

SHS research data :
a description model for self-management

as part of the Didomena project, research data management system planning by EHESS

Overseas Korean Studies Librarian Workshop,
October 15-19, 2018, National Library of Korea

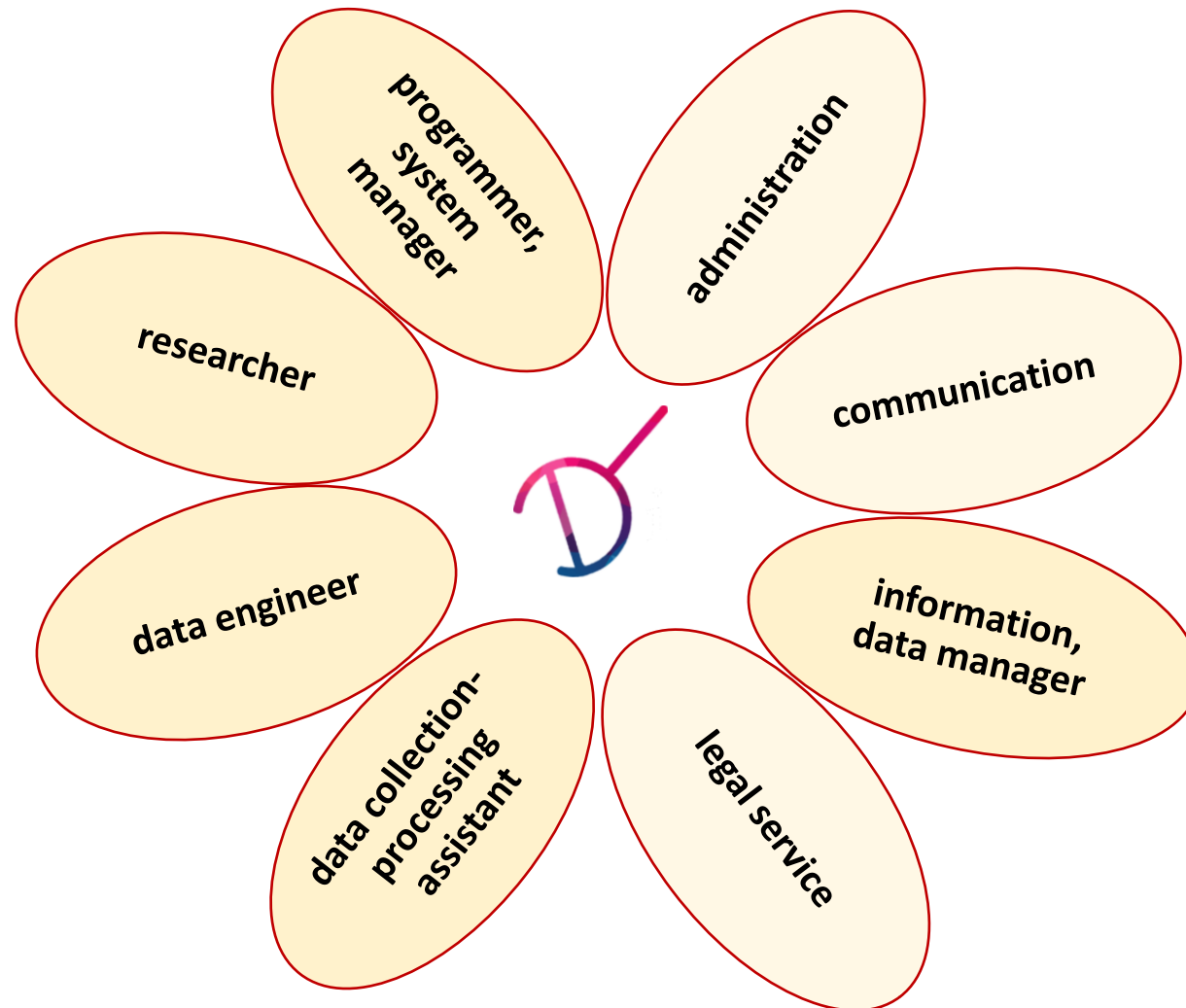
Eun-joo Carré-Na

Centre de Recherches sur la Corée (CRC), EHESS
UMR 8173 Chine, Corée, Japon, (CCJ), CNRS

Contents

- Information manager's perspective
- Research data : definition and criteria
- Research data : production type
- Research data : life cycle
- Related services and professions
- Several aspects of the description object
- Research data model for description metadata
- Research data : definition with RDA
- Composition of metadata ; selecting controlled vocabulary (RDA)
- « Type of resource » (user) vs « type of component » (administrator)
- Configuration and features

Information manager's perspective and role for planning



- global vision : general model
- metadata on certification standard : RDA
- interconnection with certification system : IdRef (abes)

« Research data » : definition and criteria

- << Research data are **factual records** (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the **scientific community** as necessary to **validate research findings**. >> (OECD, 2007)
- 4 criteria (Joachim Schöpfel, et al., 2017) :
 - 1) **Registration as a prerequisite** - material fixation of an idea on a medium as a prerequisite for certification to protect mental, intellectual work
 - 2) **Factual nature** - despite the diversity of data, all research data are based on factual nature - but as the Royal Society (2012) points out, it may be an "assumed" factual nature when research data is defined as "qualitative or quantitative statements or numbers that are (or assumed to be) factual" ;
 - 3) **Link with the scientific community** - the definition of OMB challenges the idea of an absolute concept and establishes a link with the researchers themselves through shared practices, value, methods, tools and concepts. In other words, research data is what is accepted by a group of researchers
 - 4) **Purpose of research** - the research data have academic purposes and play a role in the scientific process, especially for the validation of the hypotheses and results

Types of research data production (humanities and social science) : work type

Collect

Copy

copy / digitize materials

(**object** : materials digitized or copied through scanning, manual input, pictures etc. ; **acquisition context** and **location**)

- public archive (administrative document, medical document etc.)
- private archive (letters, diary etc.)
- non published (manuscript, grey literature, print etc.)
- publication (books, press, multimedia (website, blog, SNS etc.)

