Research resources and standardization : in the digital age

Akira MIYAZAWA

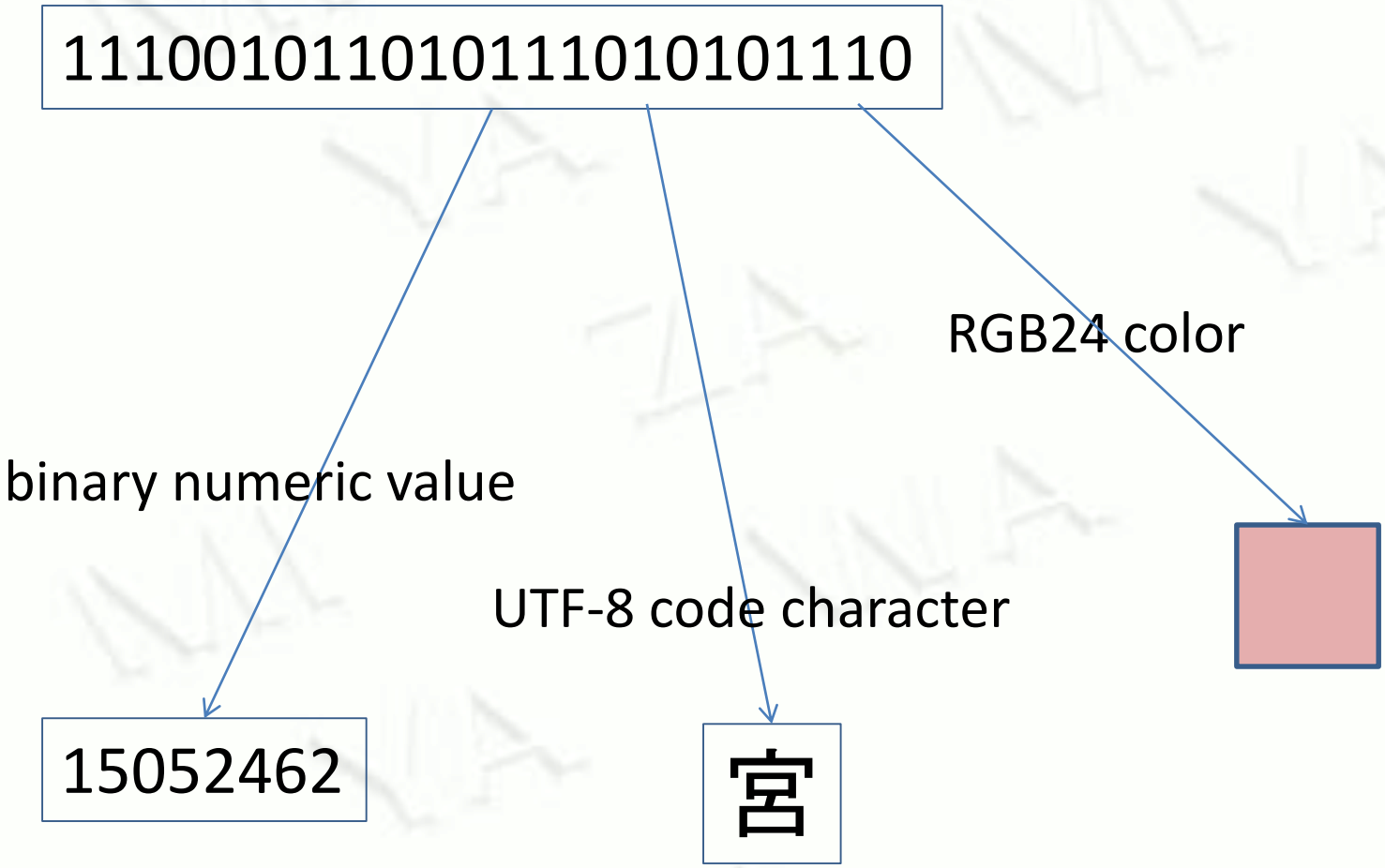
Prof. emeritus NII

Research resources

- books
- documents
- artworks
- statistical data
- experiment data
- journal articles

Originals have to be regarded, but digital copies are used in most cases.

Digital data



Data is language

My theory

- both are means of communication
  - between human beings,
  - between computers and human beings, or
  - between computers.
- But when a data format is just for one application program,
  - it 's like communication between a man and his pet.
  - There's little point in consideration.
  - There need considerable number of speakers (programs that process the data).

