

# Seven dimensions of portability

## What is portability?

**From Computer Science:** the usability of the same software in different environments. Here they say they are talking about portability of data: the usability of data in different environments.

## Why was this article written, and why is it important?

In 2003 there were:

- New general tools (word processors, hypertext processors, database packages).
- New specialized tools (Shoebox, Praat, Transcriber...).
- New specialized technology (recording devices, storage devices).

**Problem:** specialized work flows, arcane instructions for access, risk of loss of data and info at every level.

B&S identify seven problem areas or “dimensions” (with sub-dimensions) for portability in linguistics. They posit discipline-wide value statements about these dimensions, and provide recommendations for Best Practices. Readers encouraged to suggest alternate BP, or alternate values.

### 1. Content

#### (a) Coverage:

- i. We value comprehensive documentation, broad in scope and rich in detail.

**BP:** make rich records of rich interactions; document the field methods used.

#### (b) Accountability

- i. We value the ability to verify linguistic claims.

**BP:** make the full documentation available (a grammar is based on a text corpus); provide primary recordings for transcribed texts; time-align transcriptions to recordings; when a recording has been edited, provide the original.

#### (c) Terminology

- i. We value the ability to compare terminology in different resources.

**BP:** Map the underlying terminology/tags/transcription symbols to a standardized ontology (GOLD, IPA).

### 2. Format

#### (a) Openness

- i. We value the ability to make use of a resource without the need for unique or proprietary software.

**BP:** Store langdoc and description in open formats (published, nonproprietary specifications); prefer formats supported by multiple software; prefer formats supported by free tools; prefer published proprietary formats over secret proprietary formats.

#### (b) Encoding

- i. We value a character encoding that is not limited by the font used to render it.

**BP:** Use Unicode; avoid Private Use characters (or document them well); document any scheme for transliterating.

- (c) Markup
  - i. We value the potential to write new software for processing extant data in novel ways.  
**BP:** Use descriptive (not presentational) markup; use XML whenever possible with a schema or DTD; prepare and archive an explanatory document if you use some other descriptive markup.
- (d) Rendering
  - i. We value the ability to read content in a conventional way.  
**BP:** provide and archive human readable versions of your materials using common formats (HTML, txt, pdf, paper).
- 3. Discovery
  - (a) Existence
    - i. We value the ability of any potential user to learn about the existence of a resource.  
**BP:** Archive your materials in an OLAC archive; make HTML easy to find via keywords.
  - (b) Relevance
    - i. We value the ability of a potential user to judge the relevance of a resource without having to first obtain a copy.  
**BP:** Use good descriptive metadata (e.g. OLAC metadata set).
- 4. Access
  - (a) Scope
    - i. We value the ability to access a complete resource, not just a part of it.  
**BP:** Publish complete primary documentation and a method by which users can obtain it.
  - (b) Process
    - i. We value making it easy for users to gain access to resources.  
**BP:** document the process for access as part of the metadata.
  - (c) Ease
    - i. We value users being able to access resources wherever the users may be located, with whatever computer infrastructure.  
**BP:** Make CDs/DVDs available; make low-bandwidth surrogates (e.g. mp3) available online; provide print versions for the speech community with little computer access.
- 5. Citation
  - (a) Bibliography
    - i. We value being able to give credit to the creator of resources.  
**BP:** in the metadata, instruct users how to cite the resource.
  - (b) Persistence
    - i. We value the ability to locate resources even when their actual locations or filenames change.  
**BP:** Ensure that items have a persistent identifier (ISBN, DOI); ensure that the identifier resolves (=points to) either an instance of the resource or directions on how to obtain the resource.
  - (c) Immutability
    - i. We value citing versions of resources that never change.  
**BP:** distinguish versions of a resource with a distinct identifier.
  - (d) Granularity
    - i. We value being able to cite component parts of a resource.  
**BP:** provide a way to cite a part of a resource (eg timestamps).

## 6. Preservation

### (a) Longevity

- i. We value ongoing access to resources over the long term.

**BP:** Use a credible archive; digitize analog materials; migrate offline materials regularly; archive physical versions of the materials.

### (b) Safety

- i. We value ongoing access to resources over the long term.

**BP:** Ensure LOCKSS: Lots of copies keeps stuff safe; create a disaster recovery plan.

### (c) Media

- i. We value access to a resource beyond the lifespan of any particular storage medium.

**BP:** use an archive with well-maintained servers and a commitment to migrate to new media; transfer your offline data at regular intervals to new media.

## 7. Rights

### (a) Terms of use

- i. We value easy to understand restrictions on use of resources.

**BP:** fully document terms of use, including specifics of how an item may or may not be used.

### (b) Benefit

- i. We value maximal application of a resource toward the benefit of human knowledge and experience.

**BP:** Ensure that resource can be used for research and is not limited to the use of the researcher, project or agency responsible for collecting it.

### (c) Sensitivity

- i. We value the rights of the speaker community.

**BP:** Document any sensitivities in detail, and include a list of any uses that must be avoided.

### (d) Balance

- i. We value the long-term benefit of a resource, even when access has to be restricted in the short term.

**BP:** limit stipulations of sensitivity to the sensitive parts only; associate each sensitivity with an expiration or review date; when sensitivity is only for the benefit of the researcher, the expiration date should be no more than five years.