

再現と共有を可能とするデータ可視化方法について

今井 弘二^{*1}, 村山 泰啓^{*1}, 海老沢 研^{*2}, 池田 大輔^{*3}, 瀬口 大介^{*3}

Reproducible and shareable data visualization method

Koji Imai^{*1}, Yasuhiro Murayama^{*1}, Ken Ebisawa^{*2}, Daisuke Ikeda^{*3}, Daisuke Seguchi^{*3}

Abstract

Data management technology is becoming more and more important to promote scientific development in the society brimming with data. We improved the data visualization web service of earth, planetary and space sciences (Cross-Cutting Comparisons; C3) as the system controlled by the human-understandable query string (QS) to make reproducible and shareable charts. By including information of data handling procedures in the QS in an orderly manner, the chart is easy to understand, remake and share via text-based communication tools.

Keywords: reproducibility, data visualization, geoscience, cross-cutting research

概要

データが希少であった時代から溢れる時代となり、データマネジメント技術がますます重要視されるようになってきている。我々は、地球惑星科学におけるウェブサービス (Cross-Cutting Comparisons ; C3) を人間可読なクエリ文字列でコントロールするシステムとすることで、図やグラフなどのチャートを再現・共有できる新たな方法を実現した。クエリ文字列には、データの選択や処理手順などが明記されているため、チャートの理解が容易であり、また、同じクエリ文字列をテキストベースのコミュニケーションツールを利用すれば、同じチャートを再現することも可能である。

キーワード データマネジメント, データの再現性, 可視化, 地球惑星科学, C3

doi: 10.20637/JAXA-RR-17-009/0001

^{*} 平成 30 年 1 月 18 日受付 (Received January 18, 2018)

^{*1} 情報通信研究機構

(National Institute of Information and Communications Technology)

^{*2} 宇宙航空研究開発機構 宇宙科学研究所

(Japan Aerospace Exploration Agency, Institute of Space and Astronautical Science)

^{*3} 九州大学大学院システム情報科学 研究院・学府

(Graduate School of Information Science and Electrical Engineering, Kyushu University)

1 はじめに

科学技術の進歩によって、観測機器やセンサーネットワーク、計算機が小型化・高精度化・低コスト化され、膨大な量のデータが収集できるようになり、これにもとづく新しい研究や科学の発展が求められている。そして、それらのデータ資源の公開や共有などの取り組みとして、「オープンデータ」、あるいは「オープンサイエンス」が叫ばれるようになってきた¹⁾。しかし、それぞれの科学技術分野におけるデータを、国境や分野間の隔たりがなく、シームレスに活用するための社会を実現するためには、技術的、政策的、文化的、人材育成面といった乗り越えるべき壁が幾重にもあるため、研究機関や学協会などによるボトムアップな努力に加え、国際組織（Research Data Alliance, ICSU-WDS, OECD, G8/G7等）がトップダウンの政策を視野に入れて中長期的な課題解決を目指して取り組んでいる^{2, 3)}。

これらの国際組織による議論では、一拠点にデータを保有・管理する集中型から、各国の各機関がデータを保有・管理しながら相互に利用する、分散協調型のグローバルなデータマネジメントの体制を確立することが重要になると考えられている。科学技術分野では、これまでに個別のデータごとに対して、処理や可視化をするシステムが開発されてきた傾向があるが、分散協調型のグローバルなデータマネジメントの体制においては、分散配置された異分野のデータを統一的に取り扱うシステムがこれまで以上に求められると考えられる。また、近年では、第三者が再現できない論文が指摘されており、科学技術分野の発展の大きな妨げになっている^{4, 5)}。したがって、次世代のデータマネジメントシステムにはデータやチャート（ここでは、図やグラフなどの情報の視覚表現をまとめてチャートとよぶ）が、どのように処理され、どのように可視化されたかなどの作業過程を確認できる仕組みが必要である。

我々は遠隔のサーバからデータを取り込み、ブラウザ上にチャートを生成するシステムを開発した。チャートにはデータの選択や処理手順などを明示的に指定するクエリ文字列が付与されており、チャートの管理・共有が従来と比べて容易になると期待される。本稿では、第2章にこの新しい仕組みを実現した地球惑星科学のデータ可視化サービス C3（Cross-Cutting Comparisons）^{6, 7, 8)}の通信処理の流れや活用例について、第3章に本研究のまとめを記述する。

