

Managing sensitive data for public information: a use cases review of the French water information system

Alexandre Liccardi*, Laurent Coudercy, Laurent Breton, François Hisselⁱ, Pierre Lagardeⁱⁱ

*alexandre.liccardi@afbiodiversite.fr

ⁱ French Biodiversity Agency, Vincennes, ⁱⁱBRGM (French Geological Survey), Orléans, FRANCE

Abstract

Access to environmental information is a major momentum for public policies. In many ways, data publication is known as a key factor for topics appropriation by stakeholders, and can be used as an effective lever towards public opinions. Furthermore, collecting data and guaranteeing their public access are legal obligations for EU Member States since the early 2000's.

This study compares three institutional and technical processes: (i) river boat fishing monitoring and related quotas, (ii) commercial and personal water abstraction declaration facilities and (iii) the register of plant protection products sales. Two main axes are chosen to highlight differences in process and policy efficiency, as far as data are concerned: level of technical complexity and maturity of processes (for instance, *acceptance by stakeholder*).

Main conclusions echo with others data policies reviews: technical feasibility doesn't seem to be the main barrier in studied cases, whereas knowledge of purely functional and organizational trade aspects (established know-how and expertise) is a key element for success. Also, open data strategies should still prevail in public services, where partitioning data, informatics, scientific, organizational and legal task forces is a great concern. Beyond this coordination issues, because environmental phenomena don't know frontiers nor boundaries, international frameworks and initiatives should prevail on local engineering, regardless of local data strategies, as far as national secrets and policies are heeded.

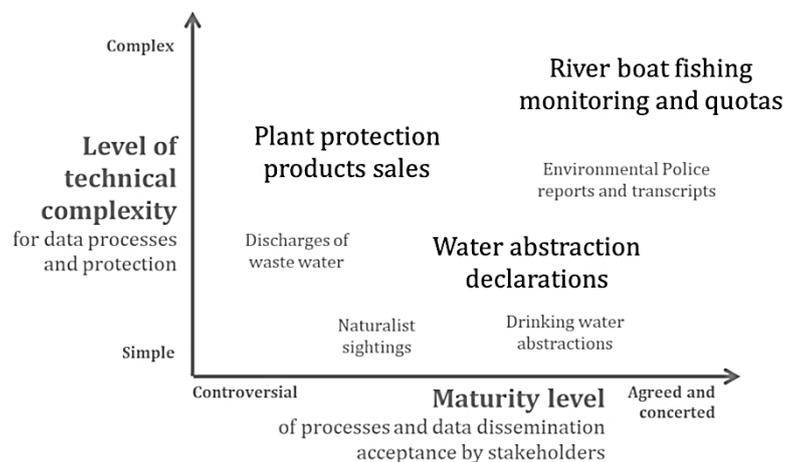
Keywords – data policy, policy acceptance, open data strategies, environmental public data, data dissemination

Introduction

The French water information system (F-WIS) is a multipartner framework led by the French Ministry of the Environment and coordinated by the French Biodiversity Agency (AFB). F-WIS meets legal obligations set by national laws and EC policies (firstly the *Water Framework Directive*) in public information on a broad scope (bathymetry, physicochemistry, hydromorphology, biology, discharge and low flows, real-time temperatures, economical issues).

This institutional and technical process initiated in the middle of the 90s, with the publication of methodological baselines and data repositories (See Lalement & Lagarde, 2005; Liccardi et al, 2017). Since then, F-WIS works on the compliance of public information policies (especially public access to environmental information and data reuse directives) and trade secrets, statistical secrecy and more widely a growing legal corpus involving economical and personal data into environmental and public decision at EC and national levels (See *SNDE*). This discussion paper focuses on a comparative study between three data processing use cases, with various levels in criticality and complexity, but sharing a strong environmental effect, a substantial political weight and great expectations from professionals, universities and general audience. River boat fishing monitoring and related quotas, commercial and personal water abstraction declaration facilities and the register of plant protection products sales (especially gardens and agricultural pesticides) are three enlightening use cases discussed below in chronological order (see Figure 1).

Figure 1. Overall use cases presentation, according to their technical complexity and the acceptance of related data dissemination processes.



