



HAL
open science

Document theory and document design

Arthur Perret

► **To cite this version:**

Arthur Perret. Document theory and document design. 2023 Annual Meeting of the Document Academy, Sep 2023, Paris, France. 10.35492/docam/10/2/6 . hal-04785507

HAL Id: hal-04785507

<https://hal.science/hal-04785507v1>

Submitted on 20 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Document theory and document design

Arthur Perret

Université Jean Moulin Lyon 3

arthur.perret2@univ-lyon3.fr

Document design or information design?

Findeli¹ defines design as diagnostic and prescriptive, an example of “projective” epistemology: designers view the world as a project to realize, rather than an object to know. This definition of design implies that document design is less about what documents *are* and more about what they *should do*. This aligns with the evolution of document theory from an ontic view of documents to a functional one, which was pioneered by the European documentation movement in the first half of the 20th century and reintroduced in the literature during the 1990s as “neo-documentation”.

Currently in the literature, document design is mostly considered to be a part of information design.² The purpose of information design is to create effective communications.³ Document design is usually taught within communications programs; in the US, in either English or fine arts departments⁴; in France, in Information and Communication Science departments.

Research on information design has very few connections with documentation. It has been criticized for its lack of theory: Carliner⁵ warned about the field’s tendency to generalize findings despite shaky conceptual foundations and a limited understanding of topics such as the World Wide Web.⁶ We believe document theory can help address some of these challenges and bring perspective to information design.

Zacklad⁷ gives us an example of this. He draws from document theory to define information design as the interplay between textualization (creating content information), documentarization (shaping the medium) and auctorialization (creating author identities). Zacklad puts documentarization at the heart of information design (see fig. 1). Documentarization can be internal (the organization of fragments into documents) or external (the relational organization of documents, e.g. classification); it can involve different roles or positions (author, editor, printer/programmer, distributor, reader); it implies significant differences between designing the systems that involve documents (“system-oriented design”) and designing the documents themselves (“author design”).

¹ “La recherche-projet en design et la question de la question de recherche,” 2015.

² The *Information Design Journal* was created in 2006 at John Benjamins by merging *Document Design* and *Information Design*.

³ Renkema, “Editorial,” 2006; see also Pettersson, *Information design*, 2002.

⁴ Shriver, *Dynamics in document design*, 1997, p. 92-93.

⁵ “Current challenges of research in information design and document design,” 2006, pp. 10-14.

⁶ On the absence of the Web in information design research, see the discussion between Shriver and Carliner (“Ten years after,” 2007, p. 167).

⁷ “Information Design,” 2019.

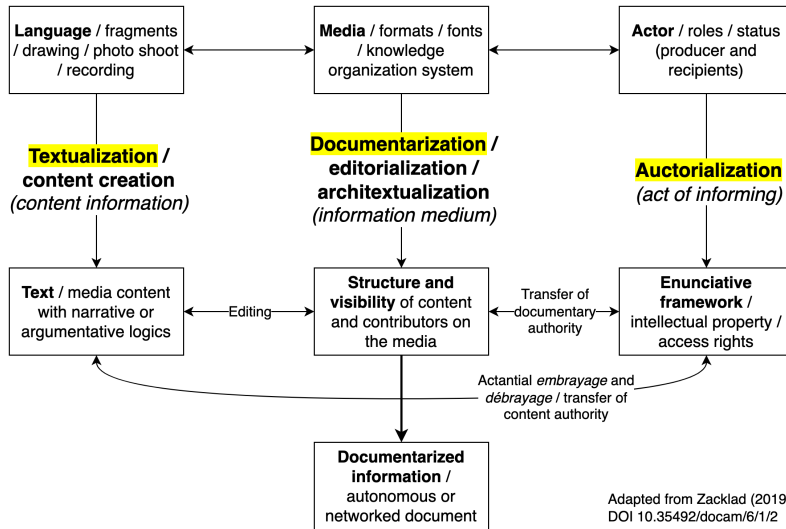


Figure 1: Information design according to Zacklad

In this paper, we explore the usefulness of document theory for system-oriented and author design for documents. Specifically, we use the example of document graphs, which are interrelated documents or fragments of documents.⁸ We present an experimental visualization program for document graphs; we describe how document theory informed the design of this program. Finally, we discuss the idea of a relational approach to document theory and its usefulness for document design.