2 C3 について

C3 は太陽から人類生存圏までの空間における様々な現象が、どのように影響を及ぼしているのかを総合的に理解するために、地球惑星科学における分野横断型の複合研究を支援するウェブサービスである。多種多様なデータ形式を取り扱っており、データマネジメントシステムを実践・評価するテストベッドとして最適である。この章では、データやチャートの再現や共有を可能としたシステムの通信処理の流れや、活用例を中心に記述する。

2.1 通信処理の流れ

C3 のシステムは、ウェブサーバとデータサーバで構成している (図 1)。ユーザは、C3 のウェブページにアクセスし、インタラクティブな入力フォームに従って必要な情報を入力/選択することで、ブラウザ上にチャートを作成することができる。

この時、ユーザのリクエストはクライアント側で確認され (図 1 の (1))、適切なクエリ文字列が動的に生成される (クエリ文字列の文法については、2.2 節を参照)。そして、生成されたクエリ文字列はウェブサーバに送信され (図 1 の (2))、CGI プログラムによってユーザのリクエストに対するデータ処理・解析が実行される (図 1 の (3))。

その後、ウェブサーバからのレスポンスとして、データとチャートを作成するためのウェブコンポーネント (例えば、HTML や JavaScript のプログラム、CSS スタイルシートなど) がクライアントに送信され (図 1 の (4))、ブラウザ上にチャートが作成される (図 1 の (5))；C3 が提供するチャートの例については、付録の図 A を参照)。このように、ウェブサーバはクライアントとデータサーバの仲介役として位置づけられており、ユーザの反応に応じて、その都度、ウェブサーバがレスポンスを返すシステムではなく、クライアントとウェブサーバ間の通信を、1つのリクエストに対して1回のレスポンスで対応することによって、ウェブサーバの負荷を低減させるとともに、ネットワークの通信速度の遅延も避けている。

データサーバは、内部用と外部用に分けている。第1章に記述したように、各機関の膨大なデータを一つのデータサーバで一括に管理することは非現実的であるため、比較的小容量の小さいデータセットは内部のデータサーバで保有し、容量の大きなデータセットは外部のデータサーバから、ユーザのリクエストに応じて取得している。

