



PatientsLikeMe.com

Developing medical research through social data



Niccolò Tempini – Information Systems and Innovation Group, LSE – N.Tempini@lse.ac.uk, @tmpncl

ABSTRACT

PatientsLikeMe is an organization that is developing a social networking platform with the aim of collecting patients’ self-reported medical data. The organization exploits the resulting data for research purposes. *How does PatientsLikeMe manage to produce real-world medical evidence from the data voluntarily contributed by dispersed patients?*

In order to be able to produce information, the organization needs to cater to the needs of a myriad of local contexts while still collecting data in sufficient quantity and richness. I theorize **two mechanisms of information cultivation** that elicit two different types of informative associations. I link the challenges I have documented to the structural configuration of the data collection arrangement, suggesting that these might highlight implications for broader organizational domains.

BACKGROUND

Case study context

- PatientsLikeMe is a for-profit company that has built and is continuing to develop an ad-free social networking platform with the aim of connecting patients while collecting self-reported medical data.
- The data from 220000+ patients informed researchers producing peer-reviewed articles, conference papers, reports. Research results included virtual clinical trials (published on *Nature Biotechnology*), perceived benefits of drugs, etc.

Theoretical framework

- I apply Bateson’s definition of information (event definition of information). Information is “a difference that makes a difference”.
 - Signs (data) are a difference (mark) that is informative if it makes a difference, i.e. if it has a meaning for someone or something.
- I build on Jacob’s analysis of the functional role of data structures:
 - Data structures that are more structured and specific have higher semantic context, i.e. they embed more information within the structure itself.
 - However, the more data models are structured the less flexible they are to local contexts, i.e. they are more difficult to adopt.

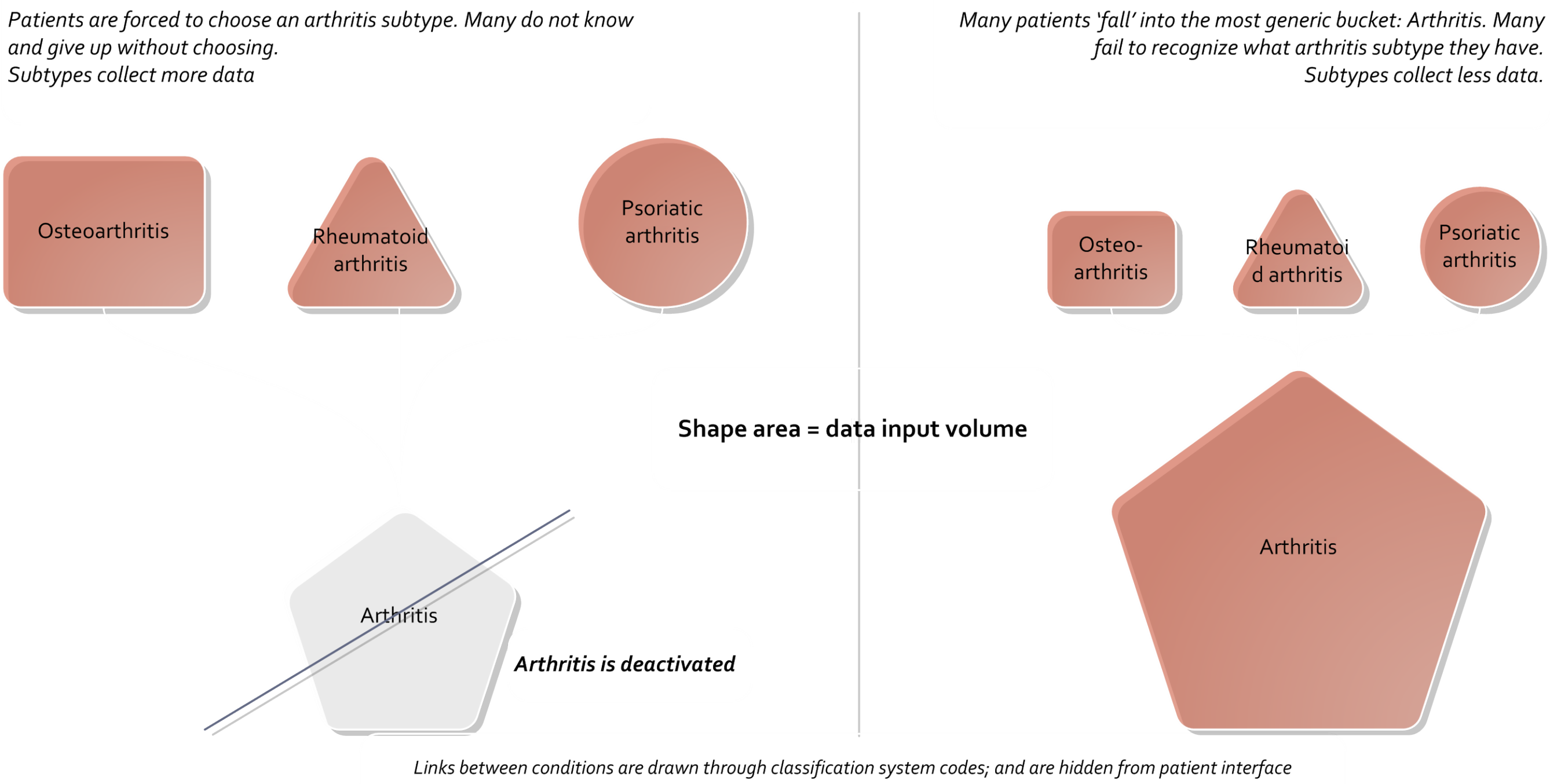
The **informative potential** of data collected from distributed patients is **function** of the **amount** and the **specificity** of the data.

