

Journée d'études en Droit Economie Gestion
Sociologie

**SOCIÉTÉS EN CRISE :
RÉVOLUTIONS, TRANSITIONS ET
TRANSFORMATIONS**

**Big Data, intégration et externalisation des données.
Vers une crise de l'infobésité ?**

John KINGSTON*, Marc BIDAN**

*Doctorant ** Professeur des Universités
Polytech' Nantes. Laboratoire d'Economie et de Management de Nantes (LEMNA).

Big data, intégration, externalisation : Vers une crise de l'infobésité ?

John Kingston, Doctorant Sciences de Gestion, Lemna
Marc Bidan, Professeur, Sciences de Gestion, Lemna
Communication aux journées scientifiques de l'Université de Nantes (1 juin 2018)



Corresponding author : John.Kingston@univ-nantes.fr

Résumé :

Cet article a pour objet d'identifier, de délimiter et de caractériser la « crise de l'infobésité » qui traverse les écosystèmes numériques actuellement et notamment un internet basé sur un stockage externalisé, systématique et ad vitam aeternam des données produites par les activités des particuliers et/ou des organisations. Il s'agit de s'appuyer sur la « théorie du millefeuille » (Kalika, et al, 2007) et sur les questionnements sur la soutenabilité environnementale des systèmes d'information et des technologies de l'information (Melville, 2010). Il s'agit aussi de déployer une démarche qualitative par approche Delphi (méthode des experts) auprès de dix personnalités compétentes pour souligner la réalité du problème et la non-durabilité d'un écosystème numérique – dont *le cloud computing* est la partie visible présenté comme indolore et vertueuse – qui apparait depuis dix ans comme à la fois énergivore, incapable et irresponsable. La question de la sortie de crise est alors posée sachant que les pratiques à risques sont le fait des concepteurs, des développeurs et/ou des utilisateurs privés (entreprises) et non privés (organisations, particuliers...)

Mots-clés :

Soutenabilité du web, informatique en nuage, données massives, chaine de block, bitcoin, infobésité, web énergivore

Abstract:

This article aims to identify, delimit and characterize the "infobesity crisis" that is going through digital ecosystems today and in particular an internet based on outsourced storage, systematic and ad vitam aeternam data produced by the activities of individuals and / or organizations. It is a question of relying on the "theory of the millefeuille" (Kalika, et al, 2004) and on the questions on the sustainability of the information technologies (Melville, 2010) and to deploy a qualitative approach by documentary and cases studies. to show the unsustainability of a digital ecosystem - of which cloud computing is the visible part - both energy-consuming, incapable and irresponsible knowing that risky practices are done by designers, developers and / or end users

Key-words:

Infobesity, digital ecosystems, cloud computing, green IT, big data

Introduction : De quelle « crise » s'agit-il donc ?

Le problème central des écosystèmes numériques actuels est celui de la production, du stockage et du traitement de données massives dont la question de la durée de vie et plus largement du cycle de vie n'est jamais abordée. Il apparaît dès lors que contrairement à toute autre production issue de l'activité humaine que 1) ces e-déchets n'ont pas été pensés pour être recyclés, traités et donc encore moins supprimés et que 2) cette surproduction de données in fine coûteuses, handicapantes et polluantes – à l'origine du terme d'infobésité - pose un problème de management de l'information et des systèmes d'information tout à fait complexe. Le problème est donc celui d'une crise de surproduction d'information – certaines d'origines humaines (les data) et d'autres d'origines mécaniques et algorithmiques (les big data) – qui est en train de mettre à mal la survie du système lui-même car il est presque devenu incapable d'envisager en l'état de continuer à les stocker, à les traiter, à les sécuriser et à les diffuser ad vitam aeternam au grés des requêtes des agents.