Is an image a data language?

Yes. It is a set of statements like:

The pixel at position (0, 0) has color .

The pixel at position (0, 1) has color .

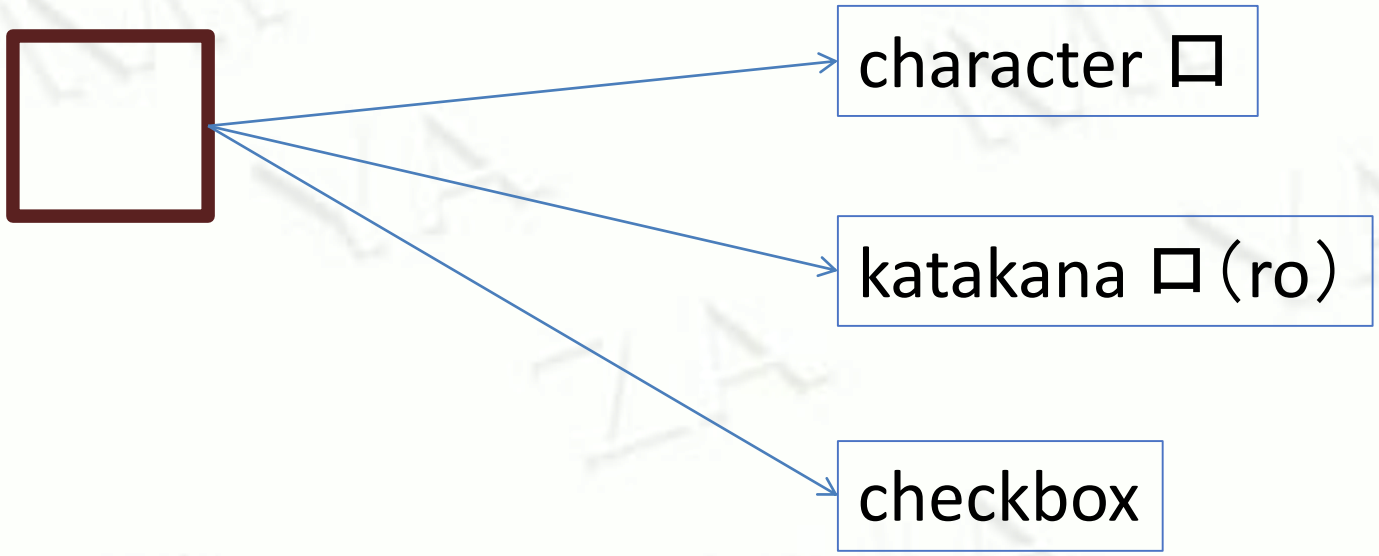
.....

The pixel at position (4, 8) has color .

.....

The pixel at position (7, 11) has color .

Arbitrariness (Saussure)



There is no necessity for the connection between the shape and the character.

A lot of rules, sometimes called language norms, work for the connection.

Data and standards

- The connection between a bit string and its interpretation has same arbitrariness.
- Therefore a lot of *standards* is necessary to interpret bit strings as a meaningful information.
- There are many other similarities between data and language.

Categorization of standards

- Measurement standards
- Terminology standards
- Standard codes
- Standard identifiers
- Evaluation and testing methods standards
- Procedural standards
- Management systems standards

Codes and Identifiers

Examples:

- ISO 3166-1 Country code
- ISO 2108 ISBN

Essentially, codes and identifiers are same.

Usually, codes are short and are given to relatively static set of things.

On the other hand, identifiers are long and are given to growing set of things.

Other types

Evaluation, testing methods

— mainly in industrial fields (chemical, material, ...)

