

Données de la recherche en SHS :
un modèle de description pour l'autogestion
dans le cadre du projet « Didomena », entrepôt de données de la recherche de l'EHESS

**Overseas Korean Studies Librarian Workshop,
Les 15-19 octobre 2018, Bibliothèque nationale de Corée**

Eun-joo Carré-Na

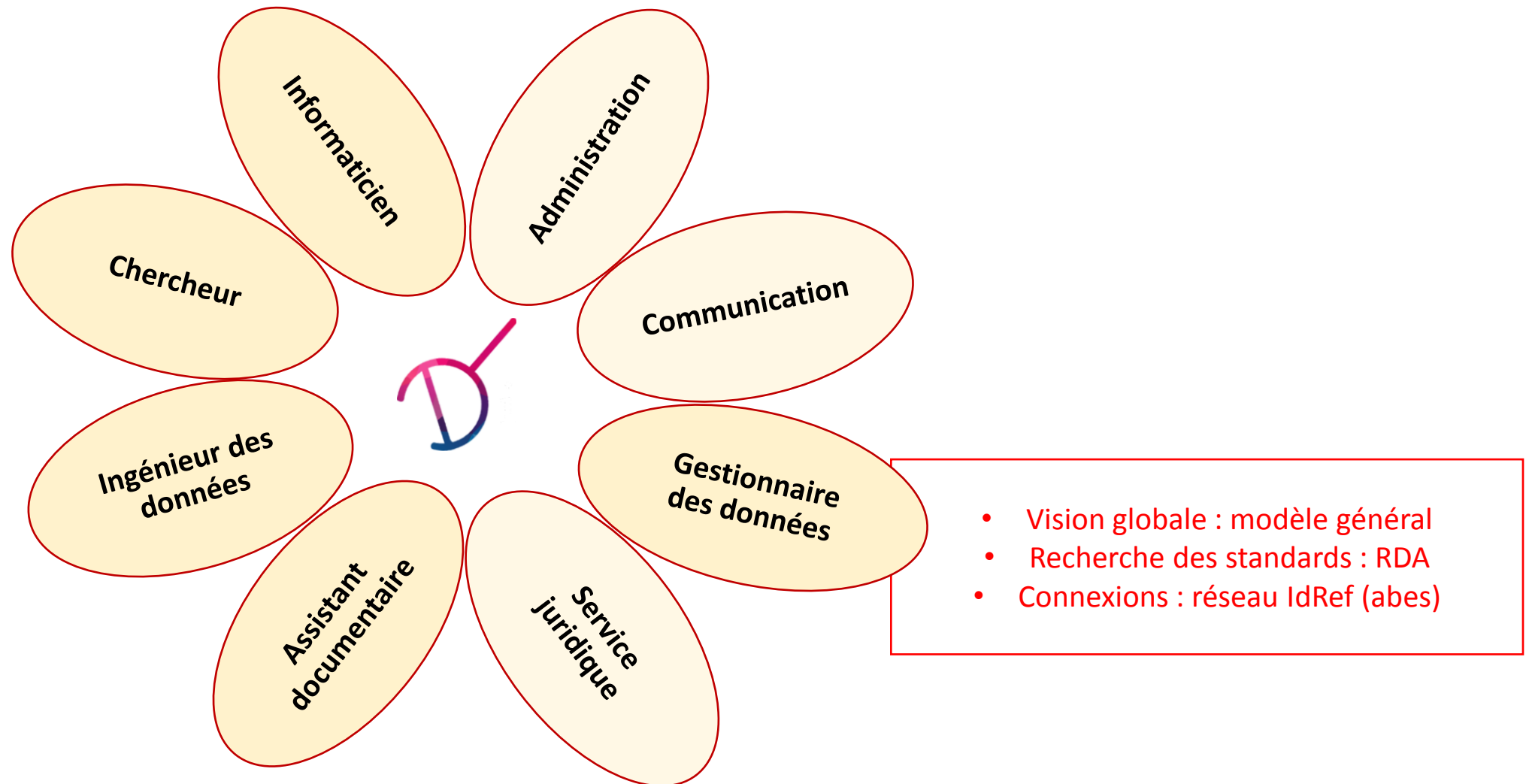
Centre de Recherches sur la Corée (CRC), EHESS

UMR 8173 Chine, Corée, Japon, (CCJ) CNRS

Sommaires

- Point de vue de gestionnaire des données
- « Données de la recherche » : définition et critères
- « Données de la recherche » : type de production
- « Données de la recherche » : cycle de vie
- Services et métiers
- Différents aspects de l'objet de description
- Modèle de données de la recherche pour métadonnées
- « Données de la recherche » : définition avec RDA
- Champs de métadonnée et expression contrôlée (RDA)
- « Types de ressource » vs « Types des composants »
- Configuration et caractéristiques

Point de vue de gestionnaire des données



« Données de la recherche » : définition et critères

- « ... les données de la recherche sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par **la communauté scientifique** comme nécessaires pour **valider les résultats** de la recherche » (OCDE, 2007)
- 4 critères pour être : « données de la recherche » (Joachim Schöpfel, et al., 2017)
 - « ... le concept des données comme matériel brut dans quatre directions :
 - **l'enregistrement comme préalable** – d'une certaine façon, il s'agit de l'équivalent de la fixation matérielle d'une idée sur un support comme préalable de la reconnaissance et protection d'une œuvre de l'esprit ;
 - **la nature factuelle**, malgré la grande diversité des données – mais comme la Royal Society (2012) précise, il peut s'agir d'une nature factuelle « supposée » quand elle définit les données de la recherche comme « qualitative or quantitative statements or numbers that are (or assumed to be) factual » ;
 - **le lien avec la communauté scientifique** – la définition de l'OMB conteste l'idée d'un concept absolu et établit le lien avec les chercheurs eux-mêmes, au sens d'un minimum de pratiques, valeurs, méthodes, outils et concepts partagés ; en d'autres termes, une donnée est ce qui est accepté par un groupe de chercheurs (« communauté ») comme telle (consensus) ;
 - et **la finalité** – les données de la recherche ont une fonction, jouent un rôle dans le processus scientifique, en particulier pour la validation des hypothèses et résultats. »

Types de travail - production

(selon le type de production de données de la recherche, avec des « acte », « objet », et « contexte »)

Collecter

Par copie

Numérisation ou copie des numériques
(**objet** de **document** numérisé ou numérique par scan, saisie manuelle, photographie ; **contexte d'acquisition** et **localisation**)

- Archive publique (document administration, médical, etc.)
- Archive privée (correspondance, journal intime, etc.)
- Non-publié (manuscrit, littérature grise, épreuve, etc.)
- Publié (livre, presse, multimédias (site-web, blog, sns, etc.)

Par capture

Capture du monde réel
(études de terrain : enregistrement, observation)

- **Objet** : réel (terrain, phénomène, personnes, société, etc.)
- **Contexte** : réel

Par enquête

Enquête
(interview, questionnaire en ligne, sondage par téléphone)

- **Objet** : Questionnaire et Réponse
- **Contexte** : prédéfini et/ou réel

Par
expérimentation

Expérimentation
(scénario, mise en scène)

- **Objet** : réel (terrain, phénomène, personnes, société, etc.)
- **Contexte** : prédéfini

Traiter

Analyse et sélection
(statistiques, corpus, base de données, anonymisation)

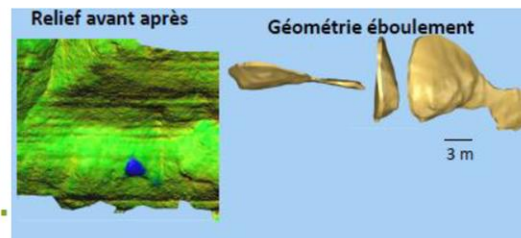
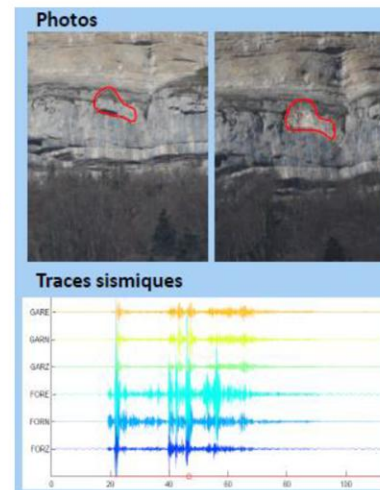
- **Objet** : Données récoltées et sélectionnées
- **Contexte** : scientifique et révision juridique

« Donnée brute » vs « Donnée élaborée » vs « Donnée dérivée »

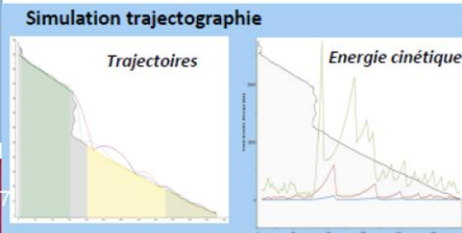
Définition : Données brutes - données élaborées (raw data – processed data)

« **Données brutes** : recueillies sur un sujet à partir d'observations ou de mesures, et qui n'ont pas encore été traitées, validées ou triées. Sensées être neutres et objectives et ne dépendant pas de leur contexte de création, d'une analyse ou de leur producteur » (Thessen & Patterson, 2011)

Données élaborées ou dérivées : ayant subi des opérations de nettoyage, de sélection, de traitement, d'analyse afin de produire des résultats.



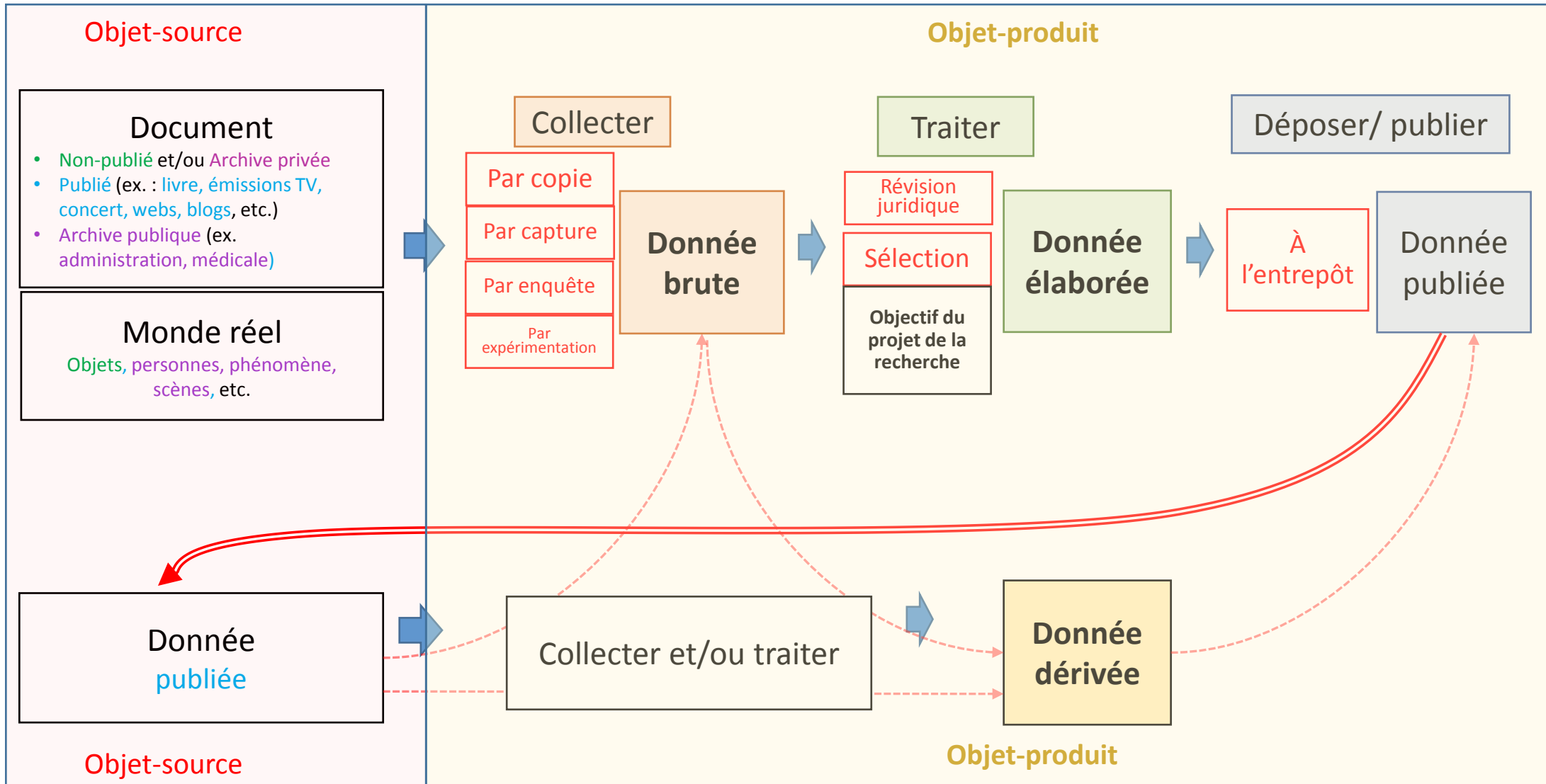
(Amitrano, Lejeune, & Hobléa, 2017)