Pour aborder frontalement cette question de l'infobésité, nous avons choisi de retenir un double support théorique et une méthodologie classique (méthode Delphi aménagée dite « des experts ») qui est fréquemment déployée face à des problématiques encore émergentes et non stabilisées

Ainsi Le recours à la « théorie du millefeuille » (Kalika et al, 2007) nous permet d'appréhender la faible substituabilité des outils de communication et donc de production de données ce qui explique a contrario que les moyens de communication et de fabrication de données (de plus en plus nombreux et performants) se superposent plutôt que se substituent. De plus, les travaux de Melville (information systems innovation for environmental sustainability) sur la soutenabilité environnement et les rôles perçus des systèmes d'information (Melville, 2010) nous permettent de faire le lien entre la production de données massives et la non soutenabilité énergétiques, éthiques et organisationnelles des systèmes d'information notamment ceux liés aux e-business models des grandes entreprises. La méthode Delphi – méthode des experts – nous permet aussi de déployer une méthodologie simple et robuste auprès d'un panel d'expert de l'économie numérique qui interviennent sur cette double question « Sommes-nous entrés en situation de crise au regard de l'infobésité perceptible au sein des écosystèmes numériques et pourquoi cette implacable production ? » et « Existe t ils des sorties de crises acceptable et déployable à court terme et si oui, lesquelles ? »

Partie 1 : Quelles sont les bases théoriques reliées au concept d'infobésité

La logique théorique est double (d'une part il s'agit d'illustrer l'infobésité et de la documenter et délimiter puis d'autre part il s'agit d'en montrer les impacts notamment environnementaux et éthiques) et nous pouvons donc mobiliser deux types de travaux pour l'illustrer.

1.1 La théorie du millefeuille

D'une part la production de données non seulement n'est pas limitée par les « nouvelles » technologies de l'information mais elle est décuplée par leur superposition et notamment dans le monde professionnel où les « manageurs » sur-consomment et sur-produisent de la data (Kalika et al, 2007) sans jamais ni détruire/recycler les outils obsolètes (fax, e-mail, téléphone fixe...) ni détruire/recycler les données produites. Les travaux de kalika et al montrent que plusieurs effets sont notables dans l'explication ce processus contre intuitif de non substituabilité des outils de communication et d'information : l'effet spirale, l'effet bureaucratique, l'effet norme de communication, l'effet « besoins sociaux », l'effet implication de la direction...Ce millefeuille technologique produit donc un « gloubiboulga » de données primaires (souvent redondantes) qui elles-mêmes vont produire un « big data » de données secondaires (souvent inutiles) issues des calculs algorithmiques et de la monétisation des données par les GAFAM et autres NATU.

1.2 La théorie de la (non) soutenabilité environnementale et comportementale des TI

D'autre part, les systèmes d'information au sens large (applications, computers, datacenter, cloud computing...) ont une responsabilité environnementale directe dans le processus actuel de réchauffement climatique actuel de par leur dimension très énergivore (électricité, eau, terres rares, métaux précieux, ...) mais aussi une responsabilité comportementale indirecte dans ce même processus de par leur perception « dématérialisée » et « hors sol » qui est d'ailleurs portée par un vocabulaire tres pernicieux comme les galvaudés et erronés « virtuel », « propre », « dématérialisé », « nuage », « non carboné », « zero papier »...) ce qui in fine pose la question du rôle et de l'impact des SI sur l'environnement et de l'indispensable (mais lente) prise en compte de cette réalité (Melville, 2010).

1.3 Le monde « dématérialisé » ne le serait donc pas tant que ça

Notre communication a également pu mobiliser les travaux de plusieurs auteurs liés aux questions de (sur)production de données, du green IT, de l'infobésité et de son impact à terme sur la soutenabilité et la durabilité des écosystèmes numériques au sens large c'est-à-dire tels que nous les connaissons actuellement. Nous les avons classés ci-après de façon simplement chronologique

Corbett et Kirsh	2001	International diffusion of ISO certification
King et Lennox	2002	Exploring the locus of profitable pollution reduction
Levy et Nault	2004	Converting technology to mitigate environnemetal damage
Barrieu, Siclair et Desgagne	2006	On precautionary policies
Bohas et Mathieu	2011	Une typologie des pratiques de systemes d'information durables
Reix, fallery, Desq et Rodhain	2016	25 ans de recherches en systèmes d'information

Ces travaux en général et nos deux supports théoriques en particulier évoquent tous la possibilité d'une « *situation de crise* » liée à cette production non maîtrisée de données numériques (data) et de e-dechet (big data) dès lors, il s'agit de savoir si « *une sortie de crise* » est encore possible et comment la mettre en œuvre ?