Capture

capture the real world

(on site research : recording, observation)

- **object** : real objects (site, people, phenomenon, society etc.)
- **context** : real

Survey

survey / interview

(interview, questionnaire / online survey, telephone survey etc.)

- **object** : questions and answers
- **context** : predefined context or reality

Experiment

experiment

(scenario, set context)

- **object** : real object (site, people, phenomenon, society etc.)
- **context** : predefined context

Process

analyze and select

(statistics, corpus, database, anonymization etc.)

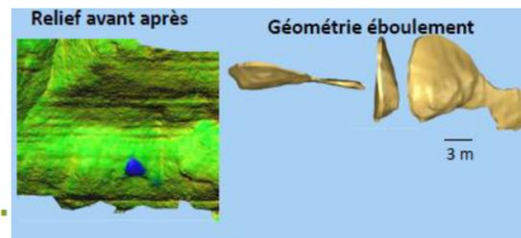
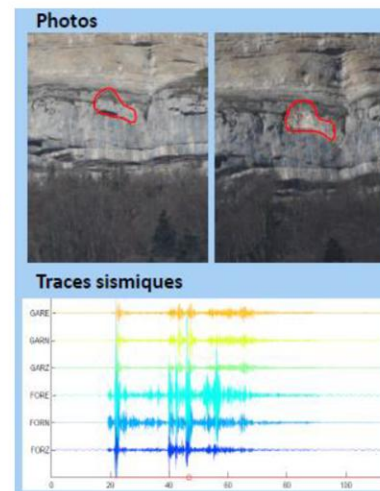
- **object** : collected raw data
- **context** : scientific and legal review

« Raw data » vs « Processed data » vs « Derived data »

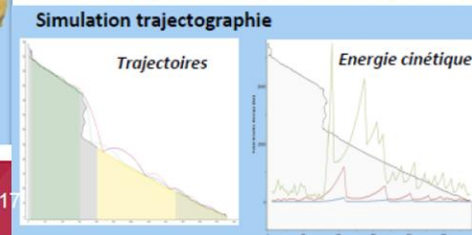
Définition : Données brutes - données élaborées (raw data – processed data)

« 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.



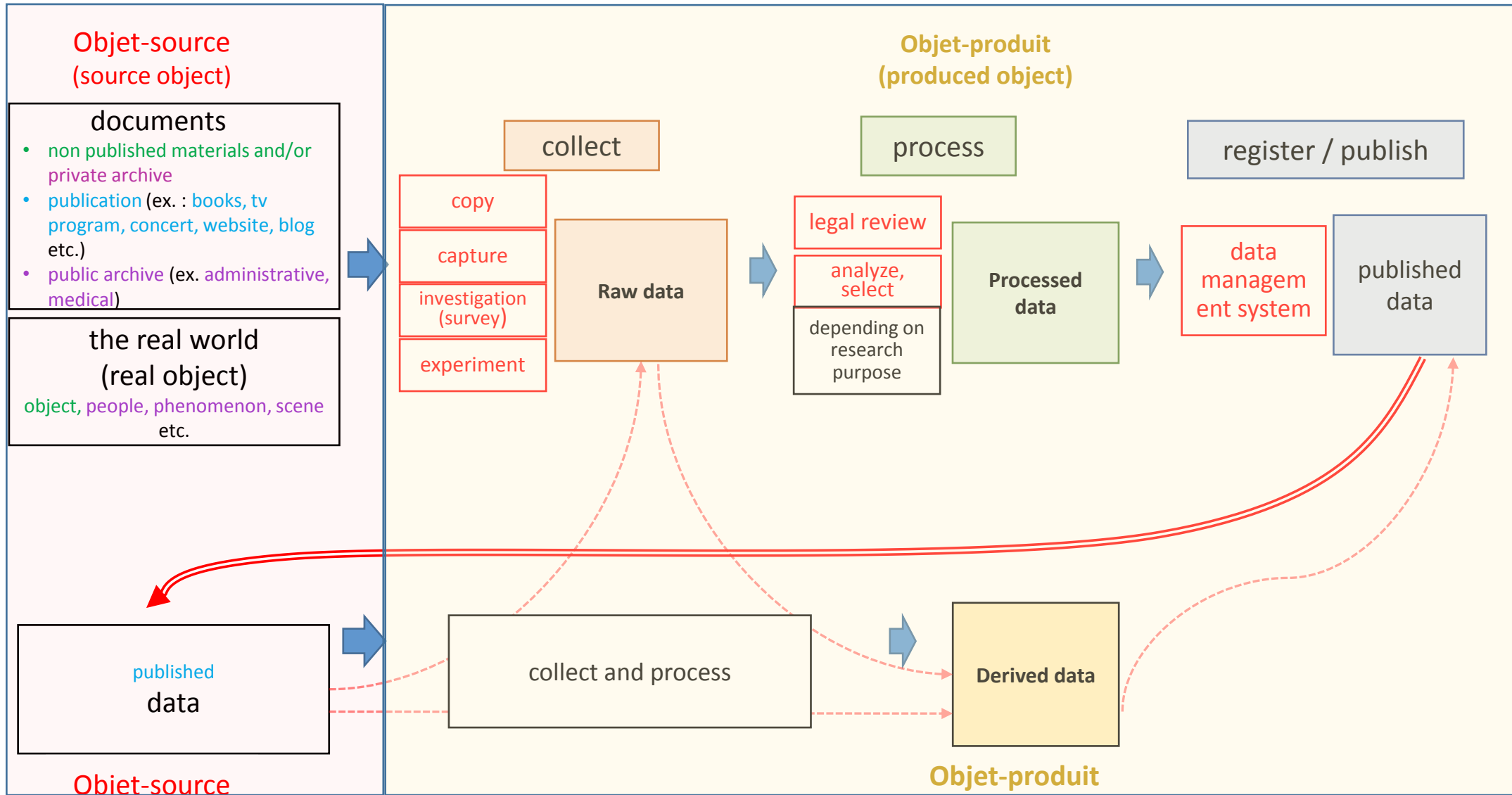
(Amitrano, Lejeune, & Hobléa, 2017)