A use cases review about managing sensitive data for public information

This study can be seen as a feedback by public services, in charge of databases engineering and coordination. Its main goal is to enlighten the success factors in environmental data collection and dissemination projects, in order to emit recommendations (see Figure 2 for detailed statistics).

Two main analysis axes have been chosen: as data is always a matter of data processing and computations (Sansone et al., 2018), the first axis is about level of technical complexity. This idea values a level of “machine and human technical effort” needed to achieve data preparation and dissemination. While this effort is commonly described with the volume of data managed in the databanks (Baru et al., 2013), we introduced different criteria, such as IT architecture needed to access data (data availability and update rate) and the weight of the data information itself - for instance, a complete description of a fishing trip, with tools, equipment and fish catches, involves more information than a single taxa located observation. Regarding data sensitivity, this axis can be seen as the required effort to reduce the risk of disclosure of confidential data, while maintaining a high availability and a high quality of environmental information.

The second axis considers the level of maturity of the coordination processes and the acceptance of data dissemination by stakeholders (see also, in private sector: Shah et al., 2012). Data gathered in this review are the cornerstone of vast economic interests and carry highly sensitive business information, so the motivation of their collection

can oppose human sectorial activities, and environmental friendly intents. These conflicts can lead to local adjustments to the Community recommendations, which can be relevant regarding trade secret, statistical secrecy and personal data protection (Tsoukala et al., 2015; Borges et al., 2018; GFIRA, 2015).

Boat fishing monitoring on French rivers¹: expertise and support to the profession to underpin an ambitious policy.

Community fisheries are highly monitored by quotas, allocated by the European Commission in order to preserve species and their stock. The CITES is strongly committed in this momentum (Washington Convention, 2009), and the EU Regulation *Data Collection Framework* builds a shared database for both scientists and administrators. In France, the 1984 law about fisheries provides tools for quotas compliance. Recreational and professional fishing is also subject to a declaration. With selling prices hitting 400 €/kg for the lever (eel juvenile), poaching impact on local economy is estimated at over millions euros (OCLAESP source, see also CITES, 2018), for local consumption or south-east Asia exports. Since the choice of fishing location and paces can be regarded as business strategies and as highly valued information in a competitive market (Jousselme, 2016; Feekings et al. 2017), trade secret is firmly propounded by stakeholders. Nonetheless, this economic stake goes along with the necessity of the implementation of fishing management plans based on better ichthyological knowledge, for which transparency and data dissemination are preconditions.

On a technical level, the data collected reveal an important inherent complexity. Behind this register,

Figure 2. Statistical summary of the three databases studied in the review, with data volume indications and main management elements.

u.c. : *under conditions*, such as specific demands ; exp. : *expected*.

	Available time span	Record's count	Number of quoted stakeholders	Major regulatory text(s)	Spatial resolution	Open data ?	API ?
Boat fishing monitoring SNPE	2003 - 2018 online reports, data series available on demand	3 844 251 fishing trips	10 173 fishermen	1984 National Law 2009 Washington Convention	Fish river unit (< 50 km)	u.c.	u.c.
Water abstraction databank bnpe.eaufrance.fr	2008 - 2016 online open data	623 009 abstraction points declaration	116 496 installations of water abstraction	1992 National Law 2000 Water Framework Directive	Coordinates of installation and/or municipality	Yes	Yes
Plant protection products sales registers BNV d (data.eaufrance.fr)	2010 - 2018 (exp.) online open data	14 197 075 sale transactions	2 633 plant protection suppliers	2000 Water Framework Directive 2009 European regulation	Sale point Municipality	No (exp.)	No

¹ See more at snpe.onema.fr.

an almost exhaustive inventory of fishing equipment and privileged geographical areas could be built, with departmental accuracy. The identification of professional agents, sometimes few in number on this scale, is not impossible for their possible malicious competitors. A specific effort of anonymization is therefore necessary, and is particularly difficult to implement because of the movements of fishermen on several management units and the concentration over time of fishing activities, which follow the behavior of migratory fish flows.

