What is the World Wide Web?

- **The World Wide Web**:
  - Is a distributed information system that provides access to hypertext documents and other objects of interest.
    - Invented by Tim Berners-Lee back in 1989.
    - Home of the first website: http://info.cern.ch/
  - **Resources** (or Web resources) is the standardized term for referring to objects of interest available on the Web.
  - The Web relies on the **client-server** communication model.

The Architectural Bases of the Web

- The Web is centered around the notion of a ***resource***.

- The Web defines three guiding principles for working with resources:
  1. Identification
  2. Representation
  3. Interaction (or operations on resources)

  - **Reference:** Architecture of the World Wide Web, Volume One https://www.w3.org/TR/2004/REC-webarch-20041215/

What Is a Resource?

- A resource is anything that is important enough to be referenced as an entity (or a thing).

- A Web resource is the source of ***Web content***.
  - It is something that can be stored on a computer and represented as a stream of bytes.

- **Two types of resources:**

1. **Static resources:** are files stored on the Web server's filesystem.
   - E.g., HTML, CSS, JavaScript, JPEG, PNG, SVG files.
2. **Dynamic resources:** are data content generated on demand by programs or scripts.
   - A row in a database, result of Web search engine, data pulled from a database, result of running an algorithm, etc.

The Web Architecture

Key Components:

1 **Identification:**

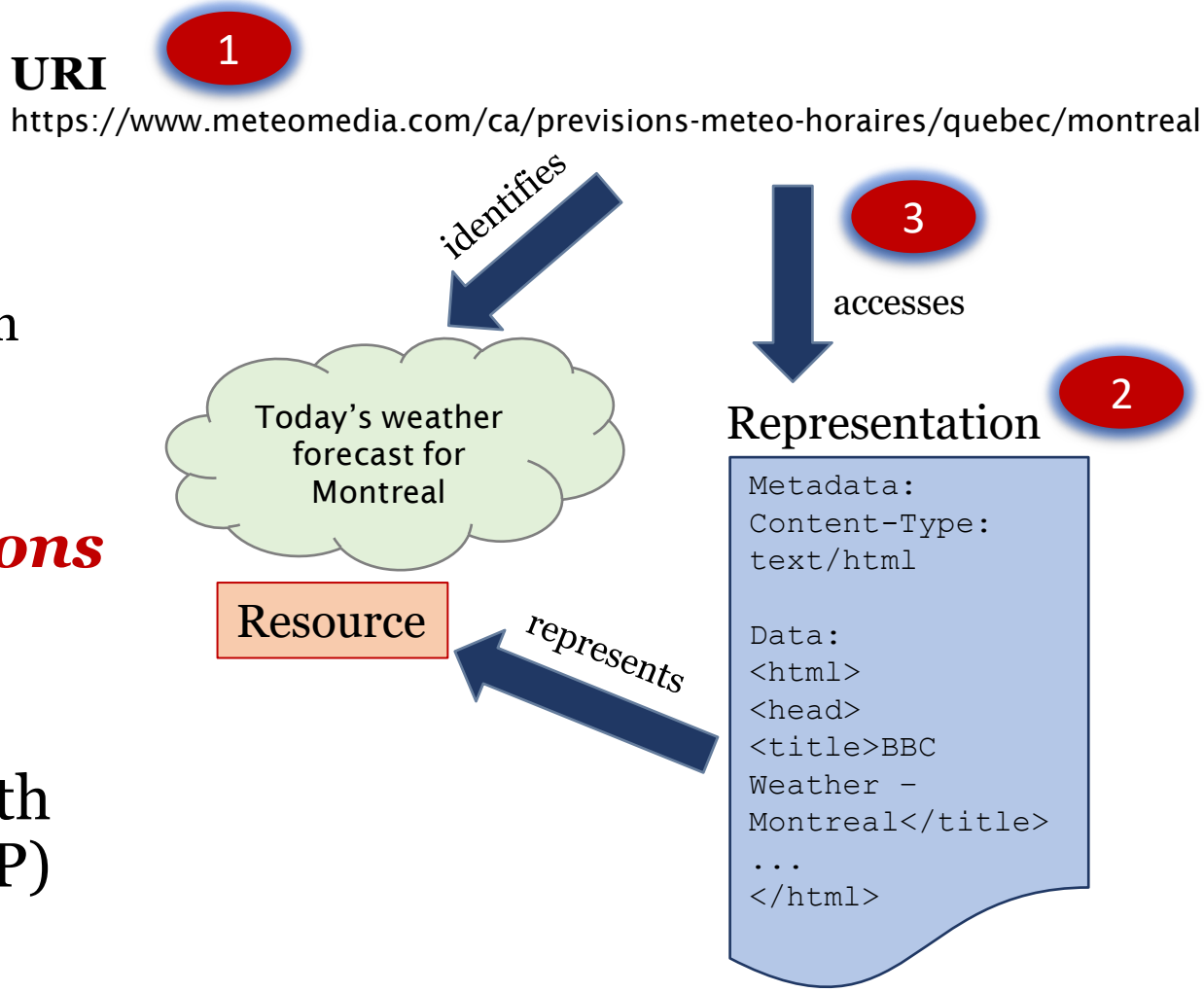
- Resources are ***identified*** by **U**niform **R**esource **I**dentifiers (URIs)

2 **Formats (representations):**

- Resources have ***representations*** in different formats

3 **Interaction**

- Resources can be interacted with using network protocols (HTTP)



Representation

- A representation is data that encodes information about resource state.
  - The type of data is specified using **Media Types**.
  - A resource representation consists of: **data content & media type**

- A resource can have more than one representation.
  - Documents available in multiple languages.
  - A collection of items can be represented in JSON or XML-based data formats.