3 états de la donnée de la recherche en SHS

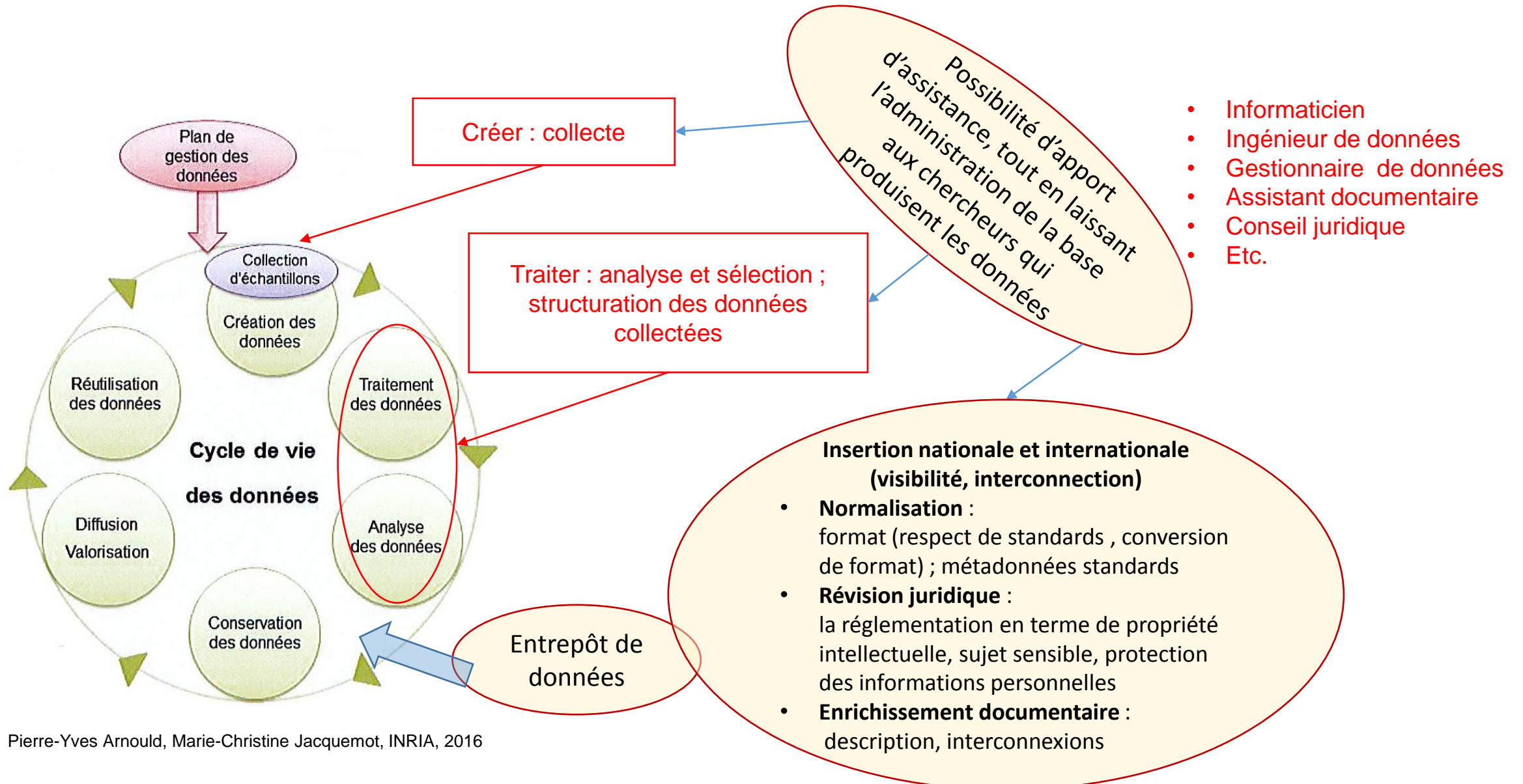
- Donnée brute
- Donnée élaborée
- Donnée dérivée

