Les données : expériences, observations et traitements

Francis Laloë

Dans un recueil de dessins publié en 1962 (éditions Denoël) et intitulé « Rien n'est simple », SEMPÉ présente une scène urbaine. Sur une première image, un camion heurte la façade d'un immeuble, renversant un cycliste et entrant en collision avec un autre véhicule. Un échafaudage s'effondre avec ses occupants. La seconde image présente l'état des lieux juste après l'accident. Des corps jonchent le sol et des gens accourent ou se précipitent aux fenêtres et aux portes pour regarder. Sur le trottoir d'en face, une dame a assisté à toute la scène. Elle se précipite alors chez une autre dame et, en lui restituant une toute petite partie de la seconde image, lui rapporte qu'elle a vu un couple, à la porte d'un bar.

Cette séquence de dessins reflète nombre des caractéristiques du contexte actuel de l'information et des données.

◆ Le nombre d'acteurs est considérable ; le dessinateur, les deux dames, le conducteur du camion, chaque lecteur, le couple, les victimes… Les rôles et les préoccupations de chacun d'entre eux sont différents. Certains sont actifs, d'autres sont passifs, parmi lesquels certains deviendront actifs à cause de l'événement décrit, à l'issue duquel nombre de points de vue sont bouleversés.

◆ Il y a des données et une restitution de ces données. Cette restitution n'a de sens que parce que les deux dames se connaissent et qu'elles ont des références communes. Bien entendu chaque lecteur fait une tout autre interprétation des images, très fournies, d'une scène urbaine, interprétation dans laquelle le dessinateur l'a égaré, à dessein, en rendant naturelle l'idée selon laquelle il s'agit là d'un grave accident. Conditionnellement à cette représentation le lecteur réalise implicitement sa propre synthèse en termes de nombre de victimes, de dégâts matériels et recherche sur l'image les données qui permettront de le faire de la façon la plus efficace possible…

◆ Si plusieurs restitutions peuvent être faites, selon plusieurs représentations différentes de la réalité présentée, toutes ne pourront pas être réalisées avec la même précision. Ainsi on pourrait se demander à quelle époque de l'année se déroule cet événement. Les gens ont des habits d'hiver, ce peut donc être entre novembre et février ? Bien sûr cette conclusion est liée au lecteur, plus particulièrement à la latitude sous laquelle il se trouve… Par ailleurs certains détails sur la façade de l'immeuble sont différents sur les deux premières images, et cette « incohérence » est bien sûr liée au fait que ces images sont elles-mêmes des restitutions faites selon un objectif et donc une représentation bien précis. Les données très nombreuses et en apparence très fidèles sont elles-mêmes une synthèse ; elles ne sont pas brutes.

On se trouve donc ici bien loin de la situation expérimentale telle que décrite par Claude BERNARD (voir LEGAY, 1997) et dans laquelle il s'agit de relier une cause à un effet, avec des hypothèses explicites, la mise en place d'une expérience cruciale et dont la répétition doit donner le même résultat. Dans ce cas, en effet, l'environnement non contrôlé de l'expérience n'a aucun impact sur le résultat et il n'y a pas lieu d'en donner une description.

On se trouve également loin de l'expérience planifiée de l'époque fishérienne (voir également LEGAY 1997) dans laquelle l'environnement non contrôlé peut avoir un impact dont on tient compte, en admettant qu'il peut exister une multitude de causes. Dans ce cas la rigueur du plan d'expérience permet de séparer les effets de ce qui est contrôlé et de ce qui ne l'est pas selon

des espaces orthogonaux. Les répétitions, en nombre élevé, deviennent alors nécessaires pour estimer un petit nombre de paramètres permettant de spécifier la distribution générale dont l'ensemble des données est une réalisation. La synthèse (statistique) des données est donc licite mais le plan d'expérience définit la forme de cette synthèse. On ne peut répondre qu'aux seules questions dont les réponses peuvent s'exprimer selon une fonction de cette synthèse.

