEAUX

**Stage facultatif effectuer du 2 mai
au 5 août 2011**

Rapport de Stage

**Réalisation automatique d'images structurées à partir de flux XML
de contenus ainsi que de ressources graphiques externes.**

Maître de Stage : Bruno Capbern.

Partie 1 :

Présentation de l'entreprise et du secteur d'activité.

L'illustration de presse est un milieu en pleine mutation. La dématérialisation des contenus et la chute des ventes des journaux papier ont été des facteurs qui ont forcé l'industrie de la presse écrite à s'adapter à un nouveau marché de l'information. Dans ce marché en pleine crise, les agences de presse qui sous traitent certains articles ou certaines illustrations pour des rédactions ont dû également s'adapter en proposant des contenus à la valeur ajoutée toujours constante mais au coût de réalisation diminué.

C'est dans cette optique que l'agence de presse NewsGraphe a souhaité la réalisation d'une application qui, à partir de données statistiques, rendra la création d'une illustration le plus automatique possible.

Adrénaline (Sarl) est une société bordelaise fondé en 1998 par Nicolas Peyrebrune et reprise en tant qu'associé par Bruno Capbern en 2006. A partir de cette date, la société centre ces activités sur la presse et commercialise ses produits sous la marque NewsGraphe.

Nicolas Peyrebrune est avant tout un illustrateur, avant de fonder NewsGraphe il travaillait en tant que correspondant de presse (illustrateur) pour le journal Sud Ouest. De son côté, Bruno Capbern a une double formation d'informaticien et en gestion industrielle. Ils se connaissent tout les deux depuis longtemps et décident de monter NewsGraphe. Le statut d'agence de presse, obtenu en 2008 est une étape de plus qui leur permet d'avoir une crédibilité plus importante vis-à-vis de client potentiels.

Depuis sa création, NewsGraphe a eu pour clients principaux Sud Ouest (journal quotidien), Onze mondial (magazine de football mensuel), le 10 sport (magazine de sport quotidien puis hebdomadaire). A la suite de remaniements et de coupes dans les budgets, les partenariats avec le 11 mondial et le 10 sport se sont arrêtés.

Avec la crise qui a frappé le secteur de la presse écrite le prix des infographies est devenu beaucoup plus difficile à négocier qu'auparavant.

Le prix d'une infographie est indexé sur plusieurs critères et de manière différente suivant les acteurs concernés.

Par exemple pour NewsGraphe, le prix d'une infographie est évalué selon deux facteurs:

- La conception, qui regroupe tout ce qui traite de la recherche d'informations, du traitement de ces informations ainsi que de la rédaction des différents textes de l'infographie. Plus l'infographie apportera d'informations, plus ce facteur sera élevé.
- La réalisation, la partie graphique du processus. Plus les techniques de réalisation seront compliquées et prendront du temps plus le facteur sera élevé. A savoir, l'infographie nécessite-t-elle une illustration? Cette dernière doit-elle être réalisée en 2D ou en 3D ? En noir et blanc ou en couleur? Faut-il ajouter et détourner des photos?



Figure 1: Une infographie parmi les plus simples que propose NewsGraphe.

Pour les rédactions qui commandent ces infographies, le calcul se fait différemment: ils jugent une infographie sur sa qualité, sur son originalité, sur sa taille - pouvant varier de quelques lignes dans une colonne à une double page -, et également sur l'exclusivité lié à cette infographie – le fait que l'agence de presse qui leur fournit l'illustration puisse revendre à une autre rédaction cette même infographie.

Il convient alors de tomber d'accord sur une grille de tarifs. Actuellement, les infographies réalisées par NewsGraphe peuvent aller de 80€ à 600€ selon les critères vus précédemment.

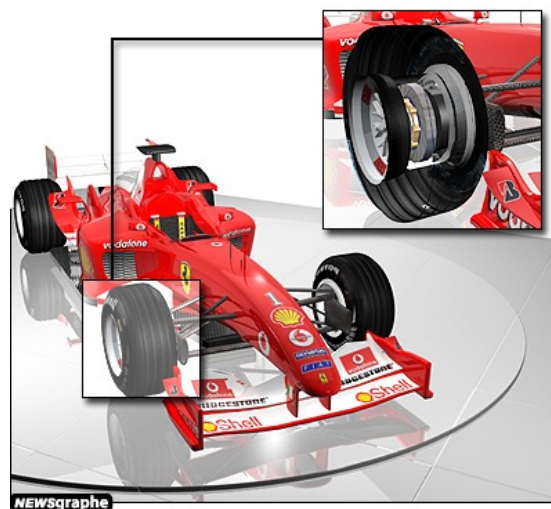


Figure 2: Exemple de l'emploi de la 3D dans une infographie.

