Why are vocabularies important for research?

Controlled or standardised vocabularies are an important part of research and scholarly communication since these rely on precise concepts with shared and structured terminology. The ability to replicate and test an experiment or communicate and verify a conclusion requires clear description and communication of concepts about which there is a shared understanding of meaning.
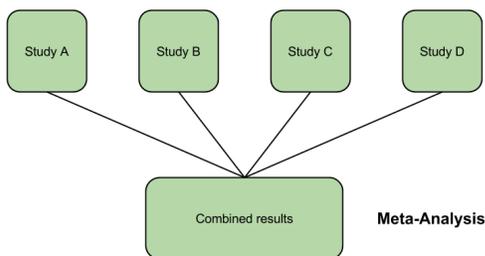
As an example, palaeontologists use an agreed vocabulary covering time periods (ucmp.berkeley.edu, 2015). This enables them to refer to periods of time in the knowledge that they agree, for example, on how the Phanerozoic Eon relates to the Cenozoic Era.

Not only in research but also in many areas of life, vocabularies are used. Medical doctors use software that incorporates vocabulary terms and definitions. They need to be able to make very precise observations about symptoms presented by a patient in order to select appropriate medicines. When using such software, the GP will not typically type in these observations but will select terms through auto-complete or point-and-click. The choices are controlled, based on a controlled vocabulary. Australia is attempting to gain agreement on sets of terms, which would enable the implementation of shared electronic patient records (Nehta.gov.au, 2015). Authorities also deal in vocabularies. In the United States of America, the Federal Highway Administration (FHWY) has an extensive authoritative classification of vehicles. This includes text descriptions and images aimed at clearly conveying what is meant by terms describing a wide range of vehicles (Onlinemanuals.txdot.gov, 2015). Vocabularies are a natural output of such authorities. In the case of the FHWY, the classification is typically used to support charging on toll roads, but such classifications also have applications to research.

Vocabularies are used ubiquitously in information systems. Consumer sites such as Amazon are structured using controlled terms (Amazon.com, 2015). An item such as a travel wallet is categorised within a hierarchical tree that starts at Clothing, Shoes & Jewelry Luggage & Travel Gear Travel Accessories Travel Wallets. Such categorisations are enabled by controlled vocabularies and are often a complement to a text-based search or indeed combined with search in a "faceted" or "filtered" search. Retail websites increasingly cater for both browse and search.

University of California Museum of Paleontology Geologic time scale (http://www.ucmp.berkeley.edu/help/timeform.php)

Vocabularies are also important in enabling data reuse. As a simple example, a tabular dataset in a spreadsheet would typically contain column headings describing the content of each table. A third party wishing to use this data would need to know what those headings mean in order to make sense of the data. To support data reuse, the data creator may supply a data dictionary to accompany the dataset. Controlled terminology is also vital if seeking to relate datasets to each other, whether this is a relatively simple join across two datasets, or a meta-analysis involving the bringing together of multiple datasets which may have been created at different time periods or geographic regions. Wherever datasets are linked or merged, the connections need to be made at points that are known to be common.