Procedural standards

— e.g. ISO 15489 Records management

Management system standards

— e.g. ISO 9000 series: Quality management

Standardization organizations
examples

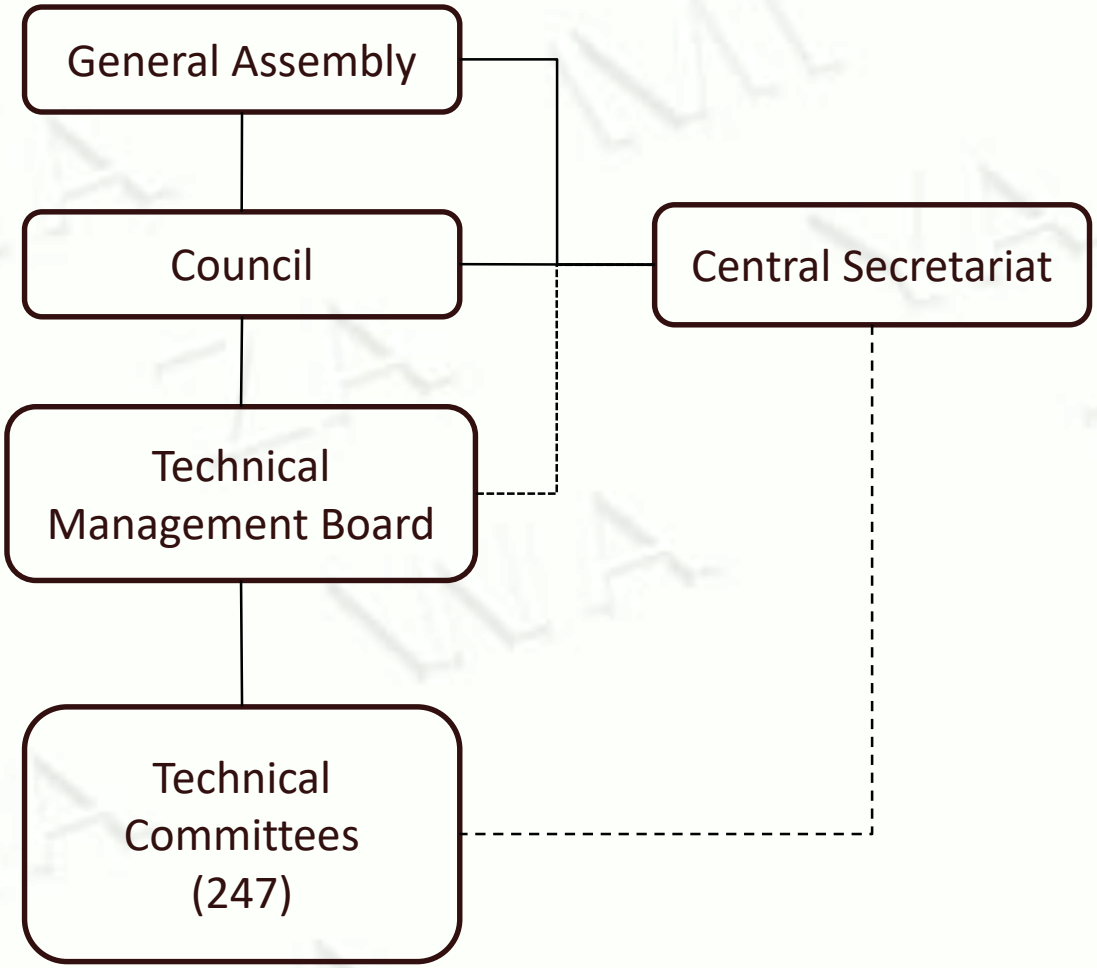
- ISO: International Organization for Standardization
  - The biggest and the most wide-ranging organization
- JISC: Japanese Industrial Standards Committee
- CEN: European Committee for Standardization
- IETF: Internet Engineering Task Force
- W3C: World Wide Web Consortium

ISO

- An NGO having about 160 member countries
- Membership of *national standards bodies*

  JISC (Japan), ANSI (US), DIN (Germany), ...
- 22000 < published standards
- 1500 < publications per a year
- 4700 < on going projects
- 20 < meetings per day

Structure of ISO



Technical Committees

TC1 Screw threads

TC2 Fasteners

TC6 Paper, board and pulps

ISO 216 Writing paper and certain classes of printed matter ... (A4 paper etc.)

ISO/IEC **JTC1** Information technology

ISO 10646 Universal Coded Character Set (UCS)