Partie 2 : Quelles sont les données empiriques mobilisables face au concept d'infobésité ?

Cette partie est le fruit d'une recherche qualitative de type Delphi qui repose sur des « round » de rencontres et entretiens avec dix experts – cadre dirigeants, DSI, key user, IT consultants, informaticiens, CIO...- auxquels nous avons posé les deux questions introduites ci-dessus et avec lesquels nous avons pour objectif d'atteindre un consensus sur ce qu'est l'infobésité, d'où elle vient et où elle va c'est-à-dire « comment la traiter »

Il en ressort que les Systèmes d'Information de gestion (SIG) – les SI appliqués aux organisations - font face depuis une vingtaine d'années à un implacable et irréversible processus d'intégration des données, des processus et des applications au sein des entreprises. Ce processus donne lieu à une forte augmentation des volumes de données stockés et manipulés au sein des bases de données des SIG. Ce qui pose alors la question de leur stockage pérenne et sécurisé. La solution est alors souvent « dans le nuage ». L'intégration des SI est alors directement et indirectement liée à un autre phénomène majeur explicatif de cette infobésité croissante, celui de l'externalisation des données. La relation entre le processus d'intégration des données puis, lorsqu'elles sont devenues trop encombrantes et trop coûteuses à héberger, leur externalisation vers la grande nébuleuse de [l'informatique en nuage](#) est essentielle pour comprendre la logique implacable de l'infobésité.

2.1 De l'intégration des données, des processus et des applications

Ce [processus d'intégration](#) est intéressant à appréhender et à analyser à plus d'un titre. D'une part, il nous permet d'aborder la vulnérabilité des fonctions SI face à la pression d'un environnement économique de plus en plus exigeant en terme de réactivité (*time to market*) et à l'attractivité des solutions « clé en main » proposées par les éditeurs (*turnkey*). D'autre part, il nous permet d'entrevoir une séquence inéluctable déclenchée par l'intégration, accompagnée par l'externalisation et préfigurant une captation de l'information. Cette révolution n'est ni technologique ni informatique et encore moins virtuelle, elle est organisationnelle. Elle pose la question de [l'immédiateté \(real time\)](#) dans une société qui a décidé un peu rapidement de s'en accommoder sans en mesurer les coûts, les conséquences et les défis.

2.2 Du rôle des progiciels de gestion intégrés dans l'intégration puis dans l'externalisation des SI

Premièrement, nous notons une centralisation et une homogénéisation des données de gestion au sein d'une base de données qui apparaît de plus en plus comme *logiquement* unique mais qui ne l'est pas physiquement.

Deuxièmement apparaît une modularisation des applications fonctionnelles qui gravitent et inter-opèrent autour de cette base de données. Schématiquement, son architecture est celle d'une étoile avec la base de données

unique au centre et ses applications fonctionnelles (ou modules) à l'extrémité de ses branches par lesquelles circulent en temps réel et simultanément les informations.

Caricaturalement, ce système d'information intégré repose sur des outils informatiques appelés [progiciels de gestion intégrés \(PGI\)](#). Ces PGI - ou encore ERP (*Enterprise Resource Planning*) selon la terminologie anglo-saxonne dominante - sont devenus l'alpha et l'oméga des SI dédiés à la gestion des organisations.

L'ossature applicative de ces outils est l'alpha de l'intégration car elle concerne leur module, c'est-à-dire les différentes tâches ou fonctions qu'ils permettent de mener à bien simultanément et en temps réel (comptabilité, contrôle de gestion, paie, marketing, achat, stock, gestion de production...). Le centre névralgique de ces outils est l'oméga de l'intégration, car il concerne leur base de données et les différentes informations plus ou moins structurées et homogènes qu'elles stockent, historisent et organisent. Les progiciels type PGI et leur architecture modulaire à base de données pivot ([SAP, Oracle, Sage, Cegid, OpenERP...](#)) sont ainsi devenus en vingt ans l'alpha et l'oméga des SI des entreprises.