Dans ces conditions d'expérience, le nombre d'acteurs est réduit, de même que la « distance » entre les données et leur(s) restitution(s). Cette distance est nulle lorsque toute différence entre jeux de données en sortie de l'expérience se traduit par des conclusions différentes. Elle est réduite lorsque la forme de la restitution est dictée par le protocole mis en place. Les données proposées par Sempé dans ses deux premières images sont une représentation décidée par lui même d'un événement « réel ». Elles relèvent d'emblée d'un tout autre contexte, dans lequel chaque lecteur va faire à son tour une sélection de données selon sa propre représentation, et tout l'intérêt de l'histoire est d'y inclure une lecture supplémentaire, selon une représentation originale… En définitive, et en utilisant des expressions scientifiquement familières, les deux premières images proposées par Sempé ne constituent pas seulement une base de données, mais une base de connaissances.

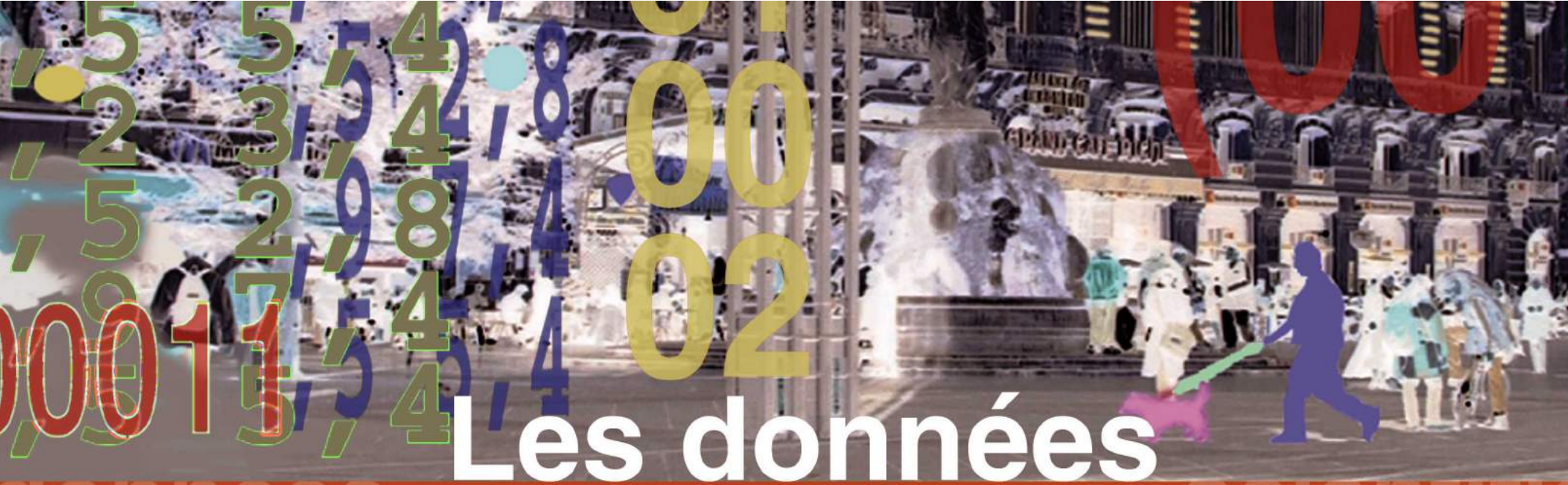
Il ne faudrait pas, sous prétexte d'un dépassement des situations expérimentales « habituelles », conclure que les méthodes et concepts associés à ces dernières devraient être mis au musée. Bien au contraire, toute synthèse, toute réponse à une question reste conditionnelle à une représentation et au protocole selon lequel les données ont été collectées. Le fait que plusieurs protocoles puissent être mis en place à partir de celui qui a conduit à fournir l'ensemble initial de données pose avec une acuité accrue le besoin de leur identification et de leur présentation rigoureuse.