Document graphs as an object of design

The concept of document graphs differs from that of knowledge graphs, which do not necessarily involve documents but in fact exist mostly as databases. Document graphs belong to hypertextual documentation. They can take various shapes and names: collections, wikis, notebooks, digital gardens, etc. In the case of fragment based document graphs, they are used for instance in XML publishing systems.

System design for document graphs defines how they can be made. It also creates tools to help with their complexity. In his early experiments with hypertextual documentation, Engelbart⁹ wrote that the biggest challenge was to keep track of all the links he was making. In our opinion, this inherent complexity is still the main issue for the design of hypertextual documentation systems.

Author design for document graphs involves analysis and synthesis. Analysis is achieved by creating documents, involving textualiza-

⁸ Arribe, *Conception des chaînes éditoriales*, 2014.

⁹ *Augmenting Human Intellect*, 1962, p. 62.

tion (taking notes, expressing ideas) and documentarization (adding metadata). Synthesis is achieved by linking documents.

Nowadays, author design of document graphs is facilitated by the increased availability of dedicated tools. Recently, the expression “tools for thought”¹⁰ has been used to describe a wave of new tools including Roam Research (2017), Zettlr (2017) and Obsidian (2020). These tools are based on lightweight markup languages, which use fewer signs than established markup languages such as HTML and XML; this simplifies hypertext writing, making it easier to create document graphs. These languages include Markdown, created as a shorthand for HTML; YAML, to express metadata; and WikiText, which has become a near-universal convention for creating internal links between plain text documents within a collection. Here is an example:

Listing 1: A digital index card written in Markdown, with a YAML metadata block at the top and WikiText links (in double brackets) pointing to other cards based on their title.

```

---
title: Bibliothécaires et documentalistes (Briet 1951)
type: publication
---

[[Suzanne Briet]] says [[Robert Pagès]] was ahead of his time when
he wrote about [[documentation]].

> « Il faut en revenir à Pagès. Son message n'a pas eu, au moment
ou il l'a lancé, tout le retentissement qu'il méritait, parce qu'
il ne trouva pas d'audience préparée à la recevoir. C'est
pourquoi, deux ans plus tard, nous avons nous-même tenté d'
expliquer ce qu'était à nos yeux la [[documentation]] : une
technique du travail intellectuel, une profession nouvelle, un
besoin de notre temps. »

```

Some of these new hypertextual tools provide navigational features that help with the complexity of document graphs. Some also support scientific and technical writing. Most offer sharing capabilities. However, none include all of these things and interoperability is limited: many of these tools are built and distributed according to a proprietary model, with design choices (such as unique syntax variants) that aim to retain users at the cost of long-term data accessibility and robustness. This creates the need for more tools, which in turns represents an opportunity to think about system design for document graphs (which entails defining how these graphs can be made, and creating tools that help with the complexity).

¹⁰ On the origins of this expression, see Matuschak and Nielsen, *How can we develop transformative tools for thought?*, 2019.

An example of system design for document graphs

With the support of the HyperOtlet research programme, we led an experiment in system design for document graphs. The result is Cosma, a program that helps visualize document graphs created with tools such as Zettlr. Here, we will briefly summarize the core design decisions, in order to later examine their relationship with document theory.¹¹

There are several ways to design an interface to work with relational data. It all depends on what the data is: how it is created and how it is meant to be used. If the intent is to crawl the web to map a controversy, one may use software like Gephi (2015) and Hyphe (2016), which use graph theory to provide network analysis capabilities. But if the intent is to create non-linear research documentation and use it as a memory aid, a medium for the synthesis of new ideas, then a different kind of interface may be more appropriate.