- Representations have metadata
  - When were they last modified (used for caching).
  - What format are they in?

- Representations of a resource are sent or received using interaction protocols.

Representation (cont'd)

- Internet Media Types, formally known as Multipurpose Internet Mail Extensions (MIME type), are used to indicate the nature and format of a document, file or assortment of bytes.
  - Media Types are defined and standardized in IETF's RFC 6838
  - IANA is responsible for all official Media Types.
  - The most up-to-date and complete list can be found at IANA's Media Types page.

- Media types have standardized structure.

- A media type is a ***two-part identifier*** for file formats and format of Web content exchanged on the Web.

Representation (cont'd)

Media Types: Structure and Examples

- Media types are hierarchical descriptions of data types.

Structure A media type consists of two parts: a ***type*** and a ***subtype*** separated by a slash (/)

1. The ***type*** represents a general category (top-level types) into which the data type falls.
   - E.g., text, image, application, audio, etc.

```
type/subtype;parameter=value
```

2. The ***subtype*** identifies the exact kind of data the specified Media type intended to represent.
   - text/plain, text/html, text/css, text/javascript, etc.
   - image/jpeg, image/png, image/gif
   - application/json, application/pdf, application/xml

NOTE Optional parameters can be added to provide metadata (additional details): e.g., a charset can be added to text/plain;charset=UTF-8

Identification

- Uniform Resource Identifiers (URIs)
  - A URI is "A compact sequence of characters that identifies an abstract or physical resource". RFC 3986
- A URI provides a simple and extensible means for identifying resources.
- Each Web resource must have at least on URI.
  - The URI is the name and address of a resource.
- General **syntax** of a URI:

  scheme://<*host*><:*port*></*path*>?<*query*>#<*fragment*>

HTTP Messages

- HTTP messages are simple, line-oriented sequence of characters.
  - Formulated in plain text (ASCII): easy for humans to read and write.
  - Defined in RFC 7230

- HTTP defines **two types** of messages:

1. **Request messages:** sent from a Web client to a Web server.
2. **Response Messages:** sent from a Web server to a Web client.

- The formats of HTTP request and response messages are very similar.

HTTP Messages: Structure

- HTTP messages consist of **three parts**:

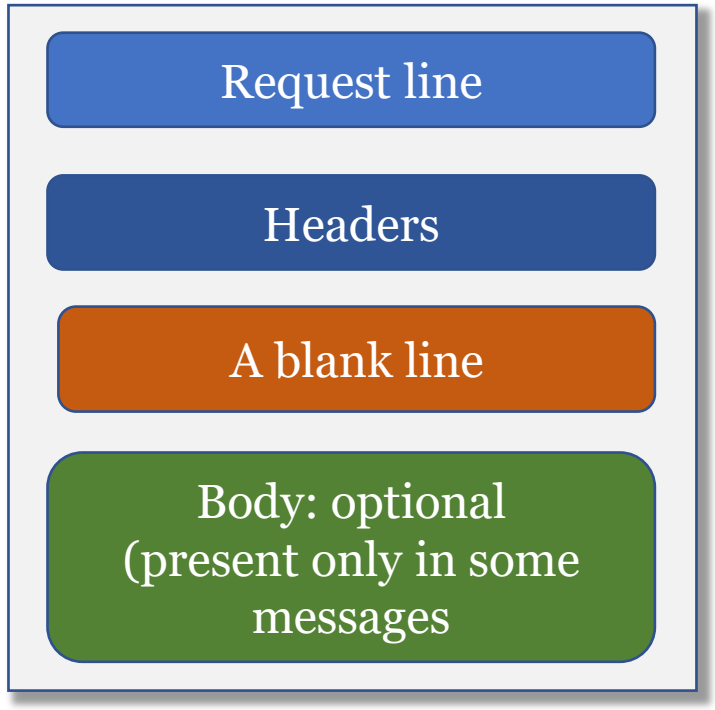
1. **Start line:** the first line of the message.
   - It indicates what to do for a request or what happened for a response.

2. **Header fields:** zero or more header fields follow the start line.
   - Each header field consists of **name/value pair** separated by a colon (:)
   - The headers fields are separated by a blank line
   - Adding a header field is as easy as adding another line.

3. **Body:** after the last blank line is an optional message body containing any kind of data.
   - **Request** bodies carry data to the Web server.
   - **Response** bodies carry data back to the client.

HTTP Messages: Structure (cont'd)

Structure of a Request message:



VS

Structure of a Response message: