**Title:** The shape of data: Manifolds, Cities and Biodiveristy

**Author(s):** Thilo Gross (123)

**Affiliation(s):** Helmholtz Institute for functional marine biodiversity (1); University of Oldenburg, Institute for Chemistry and Biology of the Marine Environement (2); Alfred-Wegener Institute, Helmholtz Centre for Marine and Polar research (3)

**Abstract:** The aim of this talk is to highlight the need for a geometrical understanding of datasets as an important challenge at the interface of complexity and data science. Consider a dataset that describes a collection of objects. Each object then corresponds to one point in the, typically high-dimensional data-space, spanned by the dimensions recorded in a dataset. For example in a dataset on lakes, each lake could be a datapoint, and the length, width and depth of the lake could be three of the axes that span the data-space. In general the datapoints do not fill the dataspace uniformly. Instead they typically lie on realtively thin, effectively low-dimensional manifolds. Thus for example the lakes in our example reveal the manifold of plausible lakes. All points correspond to lakes that could very well exist in the real world while points that are not on the manifold would likley not be recognied as lakes even if they were encountered in reality.

In this talk I am going to present how the geometric exploration of data manifolds can yield interesting insights into datasets. In particular I will consider data on cities and ecosystems to demonstrate that the natural coordinate system implied by the shape of the data manifolds yield coordinates that have an intuitive interpretation.

**Keywords**: Manifold Learning, Urban Structure, Biodiversity.

**Author Profiles (s):** exTwitter: [@thilogross](https://twitter.com/thilogross) ; homepage: [biond.org](https://biond.org) ; scholar: <eogyTQUAAAAJ> ; LinkedIn: [thilo-gross-75b9a86](https://www.linkedin.com/in/thilo-gross-75b9a86/%20) ; youtube: [@complexitypapers](https://youtube.com/%40complexitypapers)