The context of our experiment was that we had created a personal document graph with Zettlr to support our doctoral research. In creating this graph, our key finding was that linking belongs to knowledge organization processes (KOPs). The most well-known KOPs are classifying,¹² indexing¹³ and tagging.¹⁴ We found that linking can be used to classify, index and tag, as well as to compose ideas. Listings 2, 3 and 4 show examples of this.

This provided important design guidance for our experiment, the goal of which was to find the most appropriate interface to make further use of this graph. Rather than features based on graph theory (such as finding clusters, or calculating paths between nodes), it emphasized the need for features related to organization and ideation. This includes for instance being able to assign categories to documents and links, as well as to see backlinks (i.e. where a document has been linked from) and the context surrounding them.

Following this, we designed Cosma, a visualization tool that uses open, standard formats for interoperability with creation tools. It reads a collection of individual records and redocumentarizes them into a single, standalone hyperdocument, which presents the document graph as an interactive network of index cards. We named this document a “cosmoscope” as an homage to Paul Otlet.¹⁵ A cosmoscope includes features that help reduce complexity through information surfacing (graph view, contextualized backlinks) and context switching (animations, display filters).

¹¹ A more detailed account is given in Perret, *De l'héritage épistémologique de Paul Otlet à une théorie relationnelle de l'organisation des connaissances*, 2022.

¹² Hjørland, “Classification,” 2017.

¹³ Hjørland, “Indexing,” 2018.

¹⁴ Rafferty, “Tagging,” 2018.

¹⁵ The *cosmoscope* and the *cosmographe* are symbolic devices used by Otlet to describe his vision for documentation. For more details, see Le Deuff and Perret, “Hyperdocumentation,” 2019.

Listing 2: A concept card pointing to another concept using an predefined link type (“generic”). Linking is used (along with tagging) to classify.

```

---
title: Annotated bibliography
type: concept
---

An annotated bibliography is a kind of [[generic:bibliography]].

```

Listing 3: A concept card pointing to a person card. Linking is used to index.

```

---
title: Dynamic medium
type: concept
---

The concept of dynamic medium was imagined by [[Alan Kay]].

```

Listing 4: An idea card expressing a new thought by linking to concept cards. Linking (in the form of composition) is used to ideate.

```

---
title: The epistemic status of a card evolves with time
type: idea
---

The [[epistemic status]] of a [[card]] evolves with time.

```

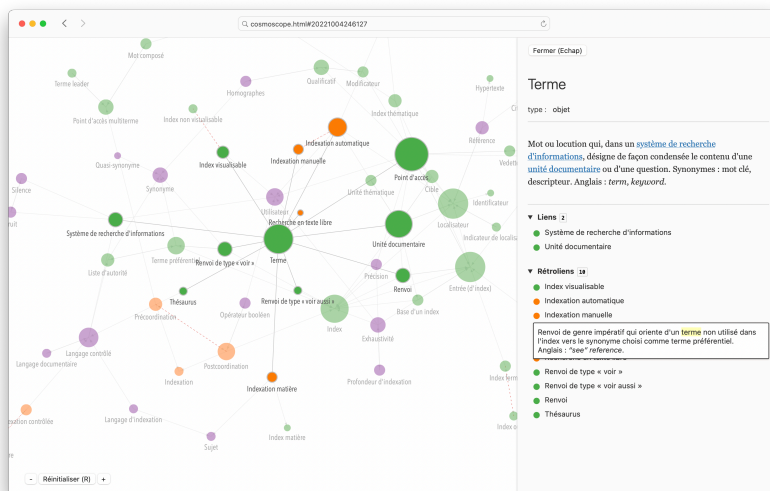


Figure 2: Example of a cosmoscope created with Cosma. A node from the graph is selected, which displays the corresponding card on the right. At the bottom of the card is a list of backlinks (incoming links from other cards). Hovering on a backlink reveals its context (the surrounding paragraph in the card it originates from).

