



# CHARMe: Sharing climate knowledge through commentary metadata and Linked Data (Jan 2013 – Dec 2014)

**Victoria Bennett**

(Slides thanks to Jon Blower)  
on behalf of CHARMe consortium  
<http://www.charme.org.uk>



Science & Technology  
Facilities Council



Royal Netherlands Meteorological Institute  
Ministry of Infrastructure and the Environment

**CGI**



Deutscher Wetterdienst  
Wetter und Klima aus einer Hand



National Centre for  
Earth Observation  
NATURAL ENVIRONMENT RESEARCH COUNCIL



National Centre for  
Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# The Satellite Application Facility on Climate Monitoring

- Overview
- Products
- Data Access
- Documentation
- Training

 >>

- Available Products
- Product Monitoring
- Ordering and Data Access
- Service Message
- Change Log
- Known Product Disruption
- Naming Convention
- Auxiliary Data
- Golden Dataset
- Add on Products
- Tools
- Digital Object Identifier (DOI)**

- Service
- Newsletter
- Contact, User Help Desk&UPR
- Staff
- FAQ's
- Glossary
- Links
- Disclaimer&Acknowledgement

Homepage > Products > Digital Object Identifier (DOI)

Title	DOI	CHARMe <sup>i</sup>
<input type="checkbox"/> <b>C</b> (59)		
Select/unselect all (0 of 12 targets selected)		
SEVIRI cloud mask dataset-Edition 1 15 minutes resolution	<a href="#">10.5676/EUM_SAF_CM/CMA_SEVIRI/V001</a>	<input type="checkbox"/> <b>C</b> (3)
Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data HOAPS 3.2 Monthly Means/6-Hourly Composites	<a href="#">10.5676/EUM_SAF_CM/HOAPS/V001</a>	<input type="checkbox"/> <b>C</b> (3)
Vertically Integrated Water Vapour from SSM/I Daily/Monthly Means	<a href="#">10.5676/EUM_SAF_CM/HTW_SSMI/V001</a>	<input type="checkbox"/> <b>C</b> (3)
CM SAF Surface Radiation MVIRI Data Set 1.0 Monthly Means/Daily Means/ Hourly Means	<a href="#">10.5676/EUM_SAF_CM/RAD_MVIRI/V001</a>	<input type="checkbox"/> <b>C</b> (4)
CLARA-A1: CM SAF Clouds, Albedo and Radiation dataset from AVHRR data Edition 1 Monthly Means/Daily Means/	<a href="#">10.5676/EUM_SAF_CM/CLARA_AVHRR/V001</a>	<input type="checkbox"/> <b>C</b> (14)



## Annotations



Viewing annotations for:



Title	Annotated by	Organisation	Date
<a href="#">Validation Report</a>	<a href="#">Nathalie Selbach</a>	<a href="#">CMSAF-DWD</a>	21 Apr 2015
<a href="#">Product User Manual (PUM)</a>	<a href="#">Nathalie Selbach</a>	<a href="#">CMSAF-DWD</a>	21 Apr 2015
<a href="#">Algorithm Theoretical Basis Document (ATBD)</a>	<a href="#">Nathalie Selbach</a>	<a href="#">CMSAF-DWD</a>	21 Apr 2015

v1.2

Login

Cancel

- Service
- [Newsletter](#)
- [Contact, User Help Desk&UPR](#)
- [Staff](#)
- [FAQ's](#)
- [Glossary](#)
- [Links](#)
- [Disclaimer&Acknowledgement](#)

CM SAF Clouds Radiation from  
Data Set 1.0  
Monthly Means/Daily Means/  
Hourly Means  (4)

CLARA-A1:  
CM SAF Clouds, Albedo and  
Radiation dataset from AVHRR  
data Edition 1  (14)