Public teams from IFREMER, INRA, AFB and local state services are involved in collecting and analyzing these daily data, using both historical and long-term registers of fish trade (before 1980) and professional networking work. Declarations and reports are gathered on *SNPE*, a common database providing geographical and temporal aggregated information for CE and national obligations. More recently, in order to provide real-time information about endangered species, dynamic aggregation solutions have been produced: handled using APIs, these algorithms provides statistics using combination functions along many business axes (tools and devices, time of the trip, position in the hydrological connected network), using the most precise aggregate allowed by trade secret and statistical secrecy.

The water abstraction databank² and its legal framework allowing the share of nearly raw data.

The quantitative aspects in water resource management are crucial to both its social and economic availability and its use by natural environments (particularly regarding ecological continuity), and is a key topic in the *Water Framework Directive* and its French transpositions. French state local services, River Basin agencies, AFB and BRGM offer with this databank a unique access point to abstraction related data, according to its economic use (domestic, agricultural, industrial including energy). French law creates declaration obligations, according to a precise framework (laws and decrees from 1992 to 2007), and defines threshold and minimal required information dataset, commonly used in the royalty calculation for water consumption. Citizens can also mobilize hundreds of thousands of yearly quantitative and georeferenced declarations, involving natural or legal persons or entities (see also Pasquier et al, 2017).

Along with this service, drinking water abstraction areas by communal services are subject to special protections in line with homeland security,

² See more and demonstrator at bnpe.eaufrance.fr.

particularly regarding drinking water and terrorist risk. In the field, the real location can't be hidden from the public, especially as their protected area is wide. In practice, the municipality centroid is transmitted: individually, each data can be easily checked and specified, but the stated objective is to make difficult the massive data analyses at the most detailed level. Data are also blurred when the scope is large, but citizen can locally point to each site using detailed urban plans. This first adjustment intends to ensure that public health and national security secrets are not violated.

The second adjustment concerns tax secrecy: like in the previous use case, the primary reason for collecting the data is the calculation of a fee. The nature of the information collected is financial, and allows, through a simple estimate, to trace the activity of the withdrawers (the tax base is set by a national decree of 2011³). In line with Community law on the dissemination of environmental information, the tax reporting framework has been modified to allow, at a very least, the communication of information at the level of the water abstraction installation, while keeping the identity of the operator secret. The National Finance Organic Law has been amended since 2013⁴.

The dissemination of water abstraction data benefits from a profound technical history, supported by both technical skills and regulatory obligations resulting from the logic of equipment and spatial planning (Richter et al., 2003; Vanneville & Werner, 2012). The social priorities related to this information directly affect public health, and refer to clearly identifiable topographical features, generating broad civic interest and promoting public achievements for greater transparency of stakeholder actions. As a result, although this database is the only one that is relevant to public safety in addition to economic aspects, it is also the only one that nowadays has complete open data engineering. However, this mission must be durably supported by a network of local and national public actors who ensure that the regulatory texts are updated and respected (Werner & O'Doherty, 2012).

Plant protection products sales registers⁵: a complex and controversial matter.

Environmental chemical pollution, and especially pesticides, is a major societal concern relayed by numerous international studies (see French

³ legifrance.gouv.fr, reference DEVL1132666A.

⁴ legifrance.gouv.fr, reference EFIX1327237L.

⁵ More info at bnvd.ineris.fr.

Environment Ministry bibliography⁶) and by large audience reports in France (Valo, *Le Monde*, 2017). In the Community, the subject of pesticides is a highly relevant one with the various authorizations or bans on the use of iconic glyphosate-based products (failure of the European vote in November 2017, in the absence of an agreement, authorizations are based on national law). In France, the implementation in January 2019 of a 2014 Law prohibits the use of non-organic plant protection products, with a significant economic and social impact on distributors and individuals.

The most relevant information source is the national register for the declaration of plant protection products sales (*BNVD*) and is administrated by AFB, INERIS and Water Basin agencies since 2007. This register has originally been built for calculation of the water diffuse pollution fee and provides the yearly declaration of each sale transaction, at individual level, for professional and private use, from major distributors or specialty stores. For each product and substance, and for each transaction, the volume sold and the buyer zip code are known. Each year, several millions of transactions are registered.