Cycle de vie de données de la recherche en SHS



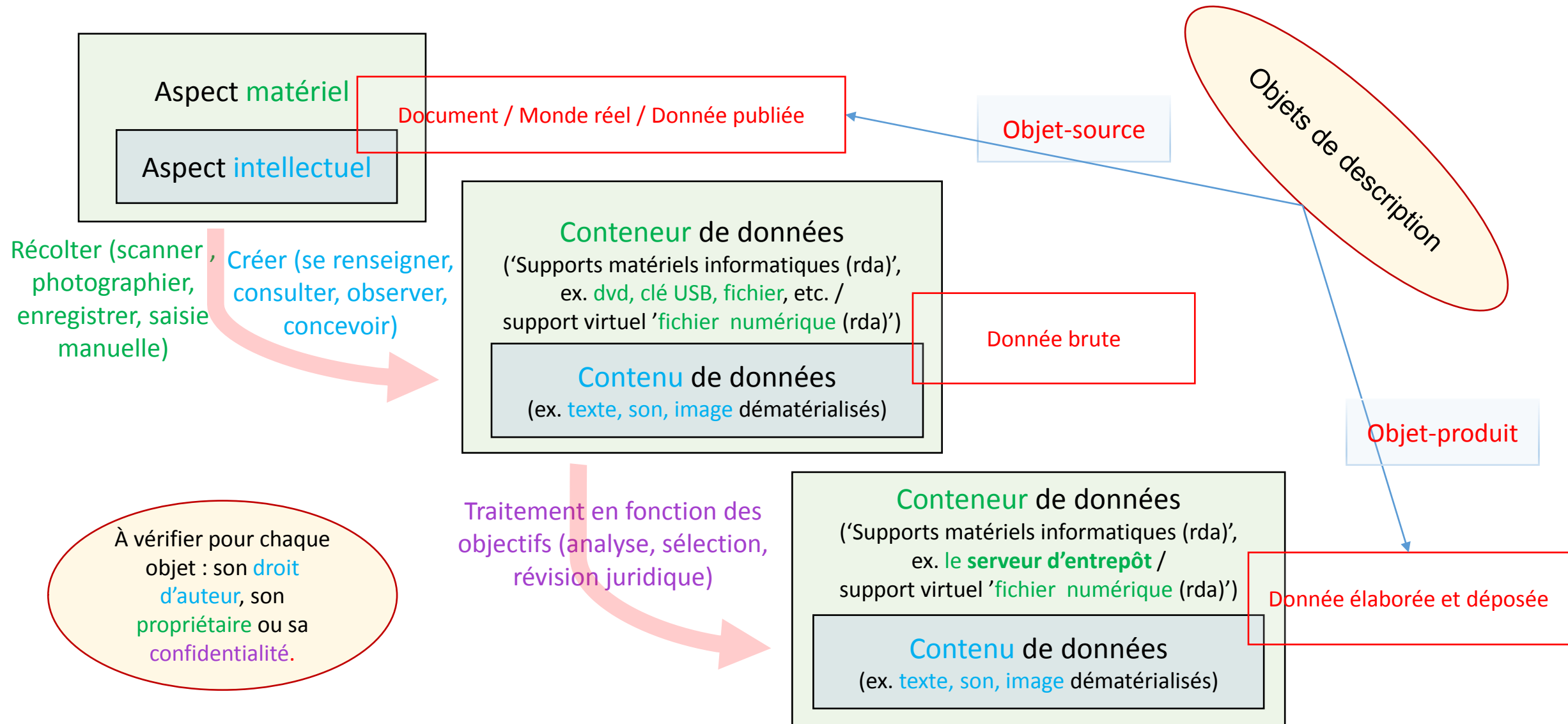
À vérifier pour chaque objet : son **droit d'auteur**, son **propriétaire** ou sa **confidentialité**.

Services et métiers : missions des gestionnaires

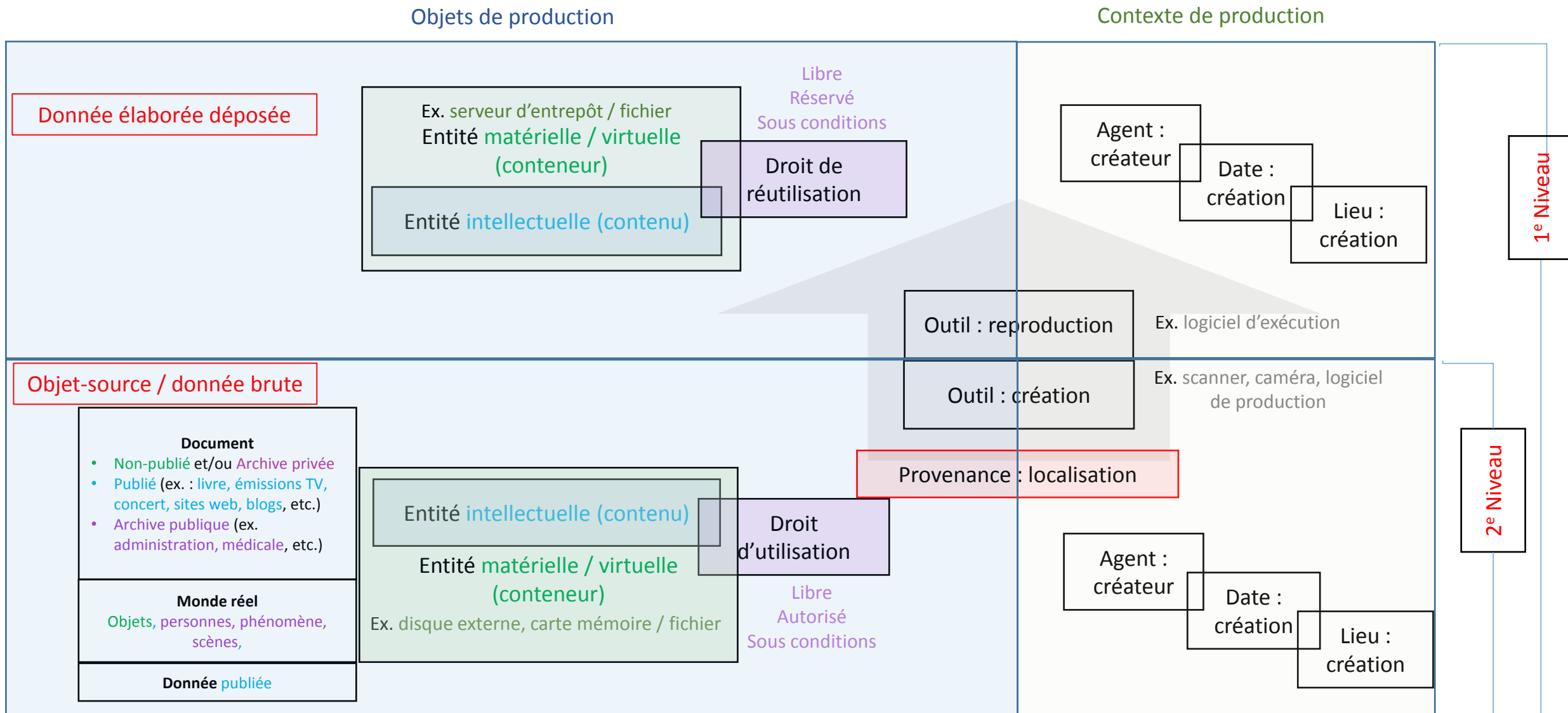


Pierre-Yves Arnould, Marie-Christine Jacquemot, INRIA, 2016

Modèle de description : plusieurs aspects de l'objet : « conteneur » et « contenu » vs « matériel », « virtuel » et « intellectuel »