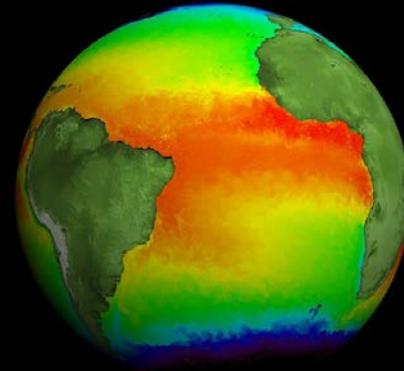
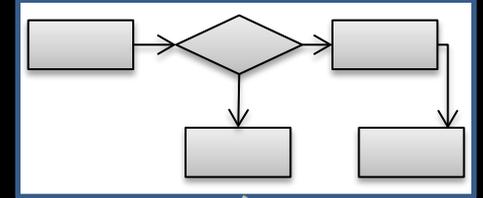


platform

instrument



algorithm



dataset

**THE ESA CLIMATE CHANGE INITIATIVE**  
 Satellite Data Records for Essential Climate Variables

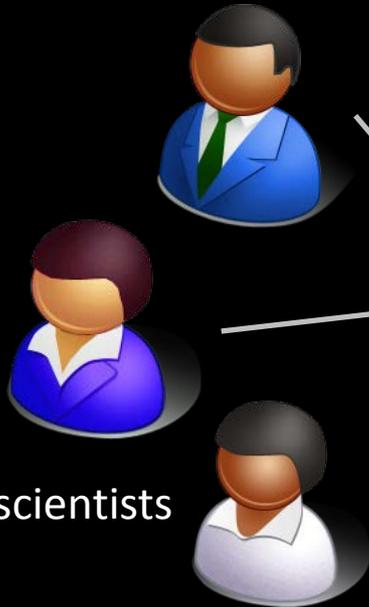
BY R. HOLLMANN, C. J. MERCHANT, R. SAUNDERS, C. DOWNY, M. BUCHWITZ, A. CAZENAVE,  
 E. CHIRIACCO, P. DEFOURNY, G. DE LEEUW, R. FORSBERG, T. HOLZER-POPP, F. PAUL, S. SANDEWEN,  
 S. SATHYENDRANATH, M. VAN ROOZENEGEL, AND W. WAGNER

The ESA's Climate Change Initiative is reprocessing and reassessing over 40 years of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.

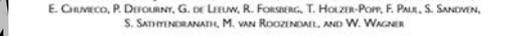
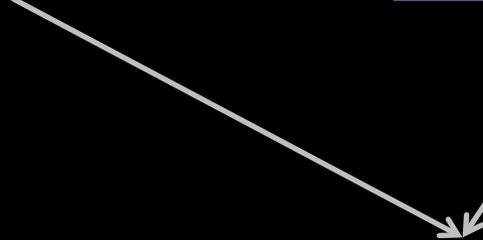
Sustained observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—one of the most pressing needs of climate science, destination key variables

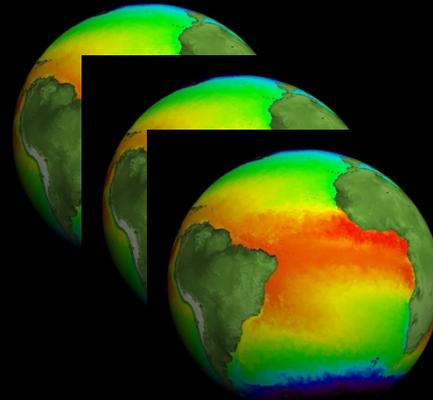
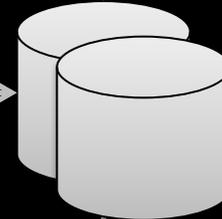
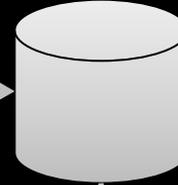
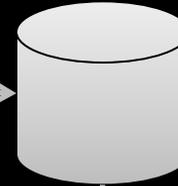
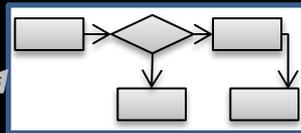
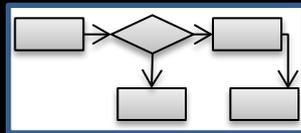
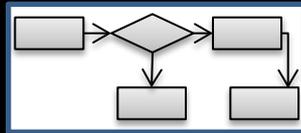
The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, destination key variables