Bordeaux - Monaco : duel au sommet

Diarra ou Nené ? Le leader ou son dauphin ? Bordeaux ou Monaco ? Bien malin qui saura trouver le résultat ! Bordeaux aura l'avantage du terrain, mais peut-être la tête à Munich. De son côté, Monaco vient d'aligner 3 victoires. Avouons-le, nous sommes tentés de suivre notre ordinateur qui pronostique un match nul avec deux données phares : fatigue proche (49 % pour Bordeaux et 51% pour Monaco) et cote identique (49,99 % pour Bordeaux contre 50,01% à Monaco).

NOTRE CLASSEMENT AVANT LA 11^E JOURNÉE

1	 DIARRA BORDEAUX Note : 67,43 Titularisations : 10 Pts/fit. : 2,2	6	 COULIBALY AUXERRE Note : 64,9 Titularisations : 10 Pts/fit. : 1,7	11	FANNI	RENNES	63,23
2	 NENÉ MONACO Note : 67,24 Titularisations : 10 Pts/fit. : 2,1	7	 HENGBART AUXERRE Note : 64,28 Titularisations : 10 Pts/fit. : 1,7	12	PUYGRENIER	MONACO	63,02
3	 DU COURTIOUX VALENCIENNES Note : 66,73 Titularisations : 6 Pts/fit. : 2,5	8	 GOURCUFF BORDEAUX Note : 63,95 Titularisations : 7 Pts/fit. : 2,7	13	PITAU	MONTPELLIER	62,85
4	 HANSSON RENNES Note : 65,32 Titularisations : 7 Pts/fit. : 2,1	9	 PEDRETTI AUXERRE Note : 63,43 Titularisations : 10 Pts/fit. : 1,7	14	CHAMAKH	BORDEAUX	62,63
5	 MONGONGU MONACO Note : 65,21 Titularisations : 10 Pts/fit. : 2,1	10	 MONTAÑO MONTPELLIER Note : 63,4 Titularisations : 10 Pts/fit. : 2	15	SANCHEZ	VALENCIENNES	62,63
				16	GRICHTING	AUXERRE	62,56
				17	MARVEAUX	MONTPELLIER	62,03
				18	CHALMÉ	BORDEAUX	61,56
				19	FATY	SOCHAUX	61,47
				20	LLORIS	LYON	61,28
				21	DZODIC	MONTPELLIER	61,26
				22	GOMIS	VALENCIENNES	61,23
				23	KOSCIELNY	LORIENT	60,92
				24	TREMOLINAS	BORDEAUX	60,86
				25	SORIN	AUXERRE	60,75
				26	BISEVAC	VALENCIENNES	60,69
				27	LEROY	RENNES	60,52
				28	FÉRET	NANCY	60,46
				29	COSTA	MONTPELLIER	60,42
				30	KALLSTROM	LYON	60,34
				31	PLANUS	BORDEAUX	60,34
				32	GIGNAC	TOULOUSE	60,09
				33	DOUCHEZ	RENNES	59,99
				34	YANGAMBIWA	MONTPELLIER	59,64
				35	BRANDAO	MARSEILLE	59,57



TOP 10 LA RÈGLE DU JEU

Après chaque journée de L1 et à partir des données officielles publiées par la LFP nous attribuons une note à l'ensemble des joueurs du championnat. Il s'agit d'évaluer la participation de chacun aux résultats de son équipe.
Le paramètre principal consiste en l'étude statistique de l'évolution des scores en fonction des périodes de présence et d'absence du terrain de chaque joueur. Les notes sont bien sûr évaluées différemment selon le poste occupé.
Nous procédons ensuite à un redressement mathématique pour que chaque joueur, qu'il soit gardien de but, défenseur, milieu de terrain ou attaquant, ait une chance équivalente de figurer au classement.