While the complexity of the information contained is mainly linked to the evolution of the data repositories of substances that must be declared for sale, the geographical scale poses an utmost functional problem and requires data preparation before dissemination. At a national level, data can be aggregated by pollutant type, as well as by distributor. Due to the fiscal and economic secrecy involved, information on distributors is not made available, although Community and international law is a motivating factor (Water Framework Directive, 2000; Kiev Protocol, 2003).

At a local level, according to stakeholders, giving access to pesticide use data at the municipal level would make it easy to identify the practices of professional actors or large private properties, which could be unexpectedly put into perspective with the environmental impact of the substances concerned. These approaches could raise suspicion of accusations on the ground, and worsen existing conflicts. On the other hand, the non-release of environmental pressure data at the finest scale available is a major obstacle to serious scientific studies and, more generally, to information for the general public (Weyers et al., 2004, Pavlis et al., 2010). This momentum has been pursued by various legal acts before national authorities and led to a European judgment in 2019, which

acknowledges the need for public information on pollutant emissions⁷.

Whereas the French context and EC frameworks tend to a better information disposal and transparency, the process is not fully acknowledged by a large part of stakeholder, especially from agricultural sector (Delaunay & al, CGEDD, 2017). The French administration has stated for the publication of these data many times (opinions from the *CADA*, the national authority for administrative documents access; Biodiversity national plan; French commitments regarding open government). However, since a public load from professionals invoke trade secret and statistical secrecy, Environment and Agriculture Ministry reconsider their position: currently online available data are aggregated at provincial level and cannot be used for the identification of economic players. But social pressure, helped by smaller stakeholder such as consultants, could lead to a wider data opening and to legal disputes, aiming at current restrictions.

Recommendations based on feedback, toward open data strategies in environmental public policies

This review contributes to a point already made in other articles (for example: Toots et al., 2017, Kassen, 2018), but broadens its scope: the design and technical implementation of architecture to ensure the efficient and up-to-date provision of massive data required for environmental information and science analysis is an achievable IT objective but requires a real knowledge of functional mechanisms and significant support from organizations in charge of policy implementation.

This three cases study not only describes a wide variety in data policies concerning environmental issues, it emphasizes the weight of the care in the deployment of technical solutions and of the legal interpretation by public services. In the end, the data manager ability to combine data processes and acceptance by stakeholder is a key factor, and the responsibility to quote economic players in view of their environmental impact still remains on national and local administrations.

In the cases studied in particular, the main obstacles are the existence of strong atomized economic interests and the representation of professional organizations formed in lobbies (agricultural and distribution sector in particular), and an unclear message from the public authorities at European and national level (lack of agreements or reversal of environmental protection positions). Otherwise, the

⁶ <https://www.ecologique-solidaire.gouv.fr/lutte-contre-pollutions-leau>

⁷ curia.europa.eu, reference ECLI:EU:T:2019:142.

provision of information to the various audiences is a catalyst for debate, a revealing indicator of social and administrative functioning, and a driving force for citizen involvement (in echo with Misucara & Viscusi, 2016; Janssen et al., 2017).

Clearly, this study highlights the need to decompartmentalize technical, functional and administrative logics in order to give a broader insight into public information strategies by community states. Beyond these general considerations, public services already have strategies for the deployment of their workforces, which must be considered jointly in order to reserve, on the one hand, the development energy for socially known and accepted projects, and on the other hand, time for dialogue and political attention on issues with a high environmental impact.

Concerning open data and environmental issues, new use cases emerge and a new framework is being built: environmental Police reports are used in artificial intelligence applications to maximize the efficiency of AFB officers (Etalab project, 2019), and naturalist sightings databases and surveys are more and more supported by crowdsourcing, gathering millions of daily collected data with the name of the observer (see gbif.org for further information).

These recent examples underline the need for technical, functional and administrative authorities to investigate synergies, such as ethics (when personal data allow the identification of environmental crime), data ownership not only as an observer but also as an observer (status of the amateur naturalist, or environmental inspector) and the provision of algorithms to identify sensitive elements within the meaning of environmental regulations (protected species whose positions must be fuzzy, details of police algorithms not to be reproduced).