Cosma is a hyperdocument creation system. The system-level design decisions have implications for the author design of these hyperdocuments. For instance, authors can decide to categorize records and links; if they do, they must create their own categories (there are no

defaults). Another example: contextualized backlinks display the surrounding text around incoming links. Using this feature can lead authors to adapt their writing in anticipation of the link context being retrieved and displayed as a memory aid—favoring e.g. dense paragraphs with multiple links, short sentences with just one link, bullet lists, etc.

The design process for these features was rooted in our theoretical focus on linking as a KOP. Going back to Zacklad’s terminology and model of information design, we analyzed linking based on document theory and knowledge organization literature. Linking can be made during writing, in which case it belongs to *textualization*, or after, in which case it belongs to *documentarization*. In both cases, it allows to *classify*, *index* or *tag* ideas. Linking belongs to *external documentarization* from the perspective of individual records but *internal documentarization* from the perspective of the document graph. And linking enables *composition* because it builds the graph, but also *fragmentation* because it provides the basis for extracting link contexts.

This analysis provided us with important design guidance at the system level: if the edges of a graph are intricately related to textualization and documentarization, then the navigational interface should revolve around knowledge organization features that facilitate the circulation of text in order to help create more documents. This leads to unexpected insights. For instance, in such an interface, information can be retrievable in the traditional sense but also emerge unsolicited, triggering serendipity (see examples in the margin). And over time, the anticipation of this information emergence can affect the way authors write. We observed something we called a “nexus” effect: knowing about the presence of contextualized backlinks encourages authors to express ideas as paragraphs which are densely packed with links, because such paragraphs create many paths back, not just to the card they belong to but to all cards that are linked in the same paragraph. A nexus is a textual context shared by several links, establishing a “many-to-many” relationship between entities in the graph. This vastly increases the opportunities for remembrance.

Does a document graph feel like a document?

This small experiment shows how system design can shape text, document and author. This in turn offers us some ground to discuss document design in relation with documentarity.

Connected by proxy—otherwise unrelated cards brought together in the backlinks of a card they all link to:

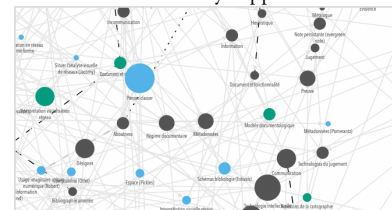
Tim Gorichanaz
 ID : 28218726171929
 TYPE : personne

▼ RETROLIENS 4

- Format
- Texte brut et stratégie
- Textuel textiel Corela
- Traitement de texte

Heim, M. (1987). *Electric language : A philosophical study of word processing*. Yale University Press. Recommandé par Tim Gorichanaz.

Unconnected but neighbors—proximity caused by the graph layout puts forgotten cards back in view by happenstance:



By the word documentarity, we refer to a quantifiable quality of documentary things. In French, *documentarité* is defined as *ce qui fait document*, which does not mean “what makes a document” but “what feels like a document.”¹⁶ The way we perceive documentarity is subjective, situational and multi-factorial, shaped by our “horizons of expectation”¹⁷, especially previous experience of genre-based rules; in the case of documents, these rules include closure, portability, structure, metadata, linearity and more.

To illustrate the subjectivity of documentarity, let us take an example. A document theorist walks into a zoological garden and sees an antelope in one of the enclosures. Recalling Briet’s famous example, they see the antelope as a document. However, most zoo visitors are not document theorists, and would probably feel that the display at the front of the enclosure (describing the species) conforms more to their experiences of what the word “document” means, rather than the animal.