publication



scientists



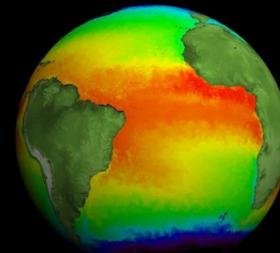
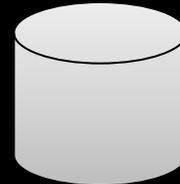
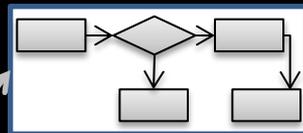


**THE ESA CLIMATE CHANGE INITIATIVE**  
 Satellite Data Records for Essential Climate Variables

Dr R. HOLMANN, C. J. MICHANG, R. SANDERS, C. DONNO, M. BUCHWITZ, A. CAZZANI, E. CHIRICO, P. DIROGARI, G. DE LEO, R. FORABONO, T. HÖRNER, F. PIVA, S. SARDINI, S. SATTENBERGER, M. VAN ROSSDORP, AND W. VIGNER

The ESA's Climate Change Initiative is reprocessing and reanalysing over 40 years of multi sensor satellite records to generate consistent, traceable, long term datasets of "essential climate variables" for the climate modeling and research communities.

Continued observations from satellites contribute vital knowledge to our understanding of Earth's climate system and how it has changed over the past decades. The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science. Cooperation between

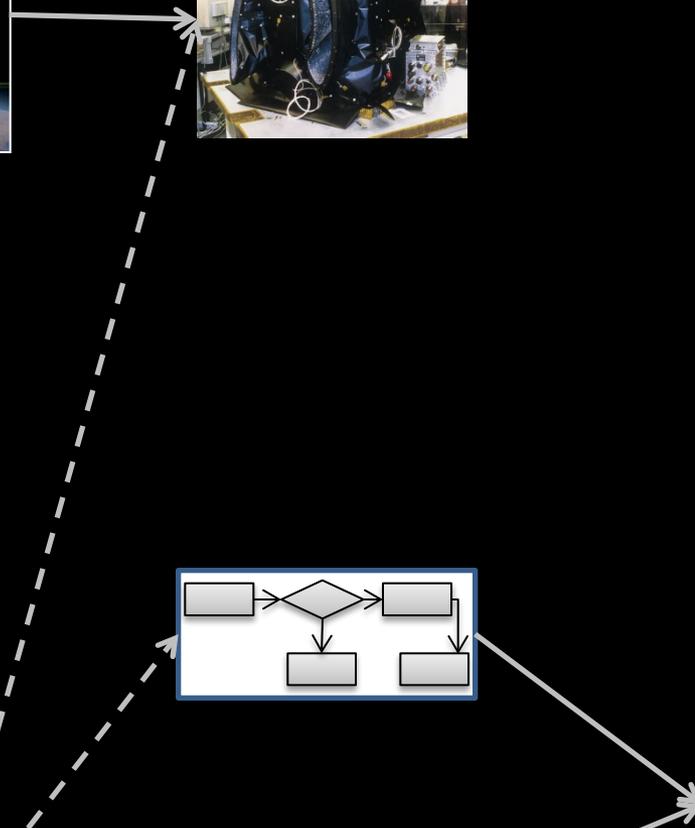


**THE ESA CLIMATE CHANGE INITIATIVE**  
 Satellite Data Records for Essential Climate Variables

Dr R. HOLMANN, C. J. MICHANG, R. SANDERS, C. DONNO, M. BUCHWITZ, A. CAZZANI, E. CHIRICO, P. DIERCKMANN, G. DE LEEUW, R. FORBES, T. HÖLZMANN, F. PIVA, S. SARDEN, S. SATHIYANARAYAN, M. VAN ROOYENDIJK, AND W. VOGEL

The ESA's Climate Change Initiative is reprocessing and reanalysing over 40 years of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.

