Quais cuidados devem ser observados ao capturar dados de um site?

1. Não fazer scrapers antes de conferir se a página possui direitos autorais.
2. Não capturar dados sensíveis, mesmo que sejam públicos.
3. Abrir o arquivo robots.txt para saber quais caminhos tem ou não acesso permitido.