METHODS

- Observational case study, embedded in PatientsLikeMe.
- Member of the R&D and Health Data Integrity teams.
- Retroductive reasoning as technique for formulating explanations.
- Triangulation of various data sources for validating tentative interpretations.

| Evidence | |
|--|---------------------------------|
| Participant observation | 26 weeks full-time office hours |
| Interviews (60 min avg. duration) | 30 |
| Other recordings (meetings, conversations) | 8 |
| Notes (snapshots, conversations, analytical reflections) | 665 |
| Meeting notes | 128 |
| E-mails (automated notifications) | 4547(2877) |

RESULTS



PatientsLikeMe faces a considerable challenge in developing an information system for **producing information** out of **data collected by distributed patients**.

Higher patient engagement is associated with an increase in the information potential of the collected data, by increasing its scale in terms of both sample size and longitude. However, **the system collects more data but these data are, taken individually, less meaningful**.

Conversely, **higher semantic context** is associated with an increase in the information potential of the collected data, by increasing its power to differentiate between phenomena. However, **the system collects more meaningful data but these data are, as a whole, fewer**.

The complexity of the problem domain (medicine) leads the company to iterate its approach to data modelling problems, as they unfold. At times, collecting **enough relatively vague data** has been **prioritized over** collecting **precise data in scarce quantity**. Conversely, **in other situations the opposite applied**.

RESULTS

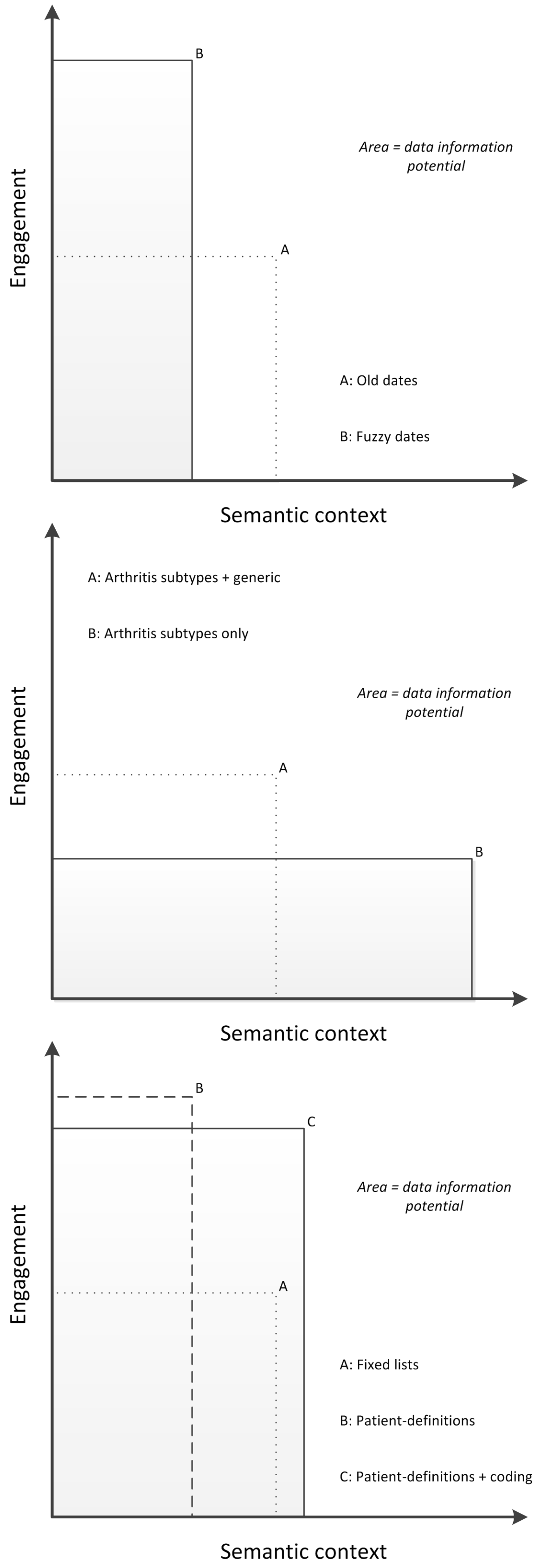
Two mechanisms of **information cultivation**:

- **Data pool extension**: information emerging from *across data models* is more likely to be expressed when the data is collected in large scale.
- **Data pool enrichment**: information emerging from *within data models* is more likely to be expressed when the data are specific.
- Mechanisms of information cultivation are **related in paradox**. Both increase the information potential of the data by strengthening one of the two factors of information production – scale and specificity – at the expense of the other.

PatientsLikeMe tries to make medical research through an **open, purely distributed** and **data-based** data collection approach.

The patient is the gateway to an experiential context of complex factors, signs and variables, that the organization cannot reach in any other way.

At the same time, PatientsLikeMe is the space that shapes the lines of connection with potentially similar others.



CONCLUSIONS

What the presented evidence suggests is that **the construction of knowledge from social data** still has **important and paradoxical challenges** to face.

Results show that the **structural configuration** of this innovative approach to knowledge production **shapes the kind evidence that is produced**. Online community management and all it involves shapes scientific results that are thus constructed.

The underlying arrangements which produce what some would perhaps call ‘**raw data**’ are very much **shifting and contingent**.

ESSENTIAL REFERENCES

- Bateson, G. 1972. *Steps to an Ecology of Mind*. London: University of Chicago Press.
- Bowker, G. C. and S. L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. London: MIT Press.
- Jacob, E. K. 2004. Classification and Categorization: A Difference that Makes a Difference. *Library Trends* 52(3): 515–540.