Continued observations from satellites contribute vital knowledge to our understanding of Earth's environment and its changing state. The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science. Cooperation between

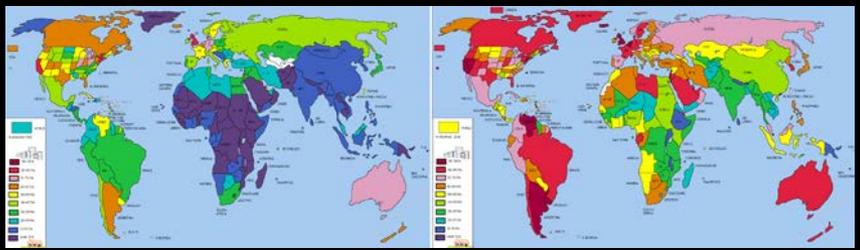
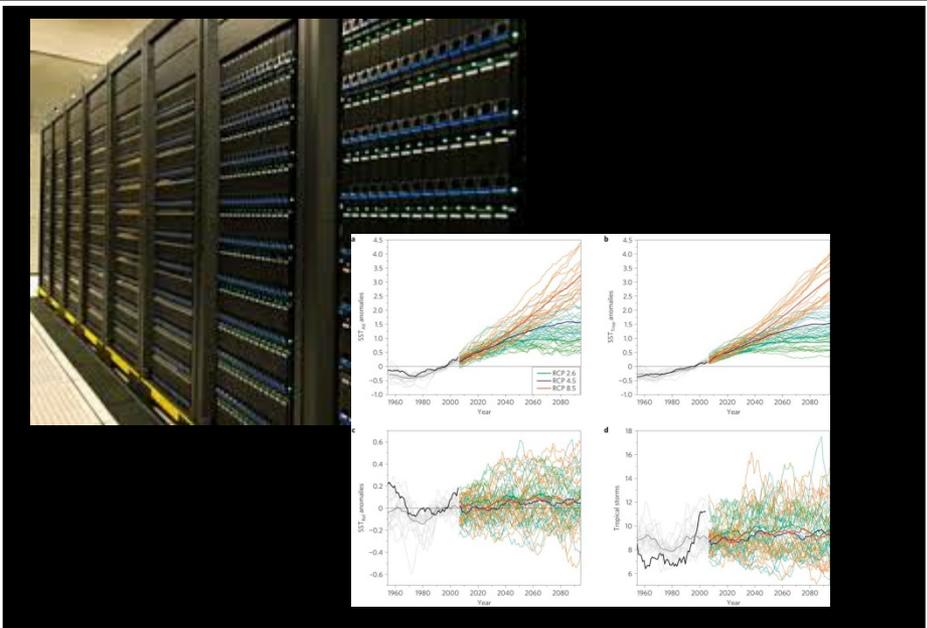
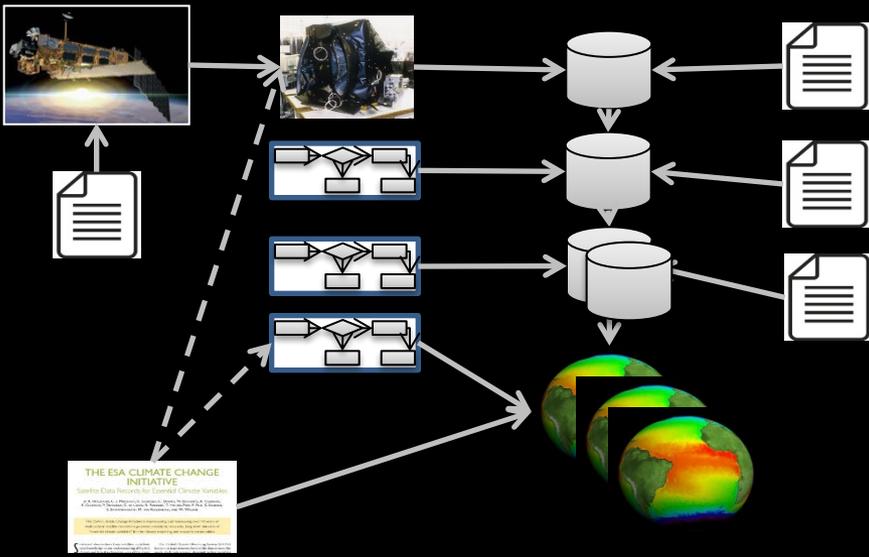