TC46 Information and documentation

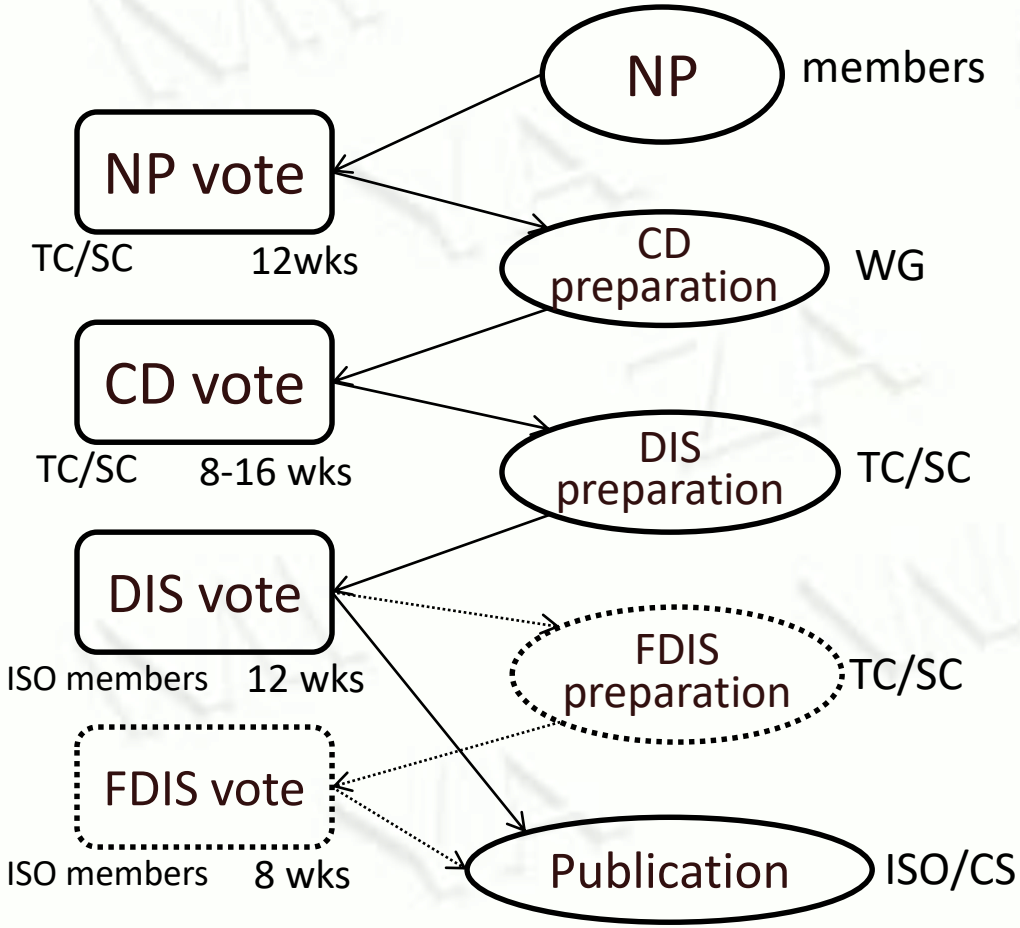
TC46 Information and documentation

41 P-members, 31 O-members

Secretariat: AFNOR (France)

- SC4: Technical interoperability
- SC8: Quality - Statistics and performance evaluation
- SC9: Identification and description
- SC10: Requirements for document storage and conditions for preservation
- SC11: Archives/records management
- WG2: Coding of country names and related entities
- WG3: Conversion of written languages

Typical standard development track



NP: New work item proposal

CD: Committee draft

DIS: Draft international standard

FDIS: Final draft international standard

Issues on standardization

- Economic base of standard development
- Standard bloat
- Maintenance costs of standards
- Usage costs

Economic base

ISO and most of national standardization bodies sell standards in book forms as their mane source of income.

But, the sales of standard books is decreasing.

Standard bloat

Number of standards are increasing.

Volume per standard is growing.

- Revision of a standard usually increase the size.
- Recent standards sometimes have several hundred pages, or more -- exceeding a thousand!

Maintenance costs for standardization bodies

Standards have to be reviewed periodically;

e.g. every 5 years

22,000 published standards require more than 4000 have to be reviewed per year (> 12/day)

Member bodies have to take on the workload of reviewing.

Usage costs for users

When going into a new business, a user may have to check several standards in the field. It may cost several hundred Euros.

Sometimes, they are very complicated, and you may need some consultants to keep compliance with the standards.
(like you need lawyers for legal matters).

Issues on digital resources

- Digital preservation
- Digital fakes

Digital preservation

Storage media, e.g. disks, may be damaged.

Software environment changes and an old version of a software, often, does not work on the new environment.

Migration and emulation are current solutions. But, ...

Digital fakes

- It is very hard to judge a digital object is a fake or genuine.
- It is not so serious to date.
- But, the technology is progressing. It will be easier to make counterfeits in future.

Thank you