TOP 10 : LE MEILLEUR PAR POSTE



ZOOM DERNIÈRE JOURNÉE

Bordeaux	3 - 0	Le Mans
Boulogne	1 - 3	Monaco
Grenoble	1 - 2	Nancy
Nice	4 - 1	Lyon
Saint-Etienne	0 - 2	Valenciennes
Sochaux	1 - 0	Lorient
Rennes	3 - 0	Montpellier
Auxerre	3 - 2	Lille
Lens	0 - 2	Toulouse
Marseille	Reporté	Paris SG

L'ÉQUIPE TYPE DE LA 10^E JOURNÉE



LES MEILLEURS DE CHAQUE CLUB

Les joueurs de l'OM et du PSG sont pénalisés par le report du Clasico. Dès la semaine prochaine, nous attribuerons un bonus temporaire aux équipes comptant des matchs en retard.






 AUXERRE 1 COULIBALY Note : 64,9 2 HENGBART : 64,28 3 PEDRETTI : 63,43	 BORDEAUX 1 DIARRA Note : 67,43 2 GOURCUFF : 63,95 3 CHAMAKH : 62,63	 BOULOGNE 1 RABUEL Note : 50,3 2 TOURÉ : 48,68 3 THIL : 47,72	 GRENOBLE 1 LJUBOJA Note : 41,74 2 BATTLES : 40,01 3 ROMAO : 39,46
 LE MANS 1 LE TALLEC Note : 55,55 2 GOULON : 53,69 3 CORCHIA : 52,66	 LENS 1 BOUKARI Note : 51,22 2 CHELLE : 50,14 3 KOVACEVIC : 47,88	 LILLE 1 RAMI Note : 55 2 BÉRIA : 52,68 3 GERVINHO : 50,99	 LORIENT 1 KOSCIELNY Note : 60,92 2 DIARRA : 57,05 3 AUDARD : 56,06
 LYON 1 LLORIS Note : 61,28 2 KALLSTROM : 60,34 3 RÉVEILLÈRE : 58,79	 MARSEILLE 1 BRANDAO Note : 59,57 2 TAIWO : 57,35 3 NIANG : 56,81	 MONACO 1 NENÉ Note : 67,24 2 MONGONGU : 65,21 3 PUYGRENIER : 63,02	 MONTPELLIER 1 MONTAÑO Note : 63,4 2 PITAU : 62,85 3 MARVEAUX : 62,03
 NANCY 1 FÉRET Note : 60,46 2 DIA : 55,32 3 LEMÂITRE : 54,34	 NICE 1 RÉMY Note : 58,07 2 MABIALA : 50,81 3 HELLEBUYCK : 49,9	 PARIS SG 1 MAKELELE Note : 53,44 2 ARMAND : 51,89 3 SAKHO : 50,87	 RENNES 1 HANSSON Note : 65,32 2 FANNI : 63,23 3 LEROY : 60,52
 SAINT-ÉTIENNE 1 MATUIDI Note : 55,77 2 LANDRIN : 54,09 3 NDAW : 52,82	 SOCHAUX 1 FATY Note : 61,47 2 BRÉCHET : 58,71 3 PERQUIS : 58,04	 TOULOUSE 1 GIGNAC Note : 60,09 2 SISSOKO : 58,11 3 CAPOUE : 55,68	 VALENCIENNES 1 DU COURTIOUX Note : 66,73 2 SANCHEZ : 62,63 3 GOMIS : 61,23

Figure 3: Un modèle d'infographie qui n'a jamais put être vendu car jugé trop complexe et trop grand.

L'automatisation de la création d'infographies permettrait de baisser les prix de la réalisation graphique de l'infographie d'environ 50%, et ainsi de pouvoir être plus compétitif sur le marché.

C'est dans ce but que NewsGraphe a développé un partenariat avec la société anglaise OptaSports Data - leader européen sur le marché de la fourniture de statistiques sportives – et a souhaité initié une dynamique de développement et de recherche dans le but de créer des applications capable de générer des infographies de différents types comme des illustrations statistiques, des cartes ou des diagrammes explicatifs.

Pour amorcé cette dynamique, NewsGraphe a donc proposé un stage qui aurai pour objectif principal le développement d'une application qui permettrai de générer une infographie sportive, à partir de données statistiques.