# How do you fill in the blanks?

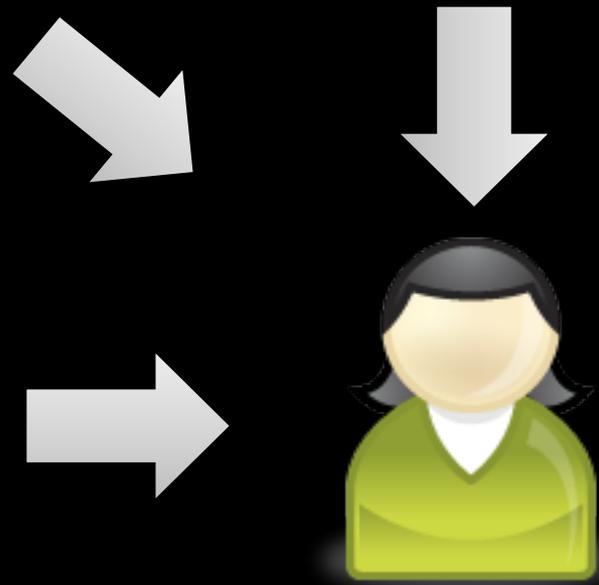
- Hope the documentation contains references
- ... to the right versions of things
- ... and using consistent terminology
  
- In restricted communities, we can gather all information into a central location and harmonize it
  
- OR we hope that a Google search throws up the right results

# Problems

1. Crossing communities is very hard



[www.fooddeserts.org](http://www.fooddeserts.org)



# Problems

1. Crossing communities is very hard
2. Lots of useful stuff isn't formally documented

# Problems

1. Crossing communities is very hard
2. Lots of useful stuff isn't formally documented
3. Unknown unknowns...

# Consequences

- Scientists often use the data that are most easily available, not necessarily the best
- Constant re-discovery of the same issues
- Very hard to share information outside communities

# The “Web of Data”

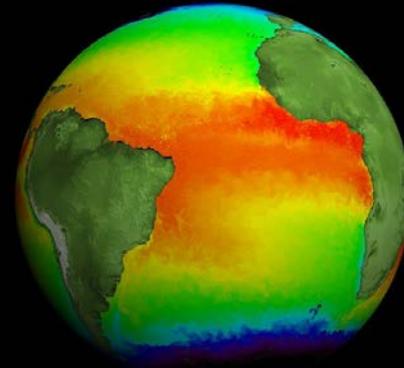
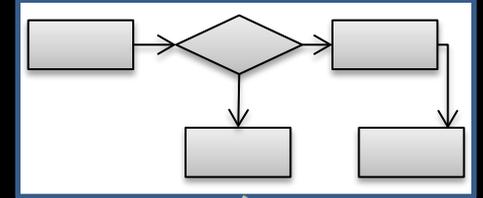