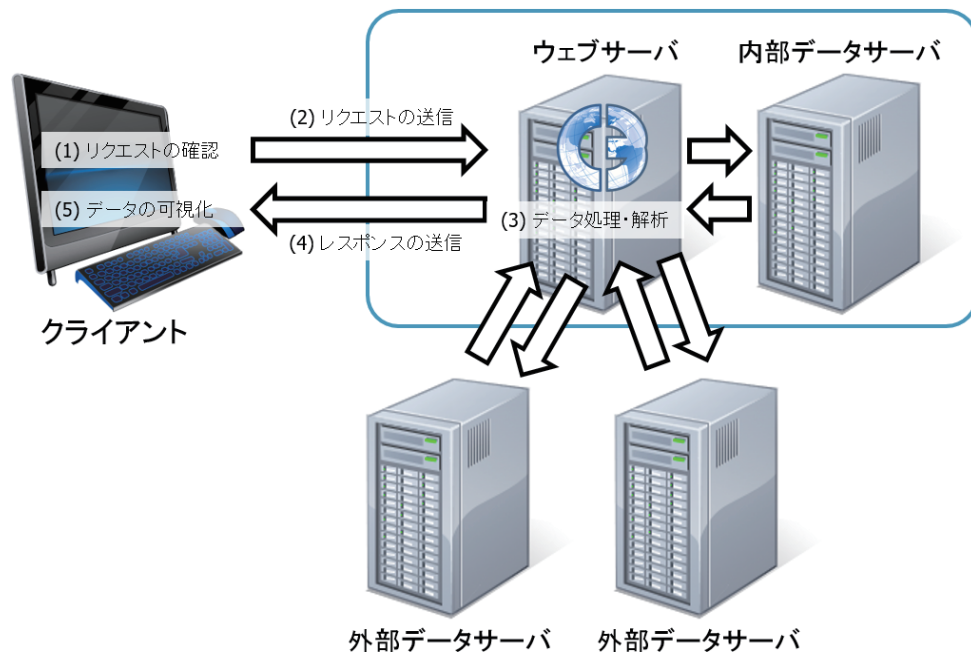


図 1 通信処理の流れ

図中の通し番号は、ユーザがチャートを作成するまでの一連の順序である。

2.2 クエリ文字列の取り扱い

地球環境における長期変動や突発的な現象の特徴を捉えるためには、データを時空間的に柔軟に表現して調査する必要がある。C3 は、HTML と JavaScript のライブラリを利用し、地球惑星科学における多種多様なデータ形式をブラウザ上に柔軟に表現している（図 A）。

ハイパーテキストシステムである WWW（World Wide Web）では、クエリ文字列は URL（Uniform Resource Locator）の一部であり、通常はクエスチョンマーク「?」に続けて記述される。そして、C3 の場合は、チャートを作成するための情報をクエリ文字列として次のように定義している。

```
C3_XXXX.cgi?en+SORCE_TSI_17_timeSeries+20100101_20100201  
C3_(システムのバージョン).cgi?ヘッダ部+データ選択部+データ抽出部+オプション部
```

上の例は、SORCE（Solar Radiation and Climate Experiment）⁸⁾が捉えた全太陽放射照度（Total Solar Irradiance ; TSI）のバージョンが 17 のデータを時系列図で、2010 年 1 月 1 日から 2010 年 2 月 1 日まで表示するためのクエリ文字列である（チャートは図 A (a) を参照）。クエリ文字列は、プラス「+」で分けられた 4 つの部分で構成している。先頭はヘッダ部であり、その後にデータ選択部、データ抽出部、そして必要であればオプション部が後に続く。

ヘッダ部には、チャートの表示言語などの情報が、データ選択部には、データのメタ情報と表示するチャートの種類が記載されている。そして、データ抽出部に選択したデータの期間や緯度、経度、高度などの抽出条件が記載されている。最後のオプション部には、チャートのタイトルやサブタイトル、データの表示形式の切替情報などが付加されている。このように、クエリ文字列にデータの選択や抽出条件などの情報を入れて、人間可読な形式とすることで、作成されるチャートの理解が容易となり、ユーザによる編集も可能となる。

C3 では、システムのバージョンによって、データの処理と可視化を、データのバージョンによって、データサーバにあるデータを制御しているため、同じクエリ文字列を含む URL（以下、URL クエリ文字列と呼ぶ）を用いれば、同じチャートが再現される。ただし、外部にあるデータは管理していないため、同じチャートを保証するわけではない。

2.3 活用例

C3 が提供する URL クエリ文字列は、選択するだけでブラウザ上にチャートを再現することができるため、URL クエリ文字列を管理しておけば、画像ファイルを保有する必要がなく、チャートを多用する研究でも、ユーザが管理するデータサイズを抑えることができる。さらに、テキストベースのコミュニケーションツール（例えば、E メールやツイッターなど）を使って同じ URL クエリ文字列を第三者にも共有することができるため、ユーザ間の通信