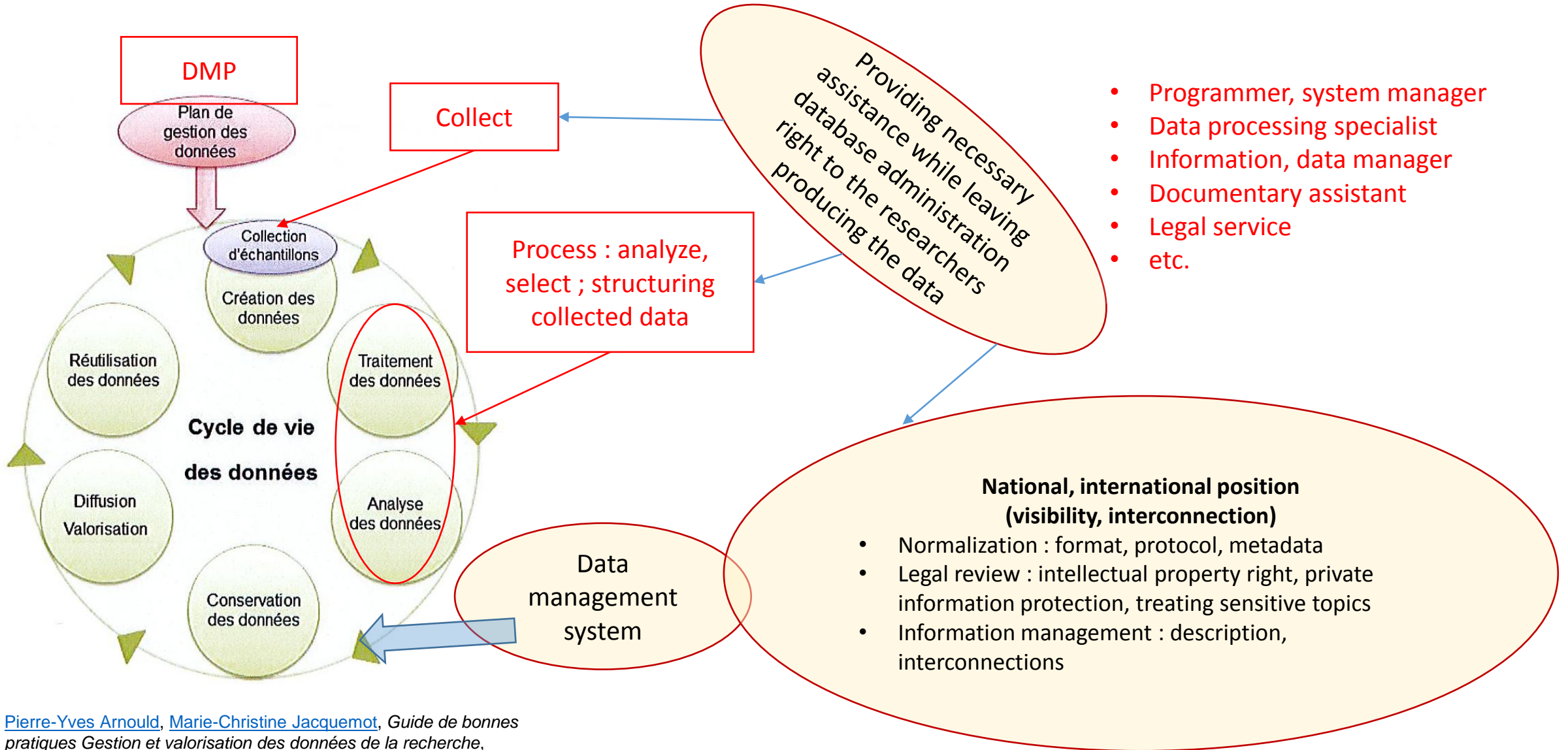
3 statuts of research data (SHS)

- Raw data
- Processed data
- Derived data

Research data life cycle (humanities and social science) (proposition)