platform

instrument



algorithm



dataset

**THE ESA CLIMATE CHANGE INITIATIVE**  
 Satellite Data Records for Essential Climate Variables

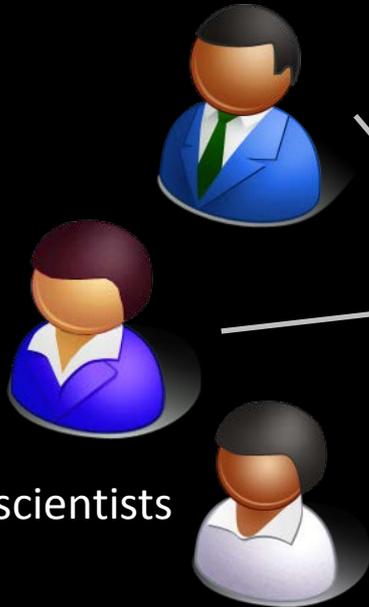
BY R. HOLLMANN, C. J. MERCHANT, R. SAUNDERS, C. DOWNY, M. BUCHWITZ, A. CAZENAVE,  
 E. CHIRIACCO, P. DEFOURNY, G. DE LEEUW, R. FORSBERG, T. HOLZER-POPP, F. PAUL, S. SANDEWY,  
 S. SATHYENDRANATH, M. VAN ROOZENEGEL, AND W. WAGNER

The ESA's Climate Change Initiative is reprocessing and reassessing over 40 years of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.

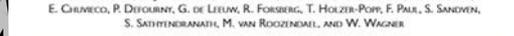
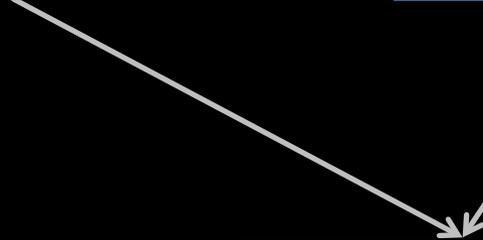
Sustained observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing—one of the most pressing needs of climate science, destination key variables

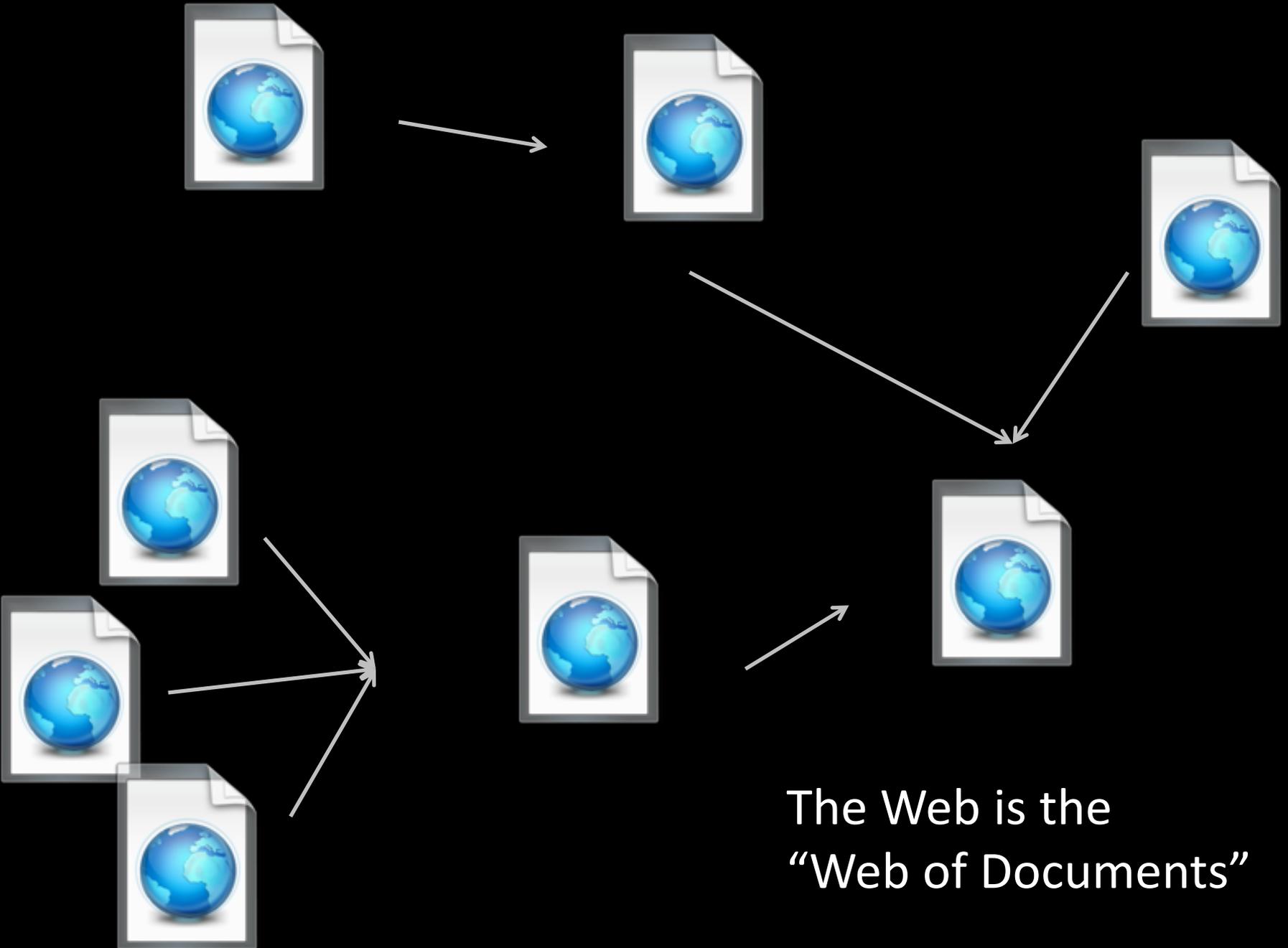
The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science, destination key variables