Do Cosma’s cosmoscopes feel like documents? Let us first introduce a relevant comparison. Multi-page static websites or single-page web applications are not usually called documents; however, the digital notebooks used in data science are often called “computational documents”. Digital notebooks are texts containing code and its output, which the reader can modify by editing the code. They are document program hybrids, run in dedicated environments (such as Jupyter or RStudio). This is not very different from a complex web application accessed from a browser; in fact, some of the environments for running notebooks are repurposed web browsers. The decisive factors that affect documentarity here are linearity and file abstraction. Contrary to a website or web application, a notebook has a starting point and its text only flows in one direction. And while a website is not easily abstracted to a single computer file, a notebook can.

Like notebooks, Cosma’s cosmoscopes are document-program hybrids. They are more portable than notebooks but they are non-linear and have no obvious starting point. So their documentarity is arguably lower and the design task could be viewed as to increase it. This can be done for instance by adding metadata such as a title, displayed in the top left corner. But we can also consider a cosmoscope as a doceme,¹⁸ meaning something that can be either a standalone document or a fragment embedded in another document. A cosmoscope can be embedded in a webpage. Then maybe we need to decrease documentarity, for instance by removing the title we previously added, so

¹⁶ Perret and Le Deuff, “Documentarité et données, instrumentation d’un concept,” 2019.

¹⁷ “Horizons of expectation” (German: *Erwartungshorizont*) was used in literary theories of reception, most notably by Jauss (“Literary History as a Challenge to Literary Theory,” 1970). This is in the same neighborhood as the concept of “literarity” (German: *literarität*; French: *littérarité*) used by Jakobson (*Huit questions de poétique*, 1977), and which influenced the French meaning of documentarity—not to be confused with Day’s (*Documentarity*, 2019).

¹⁸ Lund, “Documentation in a complementary perspective,” 2004; “Building a discipline, creating a profession,” 2007.

that it appears simply as some kind of interactive visualization embedded in a document.

This quick thought experiment suggests to us that *designing documents implies playing with documentarity*, and that depending on the situation, it potentially means setting new horizons of expectation. We call for more research on this topic and specifically usage studies: if documentarity is quantifiable, it can be measured or at least assessed, if only to reveal its variability. Linearity may be the most important factor for documentarity, as it seems to compound the effect of other factors, such as structure and metadata; but this is an intuitive statement, which could be tested.

A relational approach to document theory

This exploratory work suggests that document design and document theory should be associated more often. They benefit each other: theory orients action, while practice strengthens concepts. In particular, we believe documentarity can be a bridge between the two fields, within a more general model of what documents do.

Our foray into document design has been based on a relational, structural approach to document theory: documents (or fragments of documents) plus links equals document graph. This approach can be taken further in order to think about document design in a more holistic way. A theoretical model of what documents do can be imagined in a relational way; we illustrate this with fig. 3.

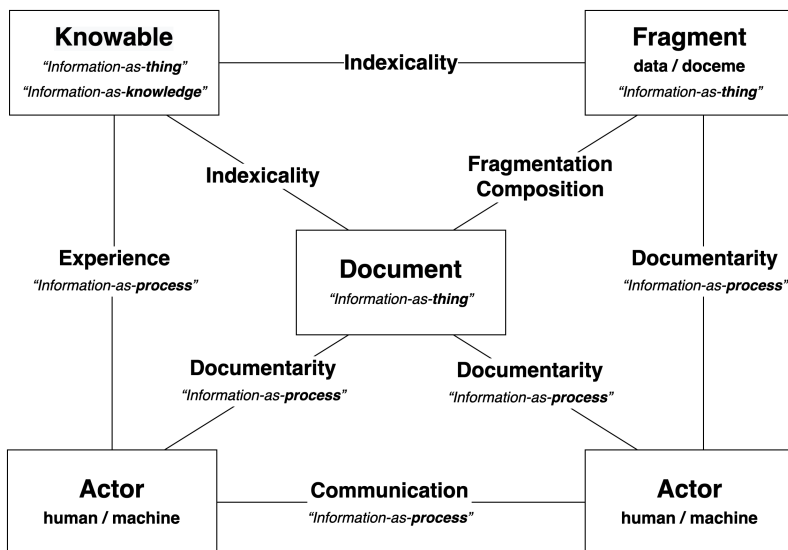


Figure 3: A relational model for document design based on document theory.