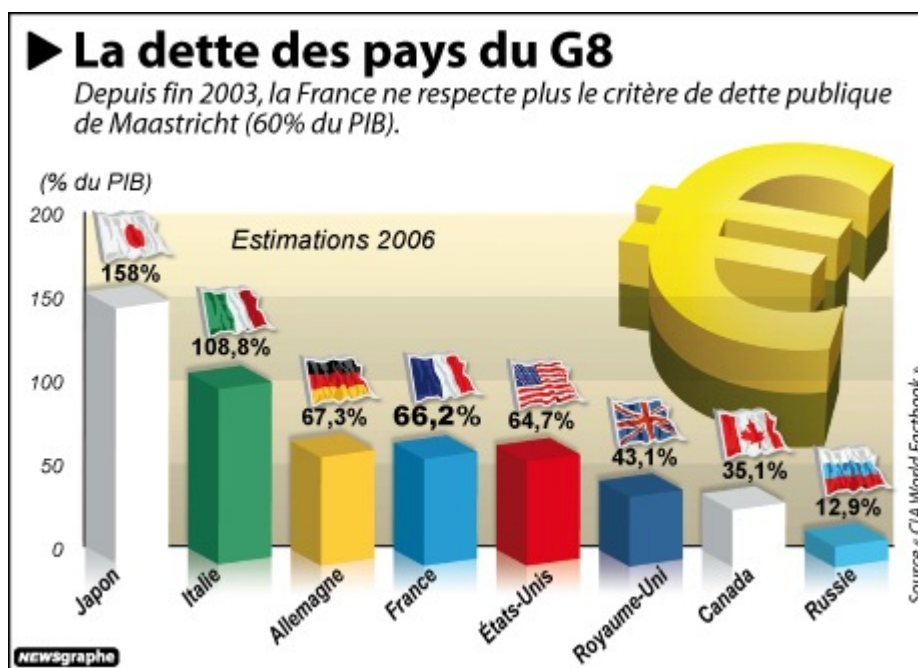


Figure 4: Un autre type d'infographie proposé par NewsGraphe.

Partie 2 :

Étude et choix des technologies.

Le premier objectif de ce stage a d'abord été de déterminer quelles technologies seraient les plus adaptées pour développer une telle application, sachant que le format des fichiers statistiques était déjà fixé et qu'il s'agissait du format XML.

2.1 Format de l'infographie :

Dans un premier temps, il a fallu choisir le format avec lequel l'infographie serait encodé. L'illustration de presse utilise différents formats vectoriels, qui permettent contrairement aux formats matriciels d'agrandir ou réduire n'importe quelle illustration sans aucune perte d'information, alors que les formats matriciels présenteront des effets de crénelage ou de flou suivant les algorithmes utilisés.

Le format EPS (Encapsulated PostScript) est un format vectoriel développé par Adobe en 1992 et basé sur le langage PostScript. C'est un format très utilisé dans le milieu de la presse. Sa spécification est disponible librement, mais il est utilisé majoritairement par le biais de la suite de logiciels d'Adobe, la Creative Suite, qui comprend notamment Adobe Illustrator et Adobe Photoshop.

Un autre format utilisé régulièrement en infographie est le format PDF (Portable Document Format), également créé par Adobe, normalisé par l'ISO en 1995.

Un troisième format, le format SVG, est un format vectoriel qui a été développé et normalisé par le W3C, sur une base de langage XML, principalement pour le web. C'est un format qui est très documenté, et contrairement aux deux autres présentés précédemment il ne nécessite pas de licences de quelque type de ce soit et s'appuie sur du XML, une structure parfaitement pris en compte par de nombreuses bibliothèques dans les principaux langages de programmation.

Tout en gardant à l'esprit que ce format ne serait utilisé que pour élaborer une version de travail de l'infographie, et qu'il serait ensuite possible de la convertir dans le format désiré, le SVG a été choisie comme format d'encodage de l'infographie.

2.2 Choix du langage de programmation :

Il a ensuite fallu déterminer quel pourrait être le langage de programmation de l'application, en étudiant les différentes bibliothèques, gérant au mieux le SVG et le XML, existantes pour chaque langage. De plus, au début du stage, l'idée était d'exécuter l'application sur un serveur, sur lequel serait déposé les fichiers statistiques XML, qui seraient ensuite utilisés par l'application pour générer l'infographie.