publication



scientists





The Web is the  
"Web of Documents"

# Fingers crossed!

- We hope that everything has a document written about it, or we have nothing to cite
- ... and we hope someone writes a new document when “it” changes
- We hope that we are citing **the right** document (i.e. the most authoritative etc)

“If ... the Web made all the online documents look like one huge book, **[Linked Data]** will make all the data in the world look like one huge database.”

Sir Tim Berners-Lee



Photo by Susan Lesch, from Wikipedia

# Why is Linked Data different?

- Focuses on **things**, not documents-about-things
- Requires things to have **unique and persistent identifiers**
  - E.g. DOIs, permanent URLs, etc.
- **Links things together** in a meaningful way
- Links are **readable by computers**

# Example: linking a paper to a dataset

(illustrative, not necessarily precise!)

subject predicate object

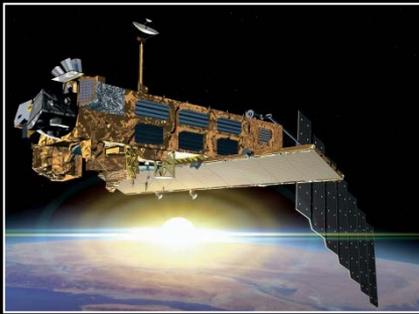
this is a “triple”

each of these terms has a **unique identifier**

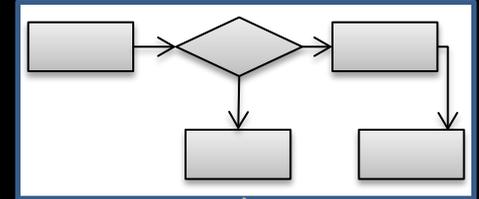
<http://dx.doi.org/12345.678910>

<http://purl.org/spar/cito/citesAsDataSource>

<http://www.badc.ac.uk/datasets/sst-cci>

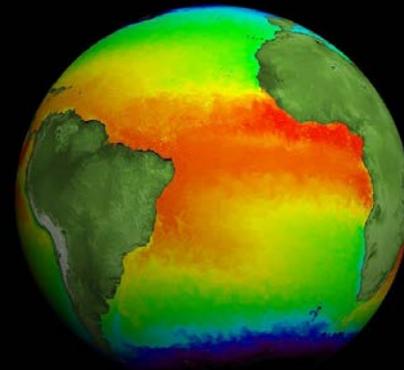


carries



produced

produced



authored

**THE ESA CLIMATE CHANGE INITIATIVE**  
 Satellite Data Records for Essential Climate Variables

by R. HEDDMANN, C. J. MARCHAND, R. SERRANO, C. DONNER, M. BUCHHEITZ, A. CAZZARI, E. CAUPELO, P. DEFOURNE, G. DE LEEUW, R. FOUMEL, T. HÖLZER-PÖRN, F. PAUL, S. SENECHIN, S. SEIFENBERGER, M. VAN ROOIJEN, and W. WINKER

The ESA's Climate Change Initiative is representing and maintaining over 95 year's of multi-sensor satellite records to generate consistent, traceable, long-term datasets of "essential climate variables" for the climate modeling and research communities.