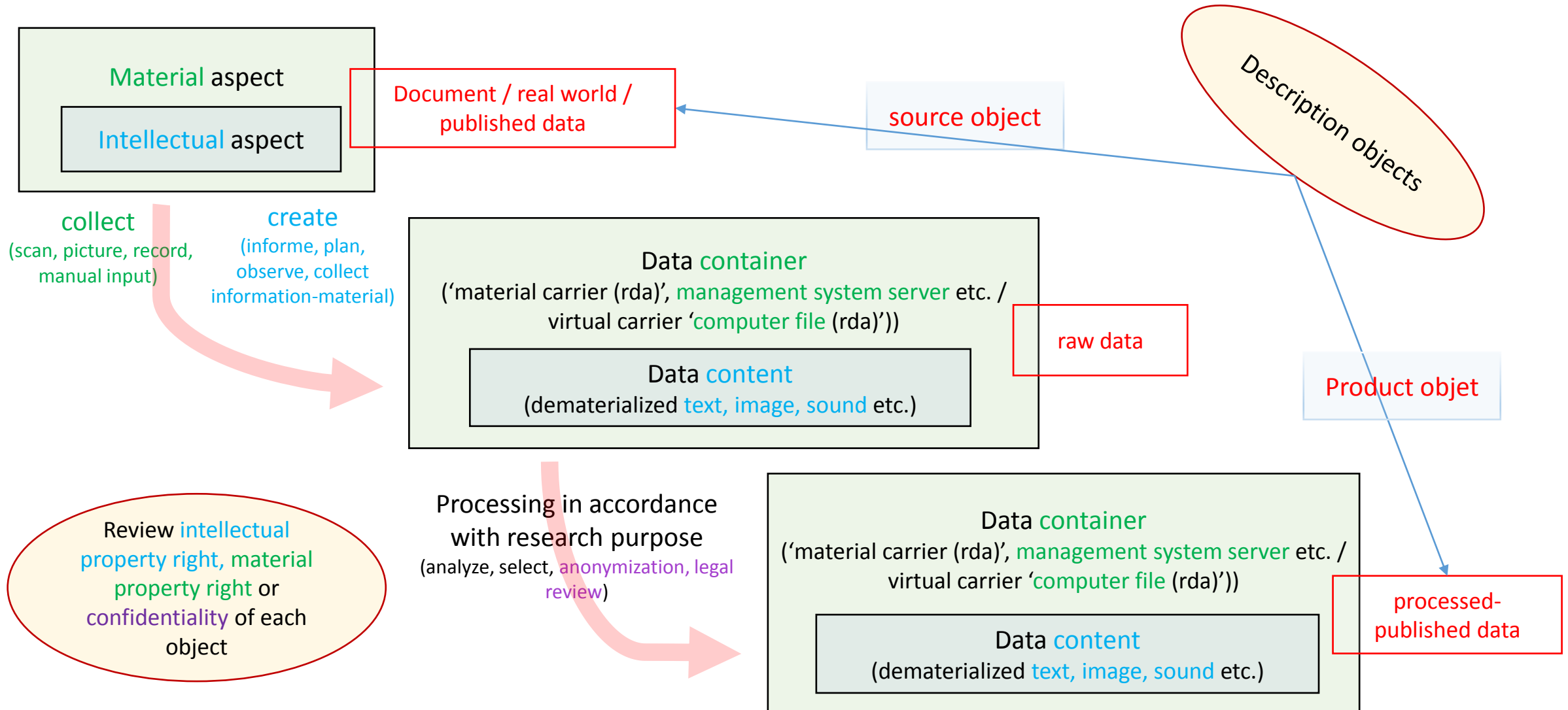


Related services and professions

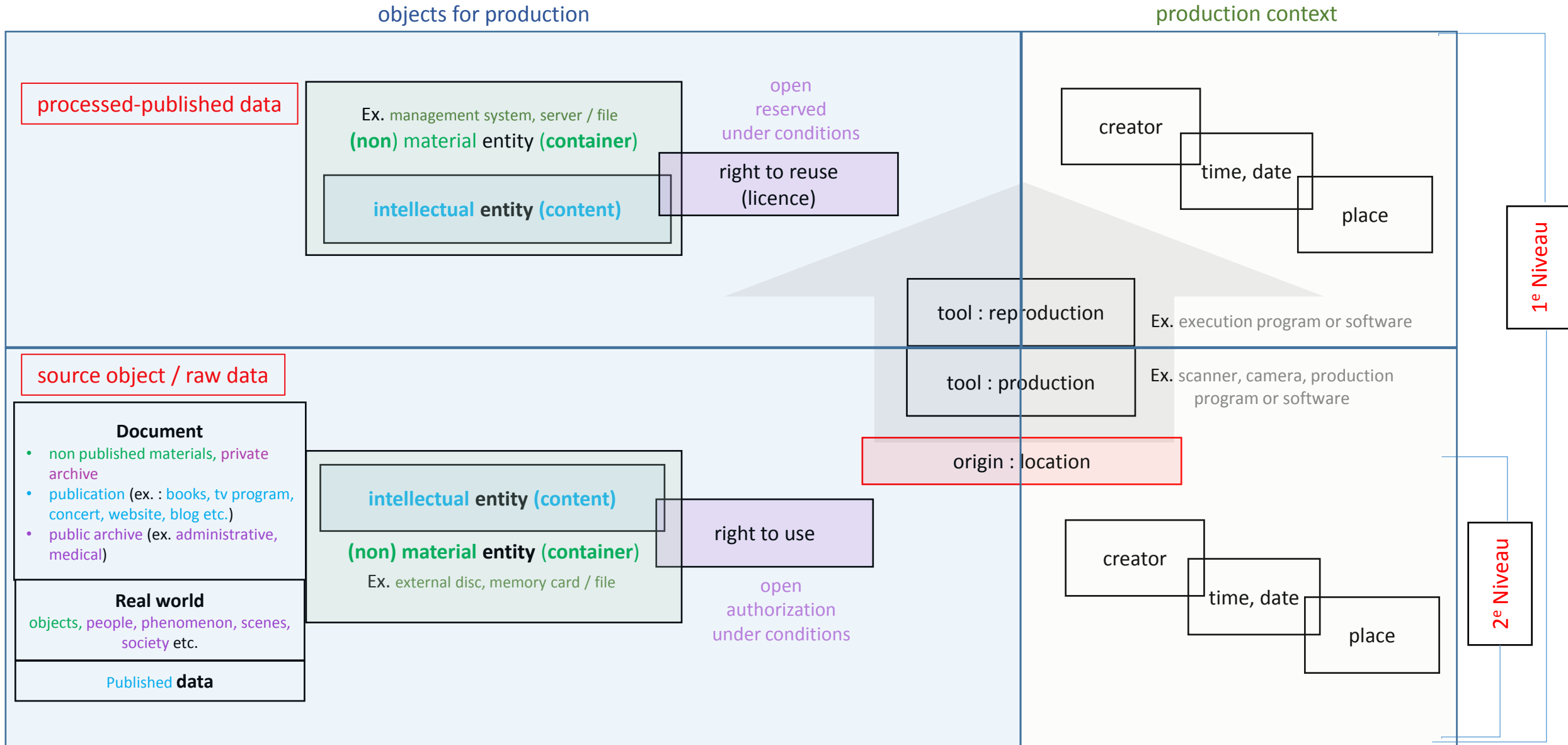


[Pierre-Yves Arnould](#), [Marie-Christine Jacquemot](#), *Guide de bonnes pratiques Gestion et valorisation des données de la recherche*,