Le choix du langage de programmation s'est donc naturellement porté vers une solution java ou PHP. A la suite de la réalisation de plusieurs prototypes en PHP ou en Java, il s'est avéré que le langage java offrait plus de possibilités en terme de bibliothèques disponibles et un plus grand confort de travail par le biais de l'IDE Eclipse. L'ajout de bibliothèques comme JDOM pour la gestion des fichiers XML ou Batik SVG Toolkit pour la conversion de l'infographie SVG en un

autre format ont conforter ce choix.

2.3 Prise en main des technologies :

Le processus de détermination des meilleurs technologies employables pour la création de l'application final a conduit au développement d'une dizaine de prototypes, permettant de tester le fonctionnement de telle ou telle bibliothèque, de créer des infographies le plus basique possible permettant de prendre en main le format SVG et ses spécifications. Il a également fallut installer tous les supports logiciels nécessaires à ces tests, comme les serveurs Tomcat et Wamp pour étudier la faisabilité d'applications en ligne java et PHP.

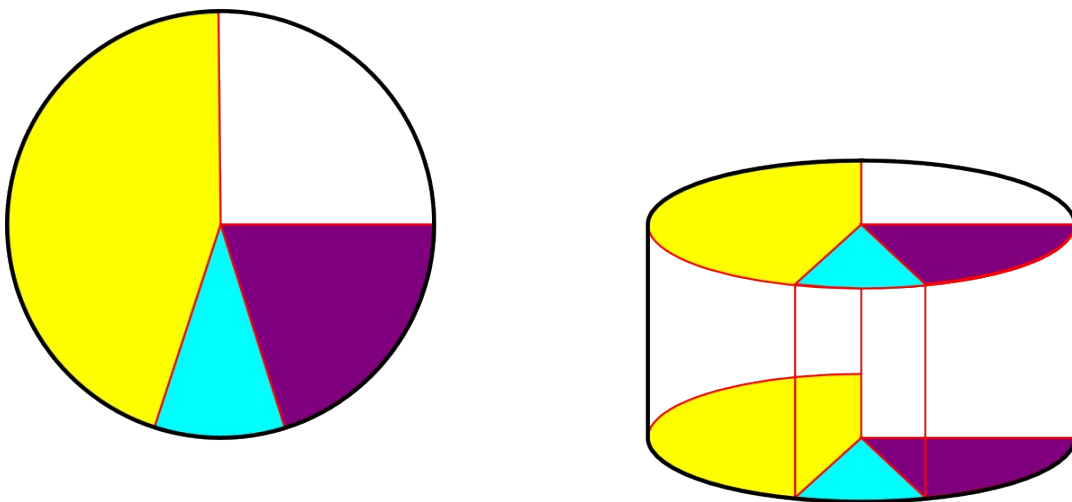


Figure 5: Infographie obtenu par l'exécution d'un des premiers prototypes sur la prise en main du format SVG.

Partie 3 :

Développement et construction des fichiers STRUCT et DATA :

Après avoir déterminé quelles seraient les technologies employées, le développement de l'application a commencé, et en même temps l'élaboration des fichiers STRUCT et DATA, des fichiers XML qui forment les deux pierres angulaire de l'application et qui sont lus et exploités par l'application à chaque génération d'une infographie.

3.1 Le fichier STRUCT :

L'application finale ayant pour but de produire une infographie, il semblait évident qu'il faudrait stocker à un moment ou un autre la structure de cette infographie. Renseigner l'emplacement des différents éléments graphiques, renseigner les polices de caractères qui apparaissent dans l'infographie, les éléments SVG qu'il va falloir importer et l'emplacement des textes. Ce fichier STRUCT fait donc office de squelette de l'infographie. Il sera donc nécessaire de créer un fichier STRUCT par type d'infographie.

3.2 Le fichier DATA :

Le fichier DATA va contenir quant à lui toutes les données de l'infographie : l'intitulé des textes, les valeurs statistiques, les valeurs de certaines couleurs qui peuvent changer. Le fichier DATA sera donc propre à chaque infographie.

