# We Know What You Read Last Summer

**Svenja Hagenhoff**
**Friedrich-Alexander-Universität Erlangen-Nürnberg**

## Overall research topic: regulation of media communication

**Main questions**

- Should we regulate media communication?
- What are the areas of regulation?
- What are the motives and arguments for regulation?
- What are adequate and targeted instruments of regulation?

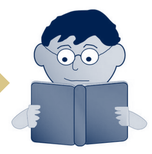
**Areas of regulation**

**Indirect regulations**

Control the communicative act indirectly via structural arrangements that enable or prevent (economic) interactions. Their nature affects whether, how and when media goods can be produced, distributed or consumed.

Regulation parameters: Pricing mechanisms, property rights, market access, product quality, conditions, obligation to contract

**Author** → **Publisher** → **Book Trade**



**Communication of content (ideas, knowledge, information,…)**

**Direct regulations**

Control the communicative act directly by authorities prohibiting, demanding or privileging the communication of certain contents

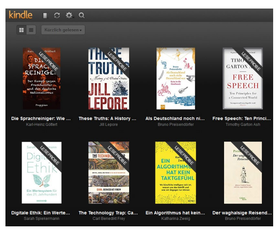
Examples: Censorship, protection of young persons, personal rights

**Typology of institutionalization** (see Saxer 2002)

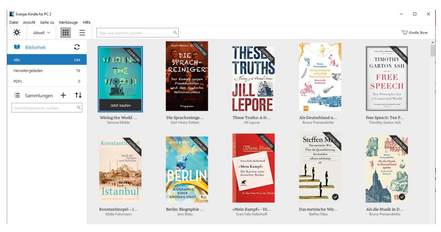
- **Totalitarian:** Absolute and effective control over media communication in all phases and forms by the ruling party
  *Example: China, North Korea, Saudi Arabia*
- **Authoritarian:** Moderately undemocratic. What the elites like politically and culturally is published. But the effectiveness often is actually poor
  *Example: Mexico, India, Russia, Turkey*
- **Democratically controlled:** Social responsibility is attributed to the media. Concern: economic power could turn into opinion or cultural concentration. Hence: equal opportunities in publishing instead of economic freedom. But: different rules for different media segments or forms (e.g. books vs. TV)
  *Example: Austria, Germany, United Kingdom*
- **Liberal:** No special media control apart from the general legal standards. Media communication may even enjoy privileges of freedom. Belief in compensatory justice and in the needs-based appropriateness of market mechanisms
  *Example: USA*

## Reader analytics: idea

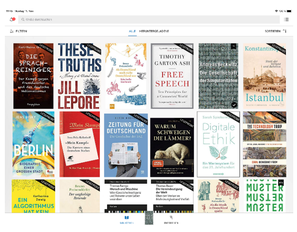
- Reader analytics = observed reading of digital media, esp. e-books
- To use an e-book a technical reading environment is needed
- Within this environment data is captured which describes the recipient's book usage



Web reader (Browser)



Desktop client



App

See Arndt et al. 2019; Hagenhoff 2019

## Two forces within the reader analytics context

**Reading as a communicative and social practice** (see Kuhn 2019, Arndt et al. 2019)

- Reading represents a social total phenomenon that permeates all levels of society and culture with decisive influence on identities, relationships between subjects, communities, and social structures
- There is a tension between the subject's desire for *privacy* to become an individual and the desire and need to *participate in social interactions*

**Data as an economic resource** (see Hagenhoff 2019)

- For companies the »world« is represented by data: control and planning of entrepreneurial activities is data based (size of markets, properties of customers,…)
- Typically, the book industry does not know very much about it's readers and their reading behavior. Reader analytics could be a chance to learn more about »reality«

## Kinds of data captured in reader analytics processes

**Data related to the individual reader based on conscious activities**

- Bookmark
- Marked passages
- Annotated passages
- Recommend the book
- Share reading progress

**Data related to the individual reader via (opaque) tracking of activities**

- Day and time of reading
- Length of each reading session
- End point of reading session
- Session distance (time between to sessions = non reading time)
- Used reading environments & devices

**Aggregated data**

- How many readers finish the book?
- Average reading speed
- Recommendation factor
- Conversion: title information ▶ reading sample ▶ read / bought
- Age and gender
- Most read title

## Empirical analysis of Kindle and Tolino (Oct – Dec 2018)

**What did we want to know and what did we analyze?**

- Which *persistent* traces occur locally on the examined system (e.g. on the hard disk)?
- What *transient* traces are created when the system is used in the network?
- Environments analyzed : Web apps Kindle & Tolino; Client app Kindle Windows; Apps Kindle & Tolino for Android

**Findings Kindle in a nutshell**

- The entire reading behavior is recorded
- Data is also collected that is not necessary for the usability of the applications (e.g. reading speed per page and for the entire book)
- Extensive, short-stacked data traffic between client and server

**Findings Tolino in a nutshell**

- Tolino collects less data than Kindle
- Traffic between client and server only takes place when it is required functionally, and only when the user is logged in

## Cited literature

Arndt, S.; Benenson, Z.; […]; Hagenhoff, S. […] et al. (2019): Analyse verbreiteter Anwendungen zum Lesen von elektronischen Büchern. Technical Reports. Hg. v. Department Informatik. Friederich-Alexander-Universität Erlangen-Nürnberg. Erlangen (in print)

Hagenhoff, Svenja (2019): 'Publishing is not a job anymore, it's a button': Algorithms as resources in value creation structures in the arena of public communication via media. In: Chr. Schumann & M. Taddicken (Hg.): Algorithms and Communication. Berlin (in print).

Kuhn, A. (2019): Reader Analytics: Vom privaten zum öffentlichen Lesen? In: Chr. Aldenhoff et al. (Ed.): Digitalität und Privatheit. Kulturelle, politisch-rechtliche und soziale Perspektiven, S. 263–281.

Saxer, U. (2002): Der gesellschaftliche Ort der Massenkommunikation. In: H. Haas & O. Jarren (Ed.): Mediensysteme im Wandel. Struktur, Organisation und Funktion der Massenmedien. 3. Auflage, Wien, S. 1–14.