Several aspects of description object : « Container » and « contents » vs « material » - « non material » and « intellectual »



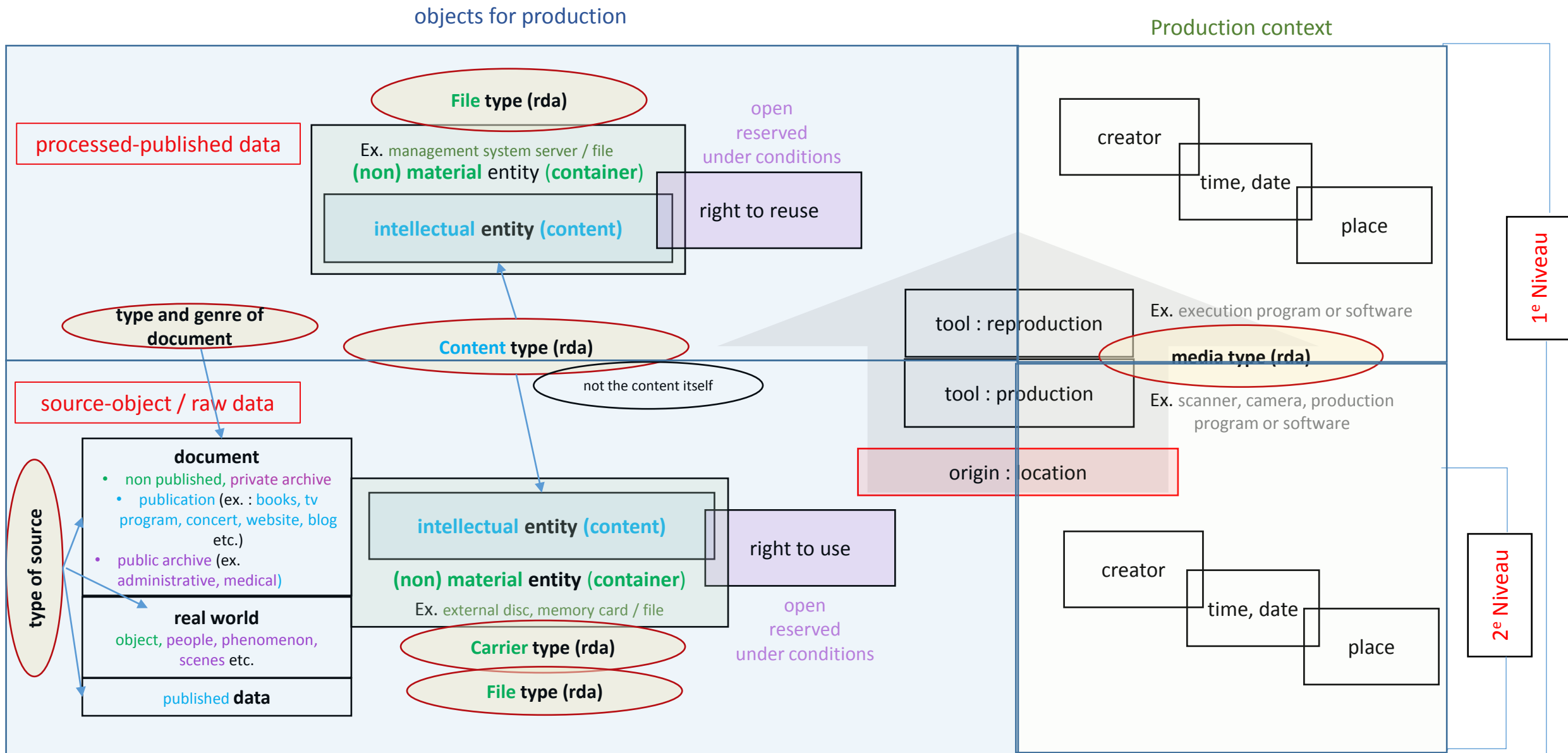
A research data model (object of description) (humanities and social science) : common metadata