Un des facteurs qu'il a fallu prendre en compte lors de l'élaboration de ces fichiers XML, est la lisibilité et la compréhension du contenu par un lecteur humain de manière à pouvoir confier l'édition de fichiers DATA à n'importe qui même s'il ne connaît pas du tout le langage XML, cela passe par des zones de commentaires explicites qui décrivent ce à quoi correspondent les différentes valeurs.

3.3 Développement :

Le développement de l'application s'est déroulé en deux parties, qui correspondent aux deux types d'infographies qu'il fallait pouvoir réaliser. Chaque type d'infographie ayant son propre fichier STRUCT ainsi qu'un contenu très différent :

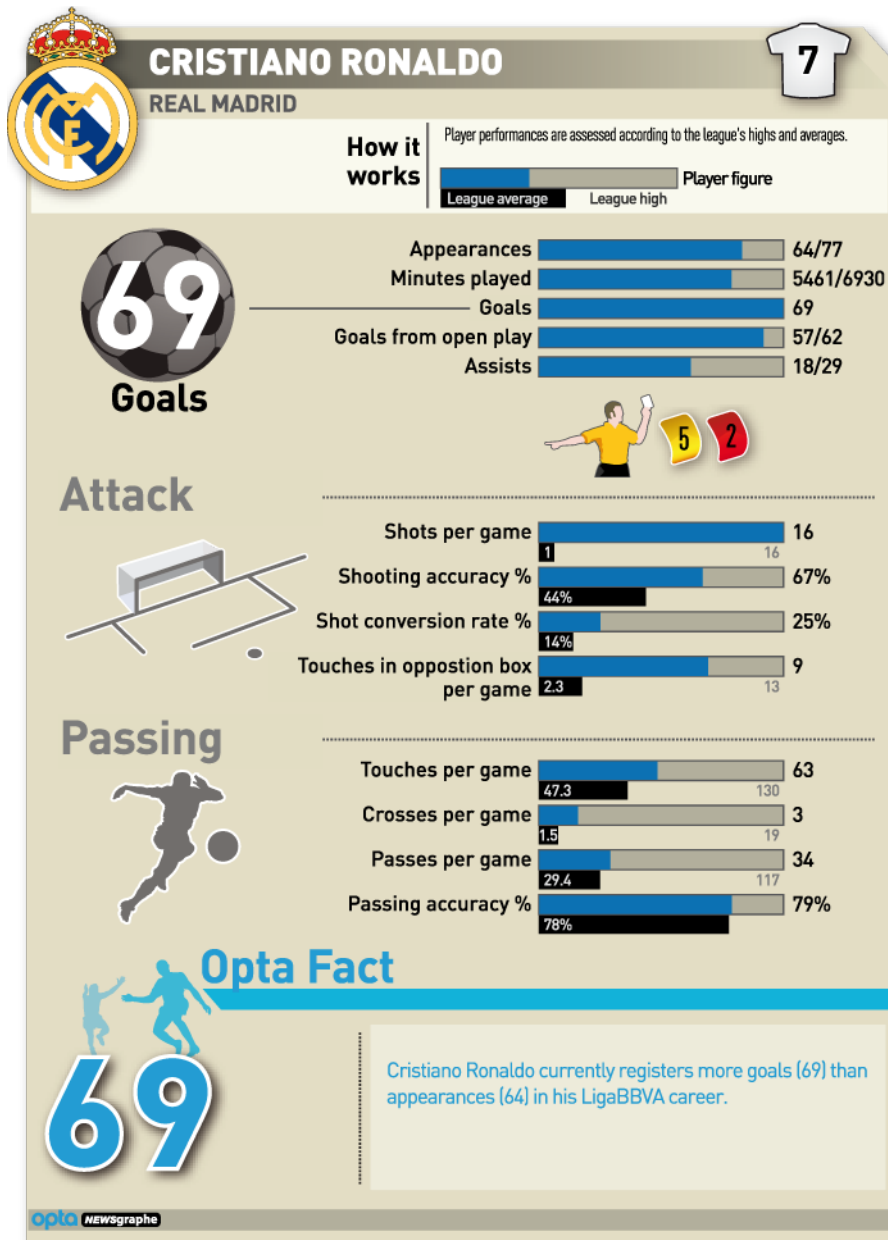


Figure 6: Premier type d'infographie réalisé.