These last impulses take also place at a world-wide scale, where regulations and legal obligations previously listed are not relevant (and the same goes with data, see Poisot et al., 2019). The supply chains for products derived from the exploitation of endangered species or, more broadly, poaching, the migration of living species and the associated physical phenomena are not limited either to national or Community borders. In the same way, the savings to be achieved by pooling technical developments between countries, on open source chains for example, are significant but still not fully adopted by administrations (Petrov & Obwegeser, 2018), even if the Community effort around the directives (Water Framework Directive in particular) has focused on the use of common tools. This review tends to demonstrate that the main efforts to be made must be on implementing and

adapting common tools to local management policies, by implementing engineering and data dissemination strategies guided by the coordination of the professions and the balanced measurement of social issues versus scientific and informative weight.

Acknowledgements

The authors would like to thank Céline Nowak and Caroline Pénil (AFB) for their insightful comments and suggestions. This paper writing wouldn't have been possible without the previous theoretical work and the long-term strategy built by the FWIS contributors, especially René Lalement (AFB).

References

Conference proceedings

Liccardi, A., & Coudercy, L., & Dembski, S., & Mauclerc, A. (2017). An open source strategy for public online algorithms and data services: the French water information system experience, Data For Policy. Zenodo. doi: 10.5281/zenodo.1025783.

Government participation

Lalement, R., & Lagarde, P. (2005). Architecture du Système d'Information sur l'Eau. Livre vert. Ministère de l'écologie et du développement durable Printing services, 36 p. La Défense, France

Regulatory text

SNDE: National Water Data Scheme <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000022734282>

Journal article

Sansone, S.-A., & Cruse, P., & Thorley, M. High-quality science requires high-quality open data infrastructure (2018). Scientific Data 5:180027 doi: 10.1038/sdata.2017.27.

Journal article

Baru, C., & Bhandarkar M., & Raghunath, N., & Poess, M., & Rabl, T. (2013). Benchmarking big data systems and the bigdata top100 list. Big Data Vol.1 N°1, 10.1089/big.2013.1509

Journal article

Shah, S., & Horne, A., & Capellá, J. (2012) Good data won't guarantee good decisions. Harvard Business Review.

Technical report

Tsoukala, V., & Angelaki, M., & Kalaitzi, V., & Wessels, B., & Price, L., & Smallwood, R., & Taylor, M., & Linde, P., & Sondervan, J., & Reilly, S., & Noorman, M., & Wyatt, S., & Bigagli, L., & Finn, R., & Sveinsdottir, T., & Wadhwa, K. (2015). Policy guidelines for open access and data dissemination and preservation, for European Commission: RECODE - Policy RECommendations for Open Access to Research Data in Europe (321463). doi: 10.5281/zenodo.1297497.

Technical report

Borges, L., & Nielsen, K., & Frangoudes, K., & Armstrong, C., & Borit, M. (2018). Conflicts and trade-

offs in implementing the Common Fisheries Policy (CFP) discard policy, for European Commission: DiscardLess - Strategies for the gradual elimination of discards in European fisheries (633680). doi: 10.5281/zenodo.1238588.

Technical report

German Federal Institute for Risk Assessment. (2015). EU Commission discusses critical issues in the assessment of the active substances in plant protection products. doi: 10.5281/zenodo.581036.

Technical report

CITES 2018: Updated information on illegal trade in European Eel, Spain submission as Chair of the Standing Committee intersessional working group on European eel (*Anguilla anguilla*).

<https://cites.org/sites/default/files/eng/com/sc/70/Inf/E-SC70-Inf-45.pdf>

Conference proceedings

Jousselle, A.-L., & Camossi, E., & Hadzagic, M., & Ray, C., & Claramunt, C., & Reardon, E., & Ilteris, M. (2016). A Fishing Monitoring Use Case in Support to Collaborative Research. In Maritime Knowledge Discovery and Anomaly Detection Workshop Proceedings (pp. 57–61). Publication Office. doi: 10.5281/zenodo.322378.

Technical report

Feeckings, J., & Savina, M., & Ulrich, C. (2017). Review on the current effect of discarding on fish stocks and marine ecosystems; on data, knowledge and models of discarding; and identification of knowledge gaps in all case studies, for European Commission: DiscardLess - Strategies for the gradual elimination of discards in European fisheries (633680). doi: 10.5281/zenodo.1204309.

Technical report

Pasquier, J.-L., & Venuat-Budon, M., & Aouir, C., & Serre, W. (2017). Les prélèvements d'eau douce en France : les grands usages en 2013 et leur évolution depuis 20 ans. Service de l'observation et des statistiques, national press, 26 p.

<https://www.eaufrance.fr/publications/les-prelevements-deau-douce-en-france-les-grands-usages-en-2013-et-leur-evolution>

Journal article

Richter, B. D. & Mathews, R., Harrison, D. L., & Wigington (2003). *Ecological Applications*, 13(1), 2003, pp. 206-224. Wiley. doi: 10.2307/3099960.

Technical report

Vanneuville, W., & and Werner, B. (2012). Water resources in Europe in the context of vulnerability EEA 2012 state of water assessment. EEA, Copenhagen, 2012. doi: 10.2800/65298

Technical report

Werner, B., & O'Doherty J. J. (2012). European waters — current status and future challenges, synthesis. EEA, Copenhagen, 2012. doi: 10.2800/63931

Journal article (large audience)

Valo, Martine. (2017) La qualité de l'eau se dégrade encore en France. *Le Monde*. https://www.lemonde.fr/pollution/article/2017/10/17/la-qualite-de-l-eau-se-degrade-encore-en-france_5201907_1652666.html

Journal article

Pavlis, M., & Cummins, E., & McDonnell, K. (2010). Groundwater Vulnerability Assessment of Plant Protection Products: A Review, Human and Ecological Risk Assessment: An International Journal, 16:3, 621-650. doi: 10.1080/10807031003788881

Journal article

Weyers, A., & Sokull-Klüttgen, B., & Knacker T., & Martin S., & Van Gestel C. A. M. (2004). Use of Terrestrial Model Ecosystem Data in Environmental Risk Assessment for Industrial Chemicals, Biocides and Plant Protection Products in the EU. *Ecotoxicology*, 2004/02 Vol.13,Iss.1-2. doi: 10.1023/b:ectx.0000012412.44625.69

Technical report

Delaunay, A., & Mir, C., & Marty-Chastan, C., & Rance, E., & Guériaux, D., & Tessier, R. (2017). Utilisation des produits pharmaceutiques. CGEDD, IGAS, national press. 94p.

http://cgedd.documentation.developpement-durable.gouv.fr/documents/cgedd/011624-01_rapport.pdf

Journal article

Toots, M., & McBride, K., & Kalvet, T., & Krimmer, R. (2017). Open Data as Enabler of Public Service Co-creation: Exploring the Drivers and Barriers. In Proceedings of the 2017 International Conference for E-Democracy and Open Government (CeDEM 2017) (102-112). Krems, Austria: IEEE Computer Society. doi:10.1109/CeDEM.2017.12

Journal article

Kassen, M. (2018). Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. *Aslib Journal of Information Management*, Vol. 70 Issue: 5, pp.518-537. doi: 10.1108/AJIM-11-2017-0250.

Journal article

Misuraca G., & Viscusi G. (2016). Is Open Data Enough? E-Governance Challenges for Open Government. *International Journal of Electronic Government Research (IJEGR)*10(1), 17p. doi:10.4018/ijegr.2014010102

Journal article

Janssen, M., & Matheus, & R., Longo, J., & Weerakkody, V. (2017) Transparency-bydesign as a foundation for open government, *Transforming Government: People, Process and Policy*, (11/1): pp.2-8. doi: 10.1108/TG-02-2017-0015

Journal article

Poisot, T., & Bruneau, A., & Gonzalez, A., & Gravel, D., & Peres-Neto, P. (2019). Ecological data should not be so hard to find and reuse. *Trends in Ecology & Evolution*, in press, doi: 10.1016/j.tree.2019.04.005.

Journal article

Petrov, D., & Obwegeser, N. (2018). Adoption Barriers of Open-Source. Software: A Systematic Review. SSRN: <https://ssrn.com/abstract=3138085>