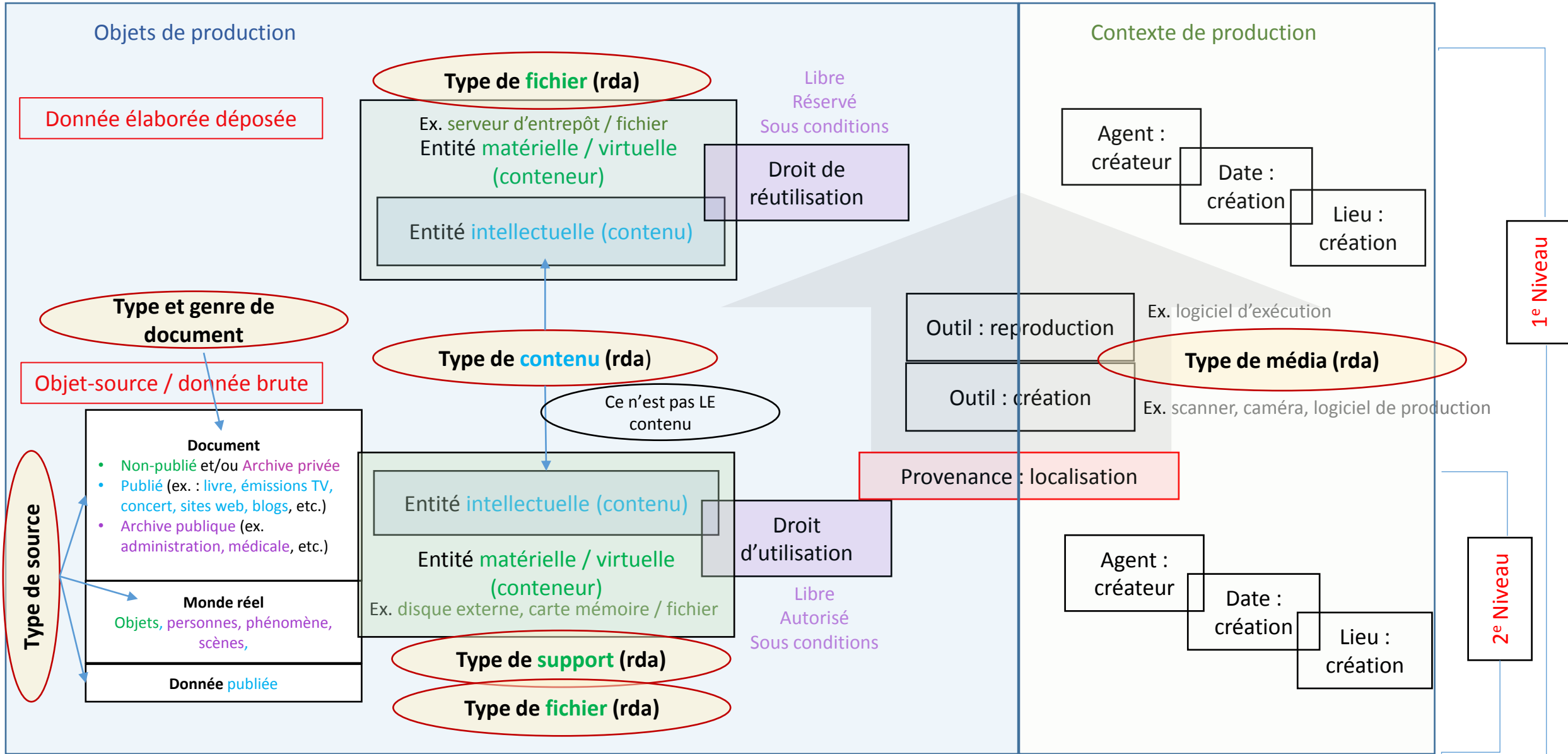
Un Modèle de « donnée de la recherche » (SHS) : Métadonnées communes