2.3 Des architectures en silos aux architectures en étoile

Ces systèmes d'information intégrés à [architecture en étoile](#) sont ceux des actuelles directions des systèmes d'information. Ils ont peu à peu remplacé les systèmes d'information hétérogènes, ceux des anciennes directions informatiques, dont les architectures étaient différentes. Elles reposaient plutôt sur une juxtaposition de silos applicatifs - composés d'une base de données avec son (ou ses) application(s) fonctionnelle(s) dédiée(s) - à l'image de tuyaux d'orgue étanches et non interopérables. Les Directions Informatique et leurs programmes conçus, développés et de plus en plus difficilement « maintenus » en interne ont, en vingt ans, laissé la place aux Directions des Systèmes d'Information. Cette évolution sémantique DI/DSI accompagne l'arrivée de solutions intégrées vendues et maintenues « clé en main » par les éditeurs.

Elle signifie surtout pour les nouvelles DSI la perte irréversible de compétences techniques et [leur remplacement par des compétences et responsabilités managériales](#) (administration, gouvernance, contrôle) afin de faire face au pouvoir de négociation des éditeurs, intégrateurs, hébergeurs et autres info-géants. Les modalités et acteurs de l'externalisation sont certes nombreux entre [outsourcing, offshoring, sous-traitance](#) et centre de services partagés

2.5 Du caractère implacable de ce processus d'intégration des SI

La première famille d'explications est économique. L'environnement d'affaires est de plus en plus [exigeant, imprévisible et complexe](#). Les clients, les fournisseurs, l'état, les banques, les actionnaires, les salariés veulent tout et tout de suite. Pour leur répondre, l'entreprise n'a pas le choix et elle doit rassembler ses forces – ses données, ses informations, ses réseaux et ses applications – en un dispositif unique, paré à la manœuvre et hyper(ré)actif.

La seconde famille d'explications est technique. Les progrès de l'informatique depuis une trentaine d'années – bien avant l'avènement de l'Internet – ont permis de rendre attrayantes et accessibles des solutions modulaires « clé en main ». Les applications fonctionnelles éparpillées dans les méandres des anciens postes de travail informatiques ont laissé la place à une solution unique plus intuitive, plus puissante et plus ergonomique accessible via l'environnement numérique de travail. Cette solution intégrée, trop complexe pour être conçue et développée en interne, est vendue, paramétrée, testée et déployée par un [intégrateur qui peut aussi en être l'éditeur voire l'hébergeur](#). Ainsi, comme dans un rêve, comme dans les nuages, comme dans un « serious game » dont tout le monde sortirait vainqueur... l'informatique offre aux entreprises les systèmes intégrés que la compétition économique exige.

2.6 De l'intégration à l'externalisation ou le syndrome d'une « extégration » déresponsabilisante

La logique de l'intégration du SI est intrinsèquement à l'origine de son externalisation. D'un côté, plus les volumes de données et de contenus manipulés vont croître et plus les opérations de maintenance vont se complexifier, plus la DSI sera obligée de s'en remettre à l'extérieur et plus les acteurs extérieurs (intégrateurs, éditeurs, info-géneurs, hébergeurs) vont devenir incontournables. La logique veut que l'externalisation partielle des données puis des applications puis des services se conclue par une externalisation totale du SI. Cette logique d'extégration est problématique, car elle se traduit par une perte de compétence des DSI, par des [coûts de réversibilité très élevés](#) et par un pouvoir accru des opérateurs extérieurs.

Cette logique est inhérente au *cloud computing* qui propose des solutions à la fois SaaS, PaaS et IaaS (X as a service) et qui devient – même si ses performances applicatives et économiques sont officiellement au rendez-vous – bien plus qu'un simple hébergeur du SI mais bien son principal opérateur.

Ce phénomène en trois temps (intégration, externalisation, captation) reste implacable et risqué, il explique une grande partie de l'infobésité actuelle due aux entreprises. Celle qui est due aux particuliers repose plutôt sur le modèle d'affaire dominant du web qui postule la « monétisation des données » et qui impose aux GAFAM de collecter « gratuitement » une masse toujours plus grande de données pour les revendre ensuite comme matière brute ou travaillée via leur algorithme. Dans les deux cas, la production de data et de big data est implacable et semble irréversible

Conclusion sur le probable non-réversibilité du modèle dominant du data-management