Ces quelques considérations ont trouvé de nombreuses illustrations dans les exposés donnés lors de ce séminaire. Les observatoires et systèmes d'information posent largement la question des restitutions à l'attention d'acteurs différents, dont la liste ou dont les attentes font elles-mêmes l'objet d'analyses renouvelant la représentation générale des systèmes étudiés et donc imposant l'adaptation régulière des protocoles de collecte des données. L'accroissement considérable des possibilités offertes par l'informatique pour le stockage et l'accès aux données, pour leurs traitements et leurs restitutions sous des formes et des supports très variés, conduit à rechercher des moyens de repérer des événements rares dans des ensembles énormes de données, puis de procéder aux analyses de ces événements en réunissant « toutes » les données disponibles utiles pour ce faire ; ce qui conduit à les rechercher dans des réseaux de sources qu'il faut construire, selon des méthodologies qui constituent également des questions de recherche d'actualité.

Bibliographie

LEGAY, J.-M., 1997 - *L'expérience et le modèle*. *Un discours sur la méthode.* Inra éditions, Sciences en questions, 112 p.

SEMPÉ, 1962 - *Rien n'est simple*. Denoël.



Les données scientifiques

Bases de progrès des connaissances

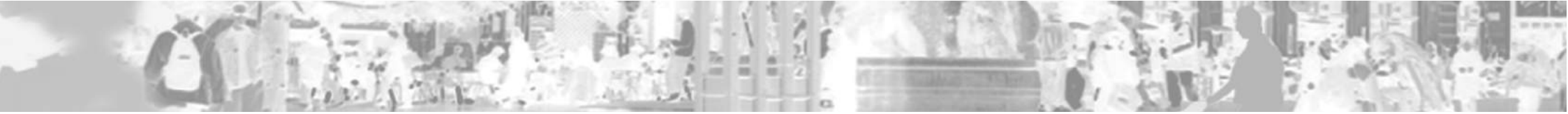
Séminaire tenu à l'IRD Ile-de France, Bondy les 4 et 5 mai 1999

Éditeurs scientifiques Jean-Michel Kornprobst, Marcel Raffy

IRD Éditions

sommaire adresses des auteurs

Sommaire



Adresse des auteurs

Jean-Luc **Boichard,** informaticien, Météo-France/Médias, BP 2102, 18, avenue E. Belin, 31401 Toulouse cedex 4.

Jeanne **Bossi**, secteur santé, CNIL, 21, rue St-Guillaume, 75007 Paris.
e-mail : jbossi@cnil.fr

Danièle **Boucon,** ingénieur CNES, 18, av. Edouard Belin, 34401 Toulouse cedex 4.

Eliane **Cubero-Castan**, informaticienne, Médias-France, BP 2102, 18, avenue Edouard Belin, 31401 Toulouse cedex 4.

Pierre **Chavance** IRD, BP 1386, Dakar, Sénégal.
e-mail : Pierre.Chavance@ird.sn

Pascal **Dallot**, assistant informatique, CEA/DAM, Analyse, surveillance, environnement, B.P. 12, 91680 Bruyères–le-Châtel.
e-mail : dallot@dase.bruyeres.cea.fr

Michel **Hoepffner**, hydrologue, IRD-Médias, BP 2102, 18, av. E. Belin, 31401 Toulouse cedex 4.
e-mail : Michel.Hoepffner@medias.cnes.fr

Claude **Huc**, ingénieur, département Valorisation et gestion des données spatiales, CNES, 18, av. Edouard Belin, 31401 Toulouse cedex 4.
e-mail : claude.huc@cnes.fr

Jean-Michel **Kornprobst**, professeur université de Nantes, vice-Président de la CS7, ISOMer, Laboratoire de chimie marine, BP 92208, 2, rue de la Houssinière, 44322 Nantes celex 3.
e-mail : jean-michel.kornprobst@wanadoo.fr

Sébastien **Lafargue,** juriste, Ifremer, Technopolis 40, 155, rue J.J. Rousseau, 92138 Issy-les-Moulineaux.
e-mail : Sebastien.lafargue@ifremer.fr

Francis **Laloë**, IRD, Halieutique et Écosystèmes Aquatiques, BP 5045, 34032 Montpellier cedex 1.
e-mail : laloe@mpl.ird.fr

Michel **Lang,** hydrologue, Cemagref, Division hydraulique, 3 bis, quai Chauveau, CP 220, 69009 Lyon cedex.
e-mail : michel.lang@cemagref.fr

Alain **Laponche**, ingénieur Sismer, Ifremer, centre de Brest, BP 70, 29280 Plouzané.