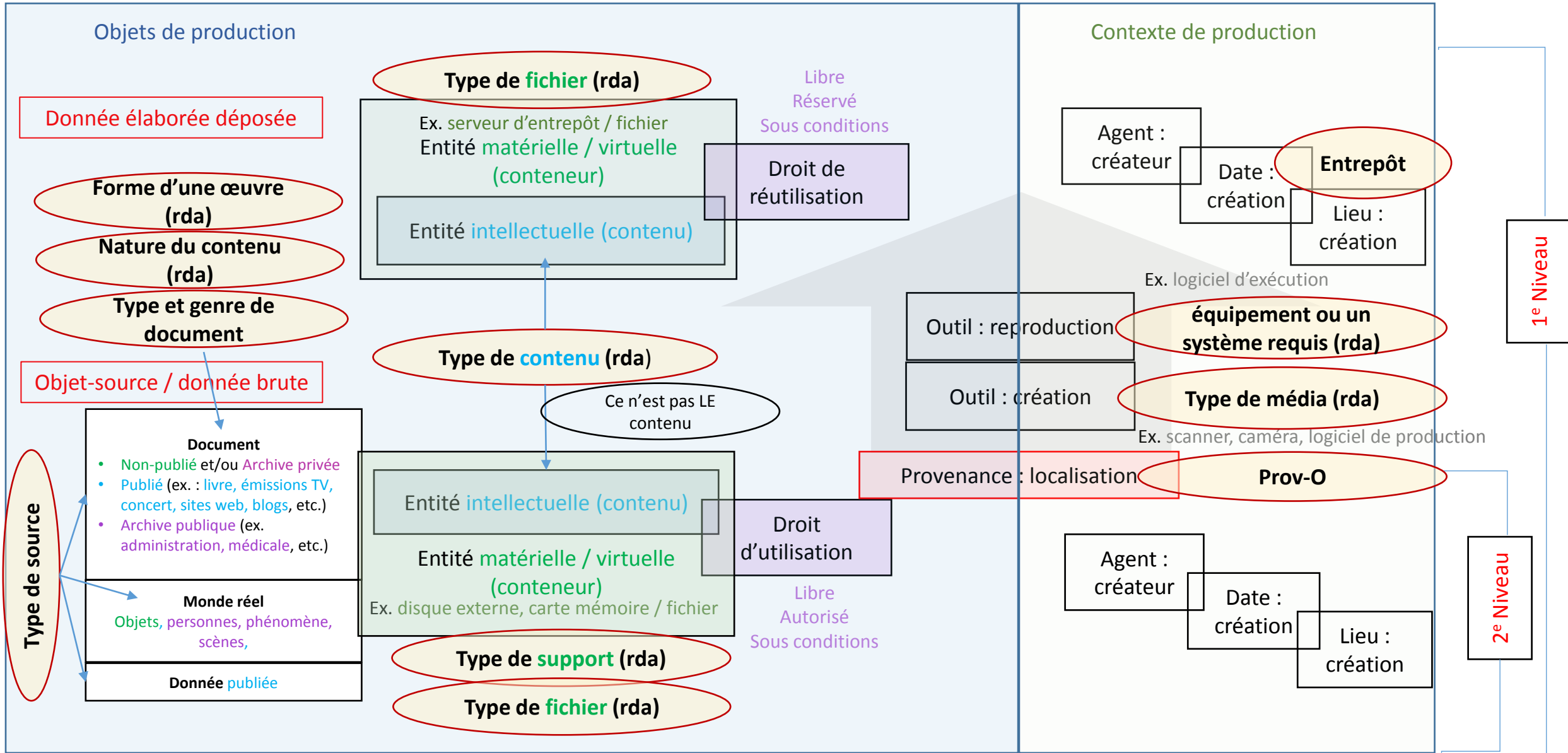
Données de l'entrepôt = « informatique (type de média) » avec « ressource en ligne (type de support) » (RDA)

	Type de support matériel RDA	Type de media RDA	Type de fichier RDA	Type de contenu RDA
<p>Supports matériels d'images projetées</p> <ul style="list-style-type: none"> • bobine de film • cartouche de film • cartouche de film fixe • cassette de film • diapositive • film fixe • film fixe court • rouleau de film • transparent pour rétroprojecteur <p>Supports matériels microformes</p> <ul style="list-style-type: none"> • bande de microfilm • bobine de microfilm • carte à fenêtre • cartouche de microfilm • cassette de microfiches • cassette de microfilm • micro-opaque • microfiche • rouleau de microfilm <p>Supports matériels microscopiques</p> <ul style="list-style-type: none"> • lame pour microscope <p>Supports matériels sans médiation</p> <ul style="list-style-type: none"> • feuille • fiche • objet • rouleau • tableau à feuilles mobiles • volume <p>Supports matériels stéréoscopiques</p> <ul style="list-style-type: none"> • carte stéréoscopique • disque stéréoscopique <p>Supports matériels vidéo</p> <ul style="list-style-type: none"> • bobine de bande vidéo • cartouche vidéo • cassette vidéo • vidéodisque 	<p>Catégorisation qui indique le format du support de stockage et du contenant d'un support matériel combiné avec le type de dispositif de médiation re équipement ou un système requis quis pour visionner, faire fonctionner, faire défiler, etc. le contenu d'une manifestation</p> <p>Objet-source</p> <p>Supports matériels informatiques</p> <ul style="list-style-type: none"> • bobine de bande informatique • carte informatique • cartouche à puce informatique • cartouche de bande informatique • cartouche de disque informatique • cassette de bande informatique • disque informatique • ressource en ligne <p>Donnée déposée :</p> <p>Matériel : serveur</p> <p>Virtuel : fichier numérique</p> <p>Supports matériels audio</p> <ul style="list-style-type: none"> • bobine de bande audio • bobine de fil audio • bobine de piste sonore • cartouche audio • cassette audio • courroie audio • cylindre audio • disque audio • rouleau audio 	<p>Catégorisation qui indique le type général de dispositif de médiation requis pour visionner, faire fonctionner, faire défiler, etc. le contenu d'une manifestation.</p> <ul style="list-style-type: none"> • audio • informatique • microforme • microscopique • projeté • sans médiation • stéréoscopique • Vidéo • autre 	<p>Type général de contenu des données encodées dans un fichier informatique.</p> <p>a) Type de fichier :</p> <ul style="list-style-type: none"> • fichier audio • fichier de données • fichier de programme • fichier image • fichier texte • fichier vidéo <p>b) format d'encodage (= MIME type)</p> <p>c) taille du fichier</p> <p>d) résolution</p> <p>e) code de région</p> <p>f) débit binaire codé</p> <p>Repérable automatiquement</p>	<p>Catégorisation qui reflète la forme fondamentale de communication sous laquelle le contenu est exprimé et le sens humain par lequel il est destiné à être perçu. Dans le cas d'un contenu exprimé sous la forme d'une ou plusieurs images, un type de contenu reflète également le nombre de dimensions spatiales au travers desquelles le contenu doit être perçu et la présence ou l'absence perceptibles de mouvement.</p> <ul style="list-style-type: none"> • données cartographiques • données informatiques • forme cartographique tridimensionnelle • forme cartographique tridimensionnelle tactile • forme tridimensionnelle • forme tridimensionnelle tactile • image animée bidimensionnelle • image animée tridimensionnelle • image cartographique • image cartographique animée • image cartographique tactile • image fixe • image tactile • mouvement noté • mouvement noté tactile • musique exécutée • musique notée • musique notée tactile • parole énoncée • programme informatique • sons • texte • texte tactile