La réversibilité – c'est-à-dire le retour à un système frugal et local – semble complexe voire impossible. Qu'il s'agisse de l'entreprise, notamment de taille modeste, ayant externalisé son SI vers un opérateur du cloud, elle sera bien seule en cas de problème dans le nuage (faille de sécurité, faillite de l'hébergeur, intrusion, perte de données, accessibilité, tarification, contrat...). Qu'il s'agisse aussi du particulier, notamment usager (sans droit) et non pas client (avec droit), confronté à la captation de ses données personnelles de navigation et à leur monétisation ensuite, il semble délicat à court terme d'envisager un web-modèle basé sur l'achat des données

personnelles par les GAFAM et autres NATU. C'est néanmoins l'idée principale – et elle n'est pas directement liée à l'arrivée du RGPD – qui se cache derrière un écosystème numérique qui serait devenu vertueux, respectueux et frugal et qui aurait le courage de bannir la folie d'un streaming très énergivore – car perçu comme non énergivore justement - pour revenir au « bon vieux cd-rom » avec une perception inverse !

Gageons que la *sortie de crise* face à cette infobésité serait 1) de prendre conscience de l'immense supercherie intellectuelle qu'est la dématérialisation de l'informatique en nuage qui n'est ni dématérialisée ni dans les nuages bien sûr (l'énergie électrique nécessaire annuellement au bon fonctionnement du web est équivalente à celle d'un pays comme la Suisse) et 2) de revenir à un usage plus responsable et raisonnable dans nos pratiques individuelles et collectives elles-mêmes et – plus tard ? - dans celle des entreprises.

Finalement, si nous avons un peu d'égard pour l'électricité famariveuse et les millions de litre d'eau qui sont engloutis dans l'exploitation des méga Datacenter de Google et autres méga opérateurs du cloud (Alibaba, Facebook, Amazon, etc....) nous nous poserions parfois de telles question : Est-ce que la vidéo de « notre gentil petit chat qui danse » que nous venons de poster sur YouTube est vraiment digne de rester ad vitam aeternam sur les étagères du web ? Et nous penserions ainsi un peu à l'impact bien réel de nos e-déchets et de leur avenir post mortem tels de véritables « datazombie » !

Bibliographie

Bidan, M. (2016) <https://theconversation.com/limplacable-integration-des-systemes-dinformation-et-son-impact-sur-lexternalisation-52089>

Bohas A et Mathieu A, (2011) Une typologie des pratiques systèmes d'information durable, AIM2011, Actes du Colloque de l'AIM à St Denis de La réunion

Desq F, Fallery B, Rodhain F et Reix R, (2016), 25 ans de recherche en Systèmes d'information, Revue Systèmes d'Information et Management

Kalika M, Boukef Charki N et Isaac H (2007), La théorie du millefeuille et l'usage des TIC dans l'entreprise, Revue Française de Gestion, p117-129

Melville, N (2010), Information Systems innovation for environmental sustainability, MISQ https://misq.org/downloadable/download/linkSample/link_id/838/

Annexe 1 : quelques informations sur la méthodologie Delphi déployée durant hivers et printemps 2018

Les deux questions (à deux niveaux) posées aux experts

Q1	« Sommes-nous entrés en situation de crise au regard de l'infobésité perceptible au sein des écosystèmes numériques et pourquoi cette implacable production ? »t
Q2	« Existe t ils des sorties de crises acceptable et déployable à court terme et si oui , lesquelles ? »

Les lieux et dates des trois rounds d'entretiens et de rencontres avec les experts

Round 1	Polytech Nantes (salle réunion)	13 Fev 2018 – 13H00
Round 2	IUT de Nantes (Salle du conseil)	5 Mars 2018 – 13H30
Round 3	Café Flessel (premier étage)	27 Mars 2018 – 18H00

Les métiers/activités des experts mobilisés sur les trois rounds

1. IT Consultant Nantes	3. DSI secteur logistique fleurs	5. Informaticiens grands comptes (sopra groupe)	7. DG start up (serious game)	9. Cadre commercial secteur big data
2. IT Consultant Paris- Asnières	4. DSI secteur hospitalier	6. Informaticiens secteurs bancaire (crédit mutuel arkéa)	8. DG adjoint start up (secteur agro(alimentaire))	10. Cadre achat secteur supply chain