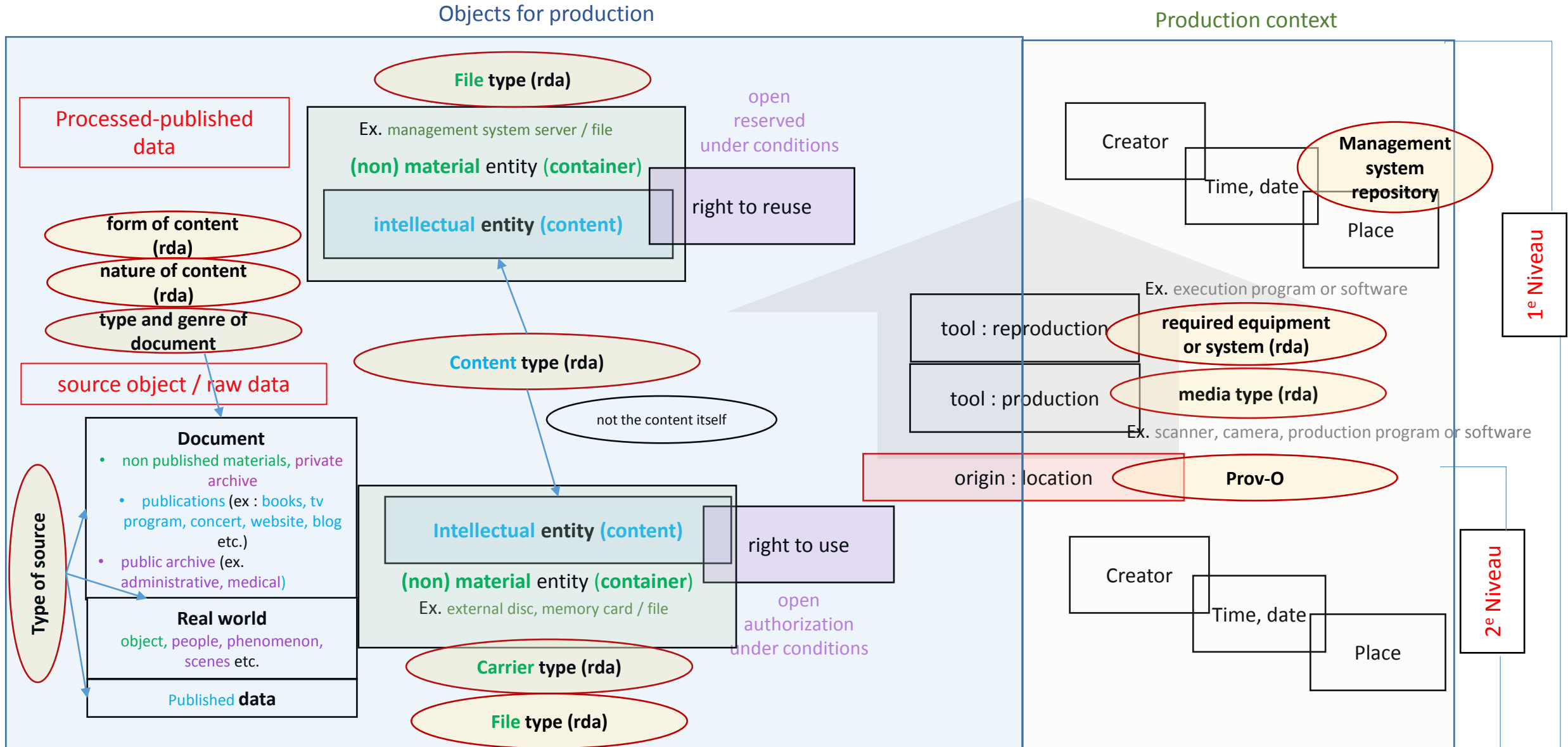
Published data = « computer (media type) » + « online resource (carrier type) » (RDA)

	Carrier type RDA	Media type RDA	File type RDA	Content type 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"> • computer tape cartridge • computer card • computer chip cartridge • computer tape cassette • computer disc cartridge • computer tape reel • computer disc • online resource <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>Published data : Material : serveur Virtual (dematerialized) : fichier numérique</p>	<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 • Computer • microforme • microscopie • projeté • Unmediated • stereoscopic • Video • autre 	<p>Type général de contenu des données encodées dans un fichier informatique.</p> <p>a) File type :</p> <ul style="list-style-type: none"> • audio file • data file • program file • image file • text file • video file <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>automatically detectable</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"> • cartographic dataset • computer dataset • cartographic tactile three-dimensional form • three-dimensional moving image • tactile three-dimensional form • image animée bidimensionnelle • image animée tridimensionnelle • cartographic image • cartographic moving image • cartographic tactile image • still image • tactile image • notated movement • tactile notated movement • performed music • notated music • tactile notated music • spoken word • computer program • Sounds • text • tactile text