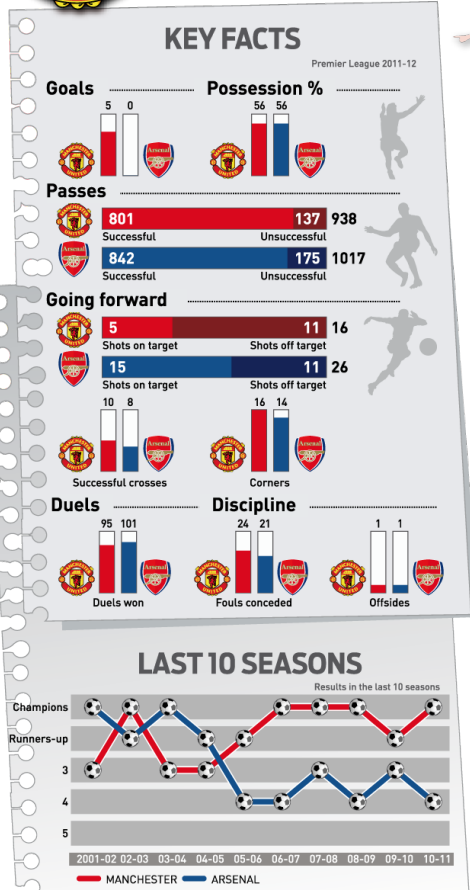
On peut constater qu'avec ce type d'infographie, il n'y a qu'une sorte d'élément qui doit être calculé : les jauges, qui ne sont finalement que de simples rectangles à largeur variable. Ce qui, en SVG est relativement simple à mettre en œuvre. Tous les éléments qui peuvent paraître compliqué comme le fond, les logos ou les cartons sont directement importés depuis des fichiers externes.



MANCHESTER



ARSENAL



Stadium : Old Trafford
Kick-off time : August 28th, 2011 16:00

PROBABLE LINEUPS



DUEL

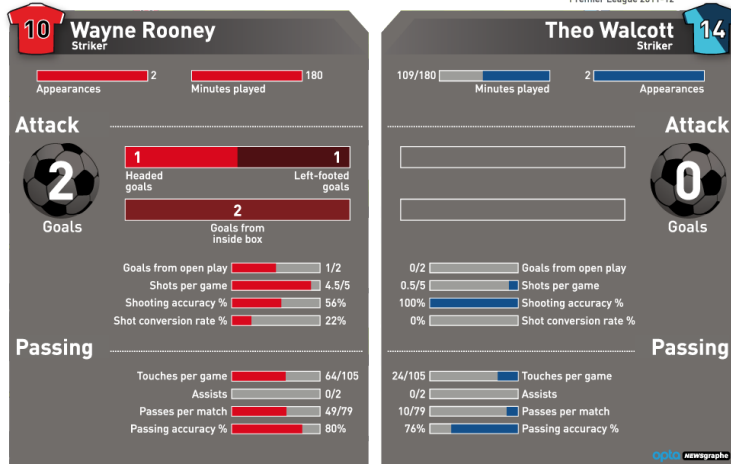


Figure 7: Deuxième type d'infographie réalisé.

Ce type d'infographie est beaucoup plus complexe et plus long à mettre en place de part le nombre plus élevé d'éléments graphiques statistiques différents. Sur l'infographie précédente on pouvait distinguer un voir deux types de jauges. Ici il y en a quatre, plus le graphique des résultats et la formation tactique de chaque équipe.

Sitographie :

NewsGraphe : www.newsgraphe.com

Site internet de la société NewsGraphe d'où sont issues les quatre premières figures.
(dernier accès : 05/09/2011)

Spécification du format EPS :

http://partners.adobe.com/public/developer/en/ps/5002.EPSF_Spec.pdf

(dernier accès : 05/09/2011)

Spécification du format SVG :

<http://www.w3.org/Graphics/SVG/>

(dernier accès : 05/09/2011)

Recommandation W3C pour SVG 1.1:

<http://www.w3.org/TR/SVG11/>

(dernier accès : 05/09/2011)

Spécification de la bibliothèque JDOM pour la gestion des fichiers XML :

<http://www.jdom.org/docs/apidocs/>

(dernier accès : 05/09/2011)