量を低減する新たなチャートの共有方法としても期待される（図 2）。

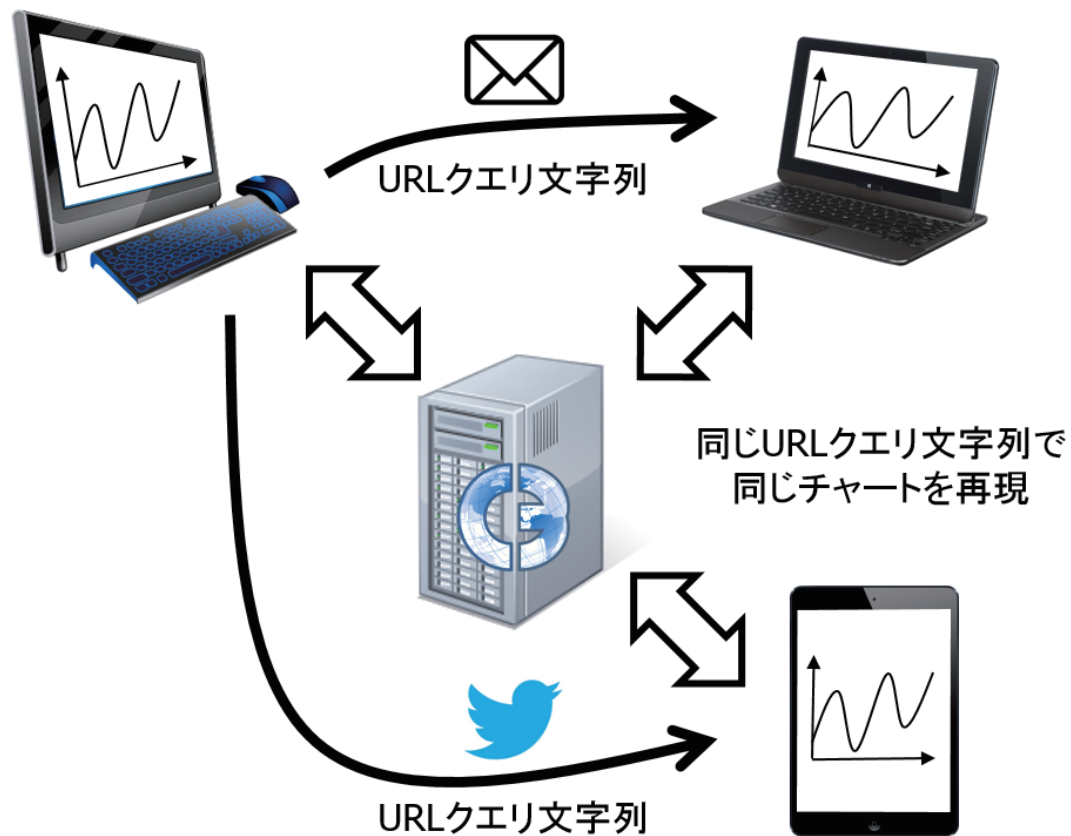


図 2 URL クエリ文字列を用いたチャートの共有

3 まとめ

データがあふれる現代となり、信頼のおけるデータを共有できるデータマネジメントを確立することが重要視されている。我々は、地球惑星科学における多種多様なデータを扱うウェブサービス C3 を、人間可読なクエリ文字列でコントロールするシステムに改良した。これによって、クエリ文字列からデータの選択や処理手順などが理解でき、ユーザによる編集も可能となった。また、テキストベースのコミュニケーションツールを利用してクエリ文字列を共有すれば、同じチャートが再現・共有できるだけでなく、ユーザ間の通信量も低減されるため、新たなチャートの共有方法として期待される。

謝辞

本研究は宇宙航空研究開発機構・宇宙科学研究所，科学衛星運用・データ利用センター（C-SODA），JSPS 科研費 15H02787 の助成を受けたものです。ここに感謝の意を表します。

参考文献

- 1) Huijboom, Noor, and Tijs Van den Broek. "Open data: an international comparison of strategies." *European journal of ePractice* 12.1 (2011): 4-16.
- 2) 渡邊堯. "ICSU 世界データシステム (WDS) について." *学術の動向* 17.6 (2012): 6_11-6_15.
- 3) ICSU World Data System, <https://www.icsu-wds.org> (参照日 2017.11.20).
- 4) Baker, Monya. "1,500 scientists lift the lid on reproducibility." *Nature* 533.7604 (2016): 452-454.
- 5) Iqbal, Shareen A., et al. "Reproducible research practices and transparency across the biomedical literature." *PLoS Biol* 14.1 (2016): e1002333.
- 6) Imai, Koji, et al. "Quick look service for geoscience." *Journal of Space Science Informatics Japan*, Vol. 5 (2016): 93-109.
- 7) 今井弘二, 村山泰啓, 松下幸市朗. "地球科学における対話形式のデジタル教材の開発とその評価の試み.", *教育情報研究*, 33 巻 1 号.
- 8) Cross-Cutting Comparisons, <https://darts.isas.jaxa.jp/C3/> (参照日 2017.11.20).
- 9) Rottman, Gary. "The SORCE mission." *Solar Physics* 230.1-2 (2005): 7-25.
- 10) Bernath, Peter F., et al. "Atmospheric chemistry experiment (ACE): mission overview." *Geophysical Research Letters* 32.15 (2005).
- 11) Kikuchi, Ken-ichi, et al. "Overview and early results of the Superconducting Submillimeter - Wave Limb - Emission Sounder (SMILES)." *Journal of Geophysical Research: Atmospheres* 115.D23 (2010).
- 12) Imai, Koji, et al. "Validation of ozone data from the Superconducting Submillimeter - Wave Limb - Emission Sounder (SMILES)." *Journal of Geophysical Research: Atmospheres* 118.11 (2013): 5750-5769.
- 13) U.S. Geological Survey, <https://www.usgs.gov/> (参照日 2017.11.20).