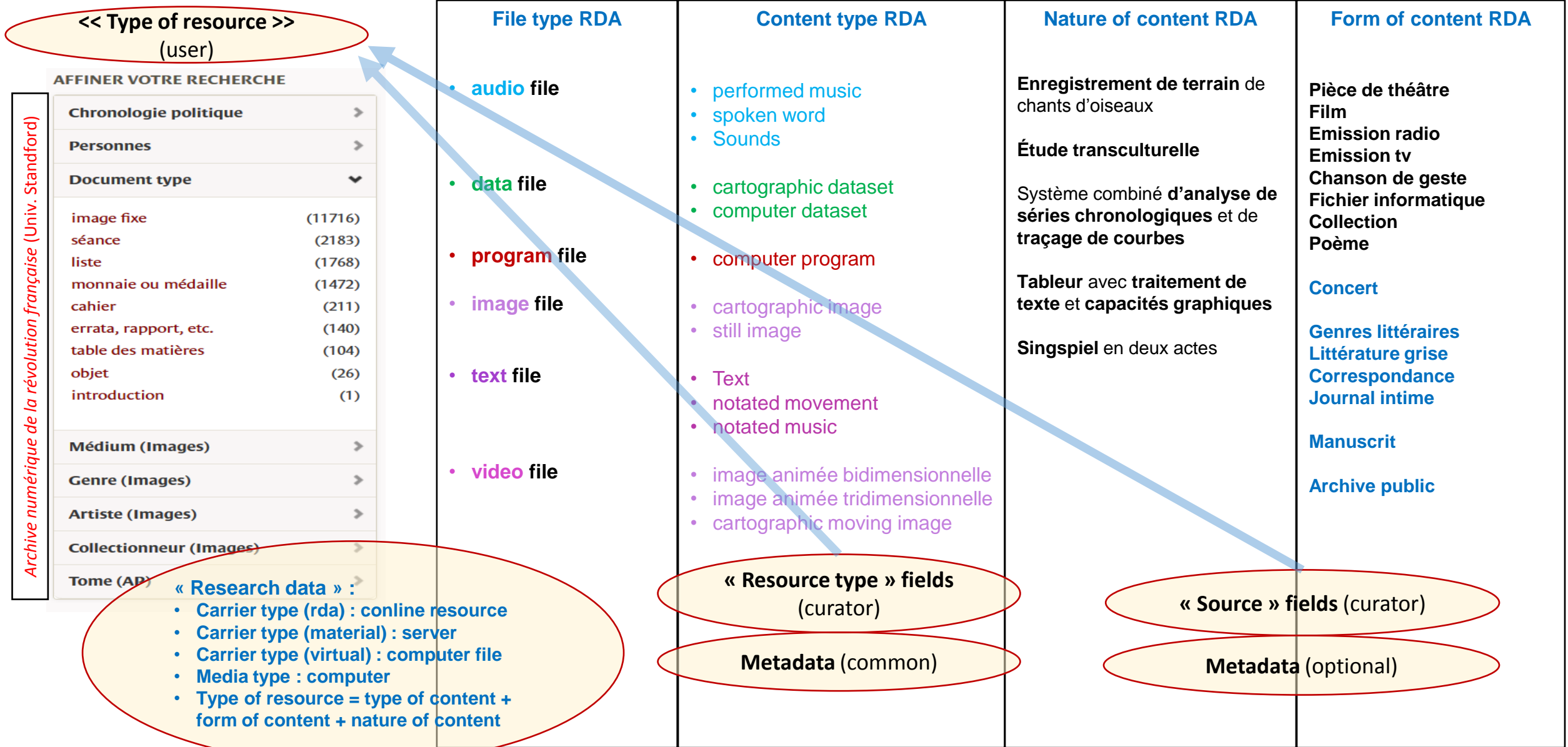
« Type of resource » (user) vs « type of component » (administrator)



Composition of metadata ; selecting controlled vocabulary



« Type of resource » = « file type » + « content type » + « natures of content » + « forms of content » (RDA)