This is a diagram of *positions* rather than entities, considering that entities can find themselves in various positions depending on the situation. For instance, the actor in a communication process can also be a document at another time. This is an attempt to place several theoretical approaches in the same conceptual space, in order to provide guidance for design. We start with our structural approach to documents and add relations studied in the literature, such as indexicality¹⁹ and information experience.²⁰ Buckland²¹'s typology of information is used throughout to characterize both positions and relations.

The goal of this diagram is not to be the unifying theory of documentation. It is reductive, as all models are, and should be used (if at all) along with other resources. For instance, our approach addresses word-based documents; images, sounds and videos may require a different approach. Our intent here is to “package” theory into an actionable tool for word-based document design, so that we may go “from the concrete to the abstract and back again”, as Hjørland²² suggested: from an experiment to a theoretical model, to another experiment—which hopefully can reveal the strengths and limits of the model. This is very much a work in progress, on which we hope to report in the future.

¹⁹ Day, *Indexing it all*, 2014.

²⁰ Bruce, Davis, Hughes, et al. (dir.), *Information experience*, 2014.

²¹ “Information as thing,” 1991.

²² “Information Science and Its Core Concepts,” 2014, p. 231.

References

- ARRIBE, Thibaut. *Conception des chaînes éditoriales : documentariser l'activité et structurer le graphe documentaire pour améliorer la maîtrise de la rééditorialisation*. Thèse de doctorat. Compiègne : Université de technologie de Compiègne, 2014. <https://tel.archives-ouvertes.fr/tel-01127476>.
- BRUCE, DAVIS, HUGHES, PARTRIDGE and STOODLEY (dir.). *Information experience: approaches to theory and practice*. Bingley : Emerald, 2014. ISBN 978-1-78350-815-0.
- BUCKLAND, Michael. "Information as thing." *Journal of the American Society for Information Science*. 1991, Vol. 42, no. 5, p. 351-360. [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C351::AID-ASI5%3E3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3).
- CARLINER, Saul. "Current challenges of research in information design and document design." Dans : Carliner, Saul, Verckens, Jan Piet and Waele, Cathy de (dir.), *Information and Document Design: Varieties on Recent Research*. [S.l.] : John Benjamins Publishing Company, 2006, p. 1-24. Document Design Companion Series, 7. ISBN 978-90-272-3207-6. <https://doi.org/10.1075/ddcs.7.03car>.
- CARLINER, Saul. "Ten years after: An interview with Karen Schriver about the lasting impact of the Dynamics of Document Design." *Information Design Journal*. 2007, Vol. 15, no. 2, p. 163-170. <https://doi.org/10.1075/idj.15.2.09car>.
- DAY, Ronald E. *Documentarity: Evidence, Ontology, and Inscription*. Cambridge : MIT Press, 2019. ISBN 978-0-262-04320-5.
- DAY, Ronald E. *Indexing it all: the subject in the age of documentation, information, and data*. Cambridge : The MIT Press, 2014. ISBN 978-0-262-02821-9.
- ENGELBART, Douglas C. *Augmenting Human Intellect: A Conceptual Framework*. SRI Summary Report no.AFOSR-3223. Stanford : Stanford Research Institute, 1962. <https://www.doungengelbart.org/content/view/138/>.
- FINDELI, Alain. "La recherche-projet en design et la question de la question de recherche : essai de clarification conceptuelle." *Sciences du Design*. 2015, Vol. 1, no. 1, p. 45-57. <https://doi.org/10.3917/sdd.001.0045>.
- HJØRLAND, Birger. "Classification." *Knowledge Organization*. 2017, Vol. 44, no. 2, p. 97-128. <https://doi.org/10.5771/0943-7444-2017-2-97>.
- HJØRLAND, Birger. "Indexing: Concepts and Theory." *Knowledge Organization*. 2018, Vol. 45, no. 7, p. 609-639. <https://doi.org/10.5771/0943-7444-2018-7-609>.
- HJØRLAND, Birger. "Information Science and Its Core Concepts: Levels of Disagreement." Dans : Ibekwe-SanJuan, Fidelia and Dousa, Thomas (dir.), *Theories of Information, Communication and Knowledge*. [S.l.] : Springer, 2014, p. 205-235. ISBN 978-94-007-6972-4. https://doi.org/10.1007/978-94-007-6973-1_9.
- JAKOBSON, Roman. *Huit questions de poétique*. Paris : Éd. du Seuil, 1977. ISBN 978-2-02-004680-0.
- JAUSS, Hans Robert. "Literary History as a Challenge to Literary Theory." *New Literary History*. 1970, Vol. 2, no. 1, p. 7-37. <https://doi.org/10.2307/468585>. JSTOR.
- LE DEUFF, Olivier and PERRET, Arthur. "Hyperdocumentation: origin and evolution of a concept." *Journal of Documentation*. 2019, Vol. 75, no. 6. <https://doi.org/10.1108/JD-03-2019-0053>.