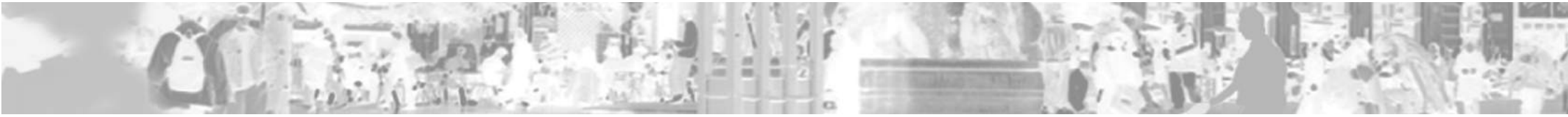
Didier **Lavoine**, ingénieur réseau, 2 bis, rue R. Le Ricollais, 44000 Nantes.

Fabrice **Lecornu**, ingénieur informaticien, Ifremer, centre de Brest, BP 70, 29280 Plouzané.
e-mail : Fabrice.Lecornu@ifremer.fr

Joseph **Le Foll**, informaticien, CEA/DSM/DAPNIA, CE Saclay, 91191 Gif-sur-Yvette cedex.
e-mail : lefoll@hep.saclay.cea.fr

François **Le Verge**, chef du service de la documentation, Ifremer, centre de Brest, BP 70, 29280 Plouzané.
e-mail : fleverge@ifremer.fr

Catherine **Mailllard**, ingénieur de recherche Ifremer, centre de Brest, BP 70, 29280 Plouzané.
e-mail : Catherine.Maillard@ifremer.fr



Bernard **Marx**, INPI, service DDI, 26 bis, rue de Saint-Pétersbourg, 75008 Paris.

Jean-Pierre **Muller**, pédologue, directeur général de l'IRD, 209-213, rue La Fayette 75480 Paris cedex 10.

Michel **Passouant**, statisticien Cirad, Campus International de Baillarguet, Bât. F, 34398 Montpellier cedex 4.
e-mail : michel.passouant@cirad.fr

Pierre **Peltre**, géographe, IRD, 32, avenue Henri-Varagnat, 93143 Bondy cedex.
e-mail : peltre@clarke.bondy.ird.fr

Marcel **Raffy**, professeur, université de Strasbourg, président de la CS7, ULP-CNRS, Parc d'innovation, 5, bd S. Brandt, 67400 Illkirch-Graffenstaden.

Armel **Rué**, ingénieur réseau, Ifremer, centre de Brest, BP 70, 29280 Plouzané.

Patrick **Séchet**, informaticien, IRD, 209-213, rue La Fayette, 75480 Paris cedex 10.
e-mail : sechet@paris.ird.fr

Éric **Simon**, directeur de recherche en informatique, Inria, BP 105, 78153 Le Chesnay.
e-mail : eric.simon@inria.fr

Françoise **Souyri**, directeur de recherche, MENRT-CSDR, 5, rue Descartes, Paris cedex 05.
e-mail : francoise.souyri@dr.education.gouv.fr

Dominique **Vuillaume**, économiste de la santé, Service du partenariat pour le Développement, Inserm, 101, rue de Tolbiac 75654 Paris cedex 13.
e-mail : vuillaume@tolbiac.inserm.fr

Philippe **Waldteufel,** climatologue, CNRS-IPSL, 10-12, avenue de l'Europe, 78140 Vélizy.
e-mail : Philippe.Waldteufel@ipsl.uvsq.fr