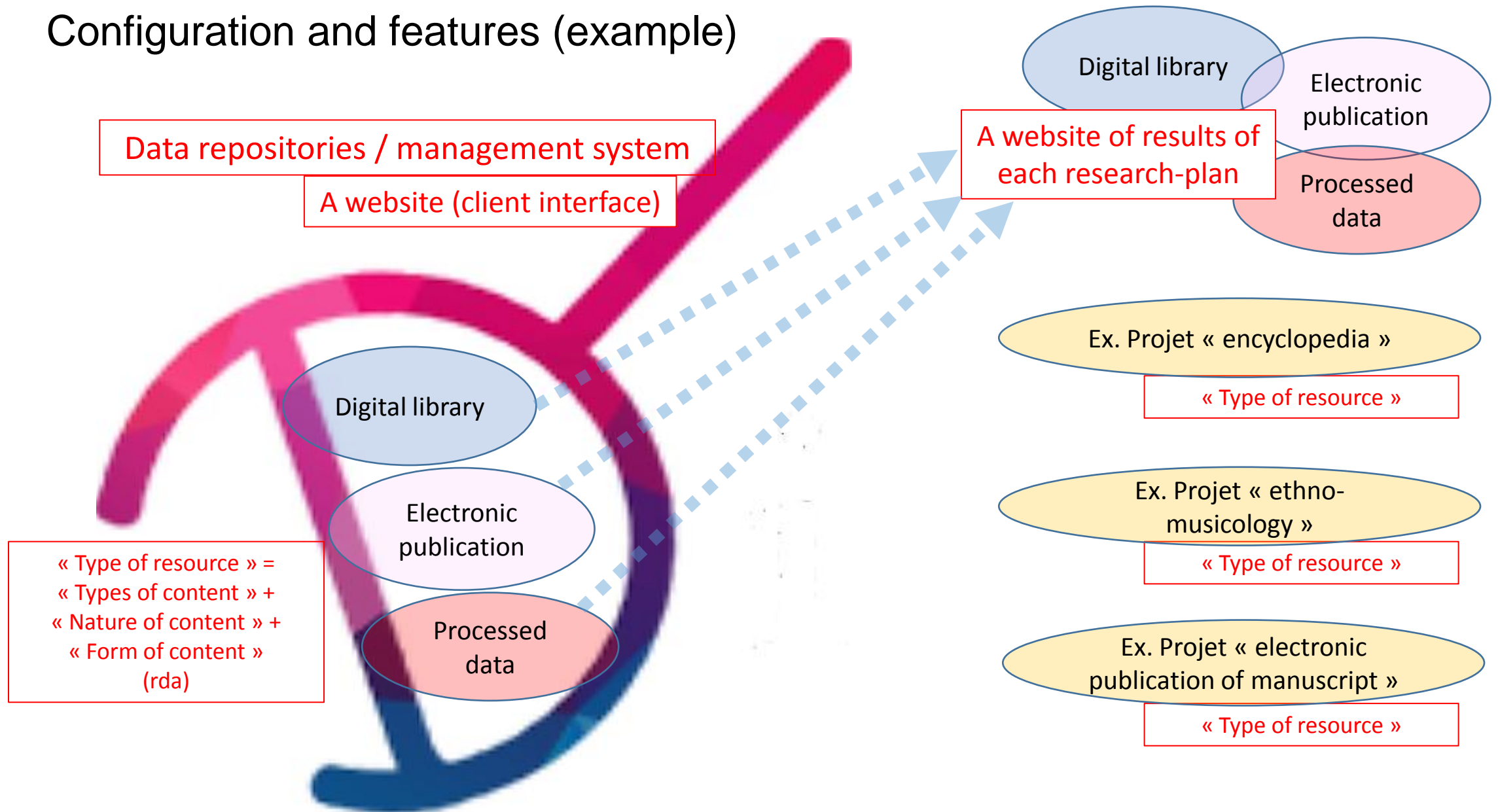


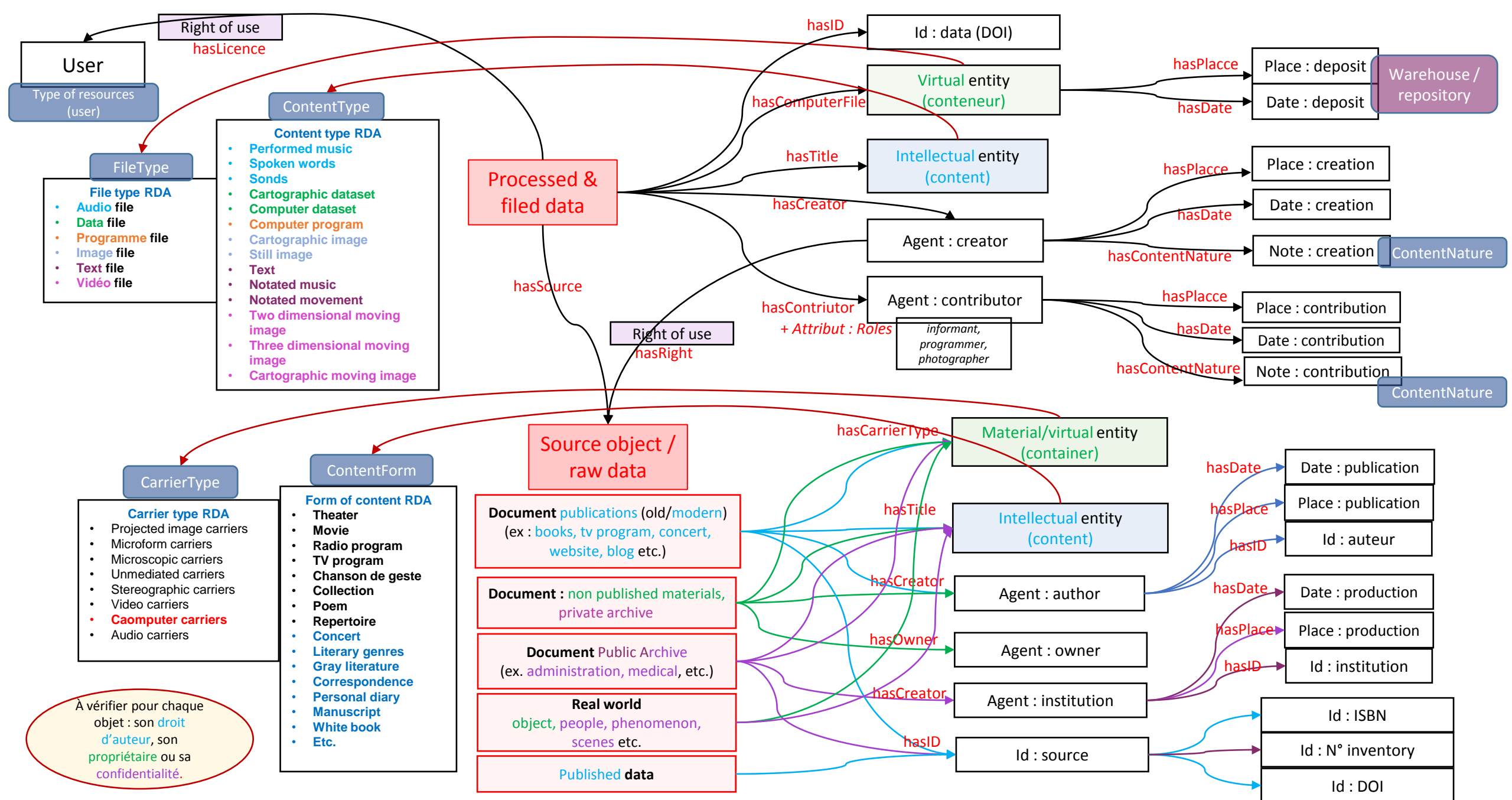
Configuration and features

Type of digitization	Objective	Content (work)	Type of carrier	Type of media	Type of content	Form of content	Examples
1	Conservation	Unchanged	Changed	Changed	Unchanged	Unchanged	Image digitization (ex. digital library, museum)
2	Reference research	Unchanged	Changed	Unchanged or changed	Unchanged	Unchanged or changed	Scanning with intervention on the content either by encoding (xml) or by OCR (ex. digital edition)
3	Scientific purpose	Changed	Changed	Changed	Changed	Changed	Production of new data (ex. statistics, corpus, geographic data)

Characteristics of the "research data" defined according to the degree of modification between the source / raw data and the elaborate-published data, according to "3 types of digitalization" (Eun-joo Carré-Na, 2016)

Configuration and features (example)





Reference

- 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://memic.ccsd.cnrs.fr/mem_01374509
- Bibliothèque de l'université Standford, *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