- LUND, Niels W. "Building a discipline, creating a profession: an essay on the childhood of 'dokvit'." Dans : Skare, Roswitha, Lund, Niels W and Vårheim, Andreas (dir.), *A document (re)turn: contributions from a research field in transition*. Frankfurt am Main : Peter Lang, 2007, p. 11–26. ISBN 978-3-631-56294-9. <https://hdl.handle.net/10037/966>.
- LUND, Niels W. "Documentation in a complementary perspective." Dans : Rayward, W. Boyd (dir.), *Aware and responsible : papers of the 2001 Nordic-International Colloquium on Social and Cultural Awareness and Responsibility in Library, Information, and Documentation Studies (SCARLID)*. Lanham : Scarecrow Press, 2004, p. 93–102. https://www.academia.edu/en/980627/Documentation_in_a_complementary_perspective.
- MATUSCHAK, Andy and NIELSEN, Michael. *How can we develop transformative tools for thought?* 2019. <https://numinous productions/ttft/>.
- PERRET, Arthur. *De l'héritage épistémologique de Paul Otlet à une théorie relationnelle de l'organisation des connaissances*. Thèse de doctorat. Bordeaux : Université Bordeaux Montaigne, 2022. <https://theses.hal.science/tel-03954967>.
- PERRET, Arthur and LE DEUFF, Olivier. "Documentarité et données, instrumentation d'un concept." Dans : *12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?* 2019. <https://hal.archives-ouvertes.fr/hal-02307039>.
- PETTERSSON, Rune. *Information design: an introduction*. Amsterdam; Philadelphia : John Benjamins Publishing Company, 2002. Document design companion series, 3. ISBN 978-1-58811-338-2.
- RAFFERTY, Pauline. "Tagging." *Knowledge Organization*. 2018, Vol. 45, no. 6, p. 500–516. <https://doi.org/10.5771/0943-7444-2018-6-500>.
- RENKEMA, Jan. "Editorial." *Information Design Journal*. 2006, Vol. 14, no. 1, p. 1–3. <https://doi.org/10.1075/idj.14.1.01edi>.
- SCHRIVER, Karen A. *Dynamics in document design*. New York : Wiley, 1997. ISBN 978-0-471-30636-8. <http://archive.org/details/dynamicsindocume0000schr>.
- ZACKLAD, Manuel. "Information Design: Textualization, Documentarization, Auctorialization." *Proceedings from the Document Academy*. 2019, Vol. 6, no. 1. <https://doi.org/10.35492/docam/6/1/2>.