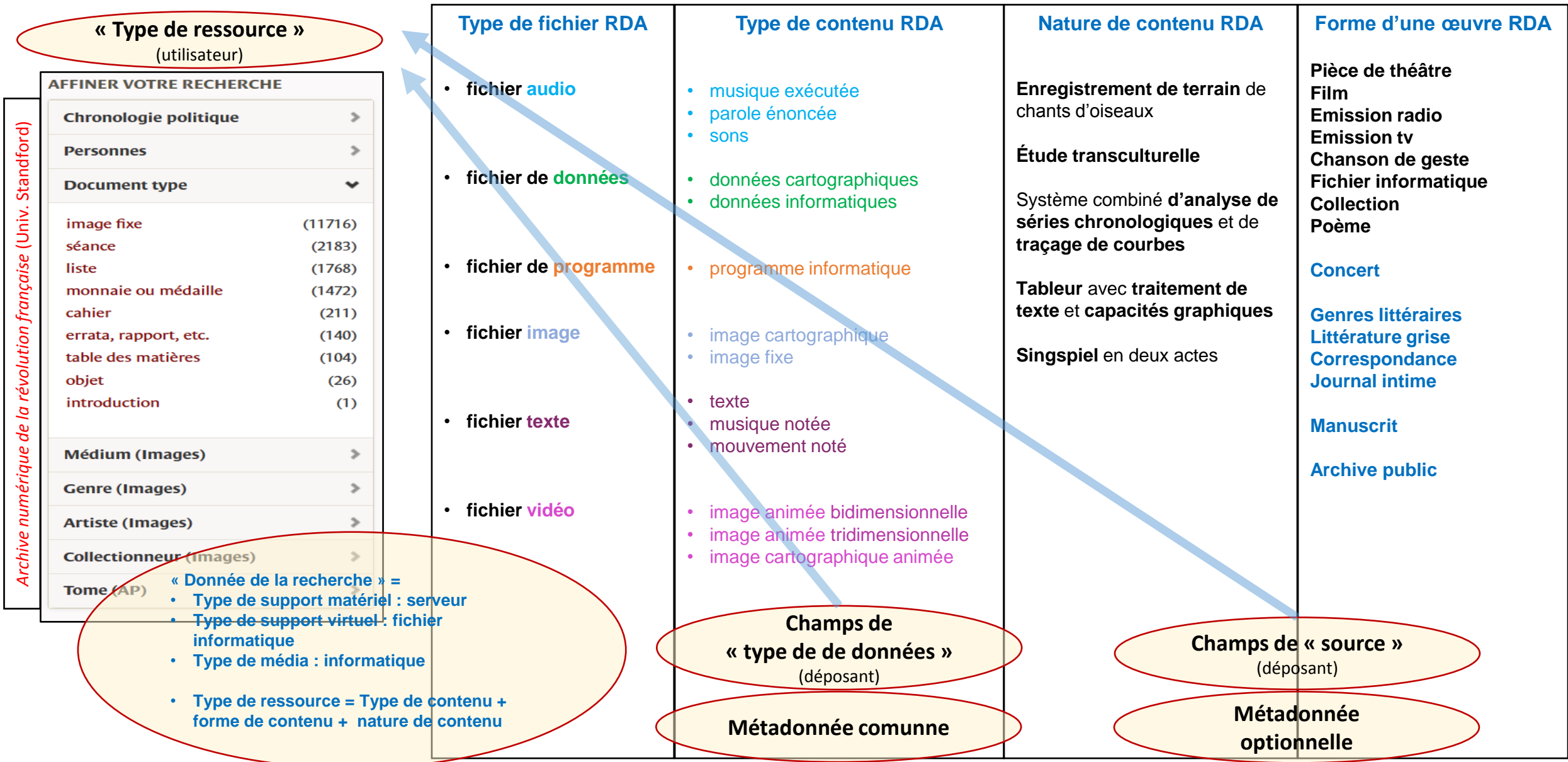
« Type de ressource » (utilisateur) vs « Type des composants » (gestionnaire)



Composition de métadonnées ; choix de vocabulaire contrôlé



« Type de ressource » = « types des « fichiers » / « contenus » » + « natures du contenu » + « formes de contenu » (RDA)

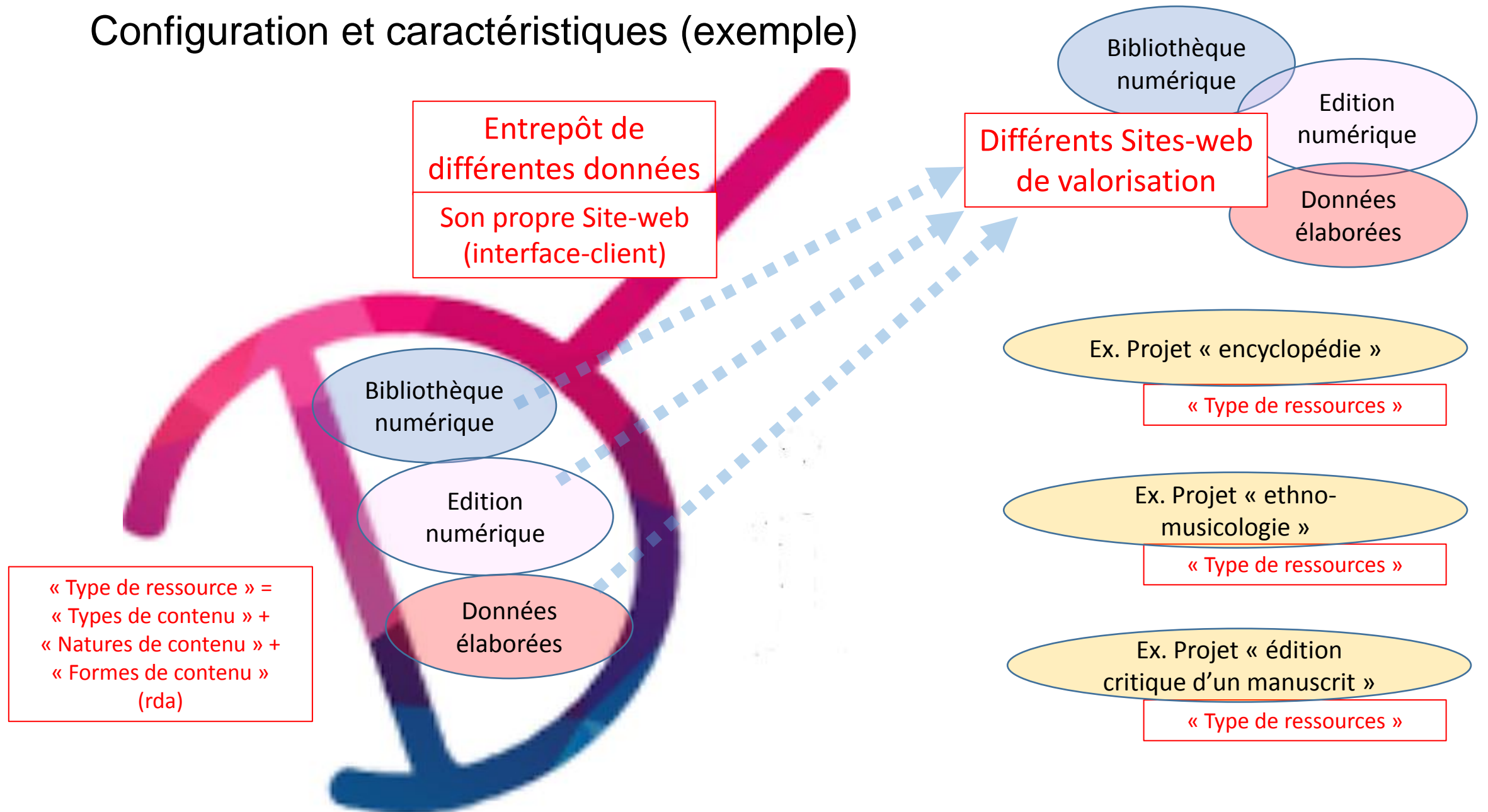


Configuration et caractéristiques

Type de numérisation	Objectif	Le contenu (œuvre)	Type de support	Type de média	Type de contenu	Forme de contenu	Exemples
1	Conservation de la source	inchangé	changé	changé	inchangé	inchangé	Numérisation en image (ex. bibliothèque numérique, musée)
2	Recherche de référence	inchangé	changé	Inchangé / changé	inchangé	Inchangé / changé	Numérisation avec intervention sur le contenu soit par encodage (xml) soit par OCR (ex. édition numérique)
3	Finalité scientifique	changé	changé	changé	changé	changé	Production de nouvelles données (ex. statistique, corpus, données géographiques)

Caractéristiques des « données de la recherche », définies selon le degré de modification entre l'objet-source / donnée brute et la donnée élaborée-publiée, d'après « 3 types de numérisation » (Eun-joo Carré-Na, 2016).

Configuration et caractéristiques (exemple)



Bibliographie

- OCDE, *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics*, 2007, Paris, <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>
- Pierre-Yves Arnould, Marie-Christine Jacquemot, *Guide de bonnes pratiques Gestion et valorisation des données de la recherche*, 2016, <https://hal.inria.fr/hal-01275841/>
- Joachim Schöpfel, Eric Kergosien, Hélène Prost, « « Pour commencer, pourriez-vous définir 'données de la recherche' ? » Une tentative de réponse », *Atelier VADOR : Valorisation et Analyse des Données de la Recherche; INFORSID 2017*, May 2017, Toulouse, France. 2017, <https://hal.univ-lille3.fr/hal-01530937>
- Sylvie Cocaud, « Les entrepôts de données de recherche », *Participer à l'organisation du management des données de la recherche, gestion de contenu et documentation des données*, Action Nationale de Formation RENATIS – MEDICI 2017, https://anfdonnees2017.sciencesconf.org/data/pages/Entrepots_ANFRenatis_2017_Cocaud_Aventurier_1.pdf
- Pain M. (2016). *Les données de la recherche et leurs entrepôts, de la documentation à la réutilisation: étude de cas pour l'archive HAL*. Mémoire de Master. Enssib, Villeurbanne. https://memsic.ccsd.cnrs.fr/mem_01374509
- Bibliothèque de l'université Stanford, *Archive numérique de la révolution française*, <https://frda.stanford.edu/fr>
- Eun-joo Carré-Na, « Réflexions sur les rôles des documentalistes dans la numérisation : le cas du « Répertoire historique de l'administration coréenne » de Maurice Courant », *10 e Rencontre annuelle du réseau DocAsie : les fonds asiatiques à l'ère du numérique*, Paris, 2016, <https://hal.archives-ouvertes.fr/hal-01337488>
- Ph. Le Pape, *Que cent zones zéro s'épanouissent*, ABES, 2014, <https://rda.abes.fr/2014/05/28/que-cent-zones-zero-sepanouissent/>
- Ph. Le Pape, *La zone zéro*, ABES, 2014, <https://rda.abes.fr/2014/05/26/la-zone-zero/>
- Ph. Le Pape, *RDA : à quoi sert l'élément Type de support ?*, ABES, 2017
- Transition bibliographique, « 4. RDA : Resource Description and Access », *Sensibilisation à l'évolution des catalogues*, RNF, 2015, https://www.transition-bibliographique.fr/wp-content/uploads/2016/01/support_formation_rda.pdf
- Transition bibliographique, « Type de médiation, Type de contenu », *Application des règles de catalogage RDA-FR*, 2017, https://www.transition-bibliographique.fr/wp-content/uploads/2017/04/RDA_FR2016_SupportFormation_Type_m%C3%A9diation_contenu.pdf
- Transition bibliographique, « III. Description des supports, matériels et des contenus », *Application des règles de catalogage RDA-FR*, 2017, https://www.transition-bibliographique.fr/wp-content/uploads/2017/04/RDA_FR2017_SupportFormation_TypedSupport.pdf