付録 A

図 A は C3 が提供するチャートの例である． 図 A (a) は SORCE⁸⁾ の全太陽放射照度の時系列図であり，データ点にマウスオーバーして（マウスカーソルを重ねて），ポップアップで詳細な情報を表示している様子を示している． 図 A (b) は ACE-FTS（Atmospheric Chemistry Experiment-Fourier Transform Spectrometer）⁹⁾ の温度の高度分布図であり，ドラッグして拡大表示している様子を示している． 図 A (c) は SMILES（Superconducting Submillimeter-Wave Limb-Emission Sounder）^{10, 11)} のオゾンの全球分布図である．そして，図 A (d) は USGS（United States Geological Survey）¹²⁾ の震源図である．データ点にマウスオーバーして詳細な情報を表示している．チャート自体をドラッグして回転させることも可能である．このように，C3 は点，線，面，そして疑似的な 3 次元まで表現し，自在なスケール調整やデータの詳細表示など，多機能なチャートを提供している．

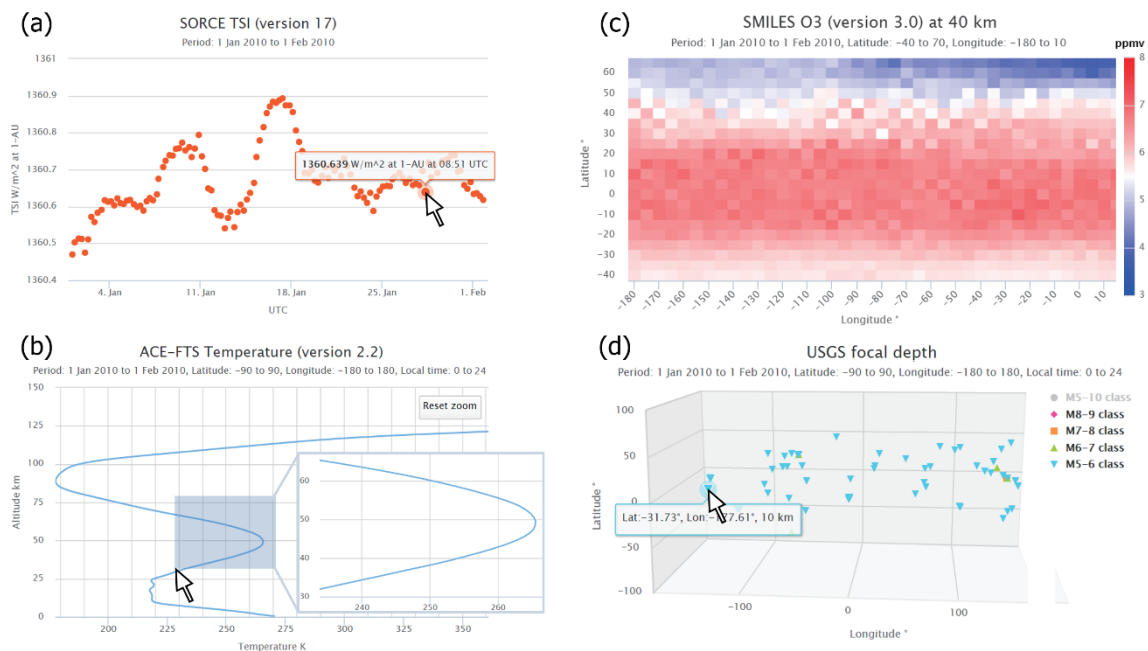


図 A 2 C3 が提供するチャートの例

- (a) 全太陽放射照度の時系列図，(b) 温度の高度分布，
(c) オゾンの緯度経度断面図，(d) 震源図