Satellite observations from satellites contribute vital knowledge to our understanding of Earth's climate and how it is changing over the course of time. The Global Climate Observing System (GCOS) has set out requirements for satellite data to meet the needs of climate science. ESA's mission is to provide the world with the most accurate and reliable satellite data available.

describes / uses / queries ...

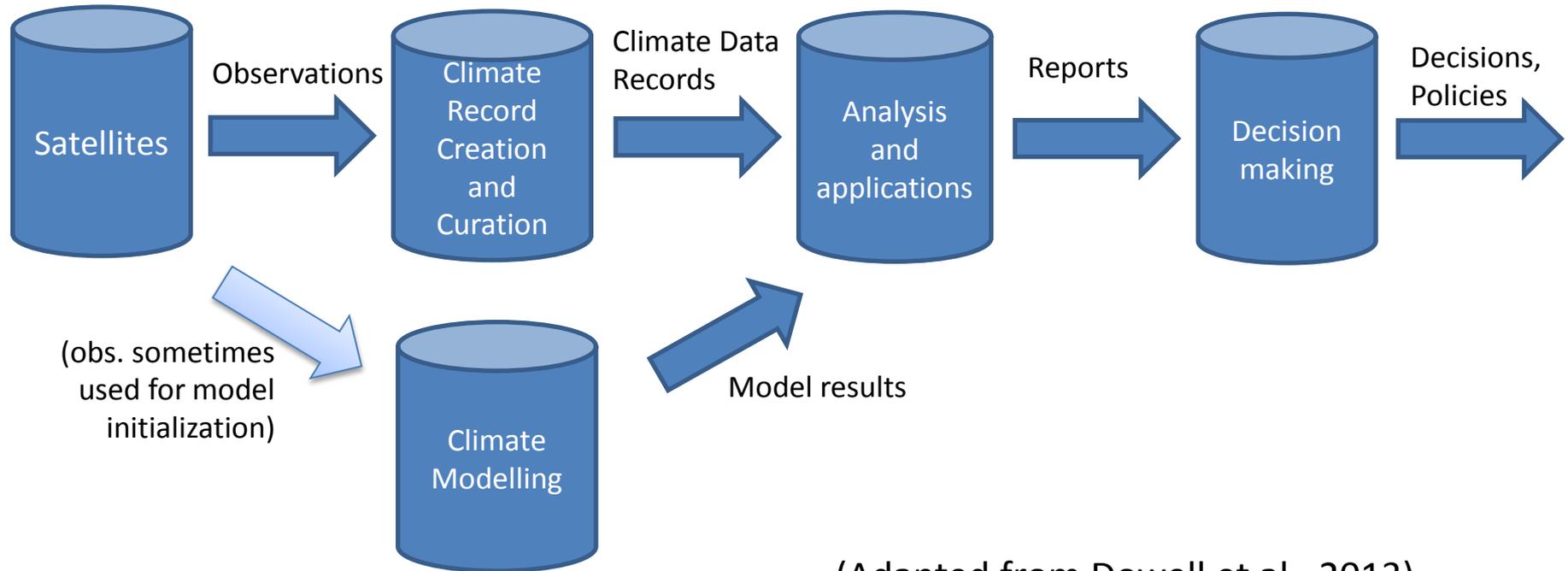
The "Web of Data"



# CHARMe: Sharing climate knowledge through commentary metadata and Linked Data (Jan 2013 – Dec 2014)



# From observations to decisions

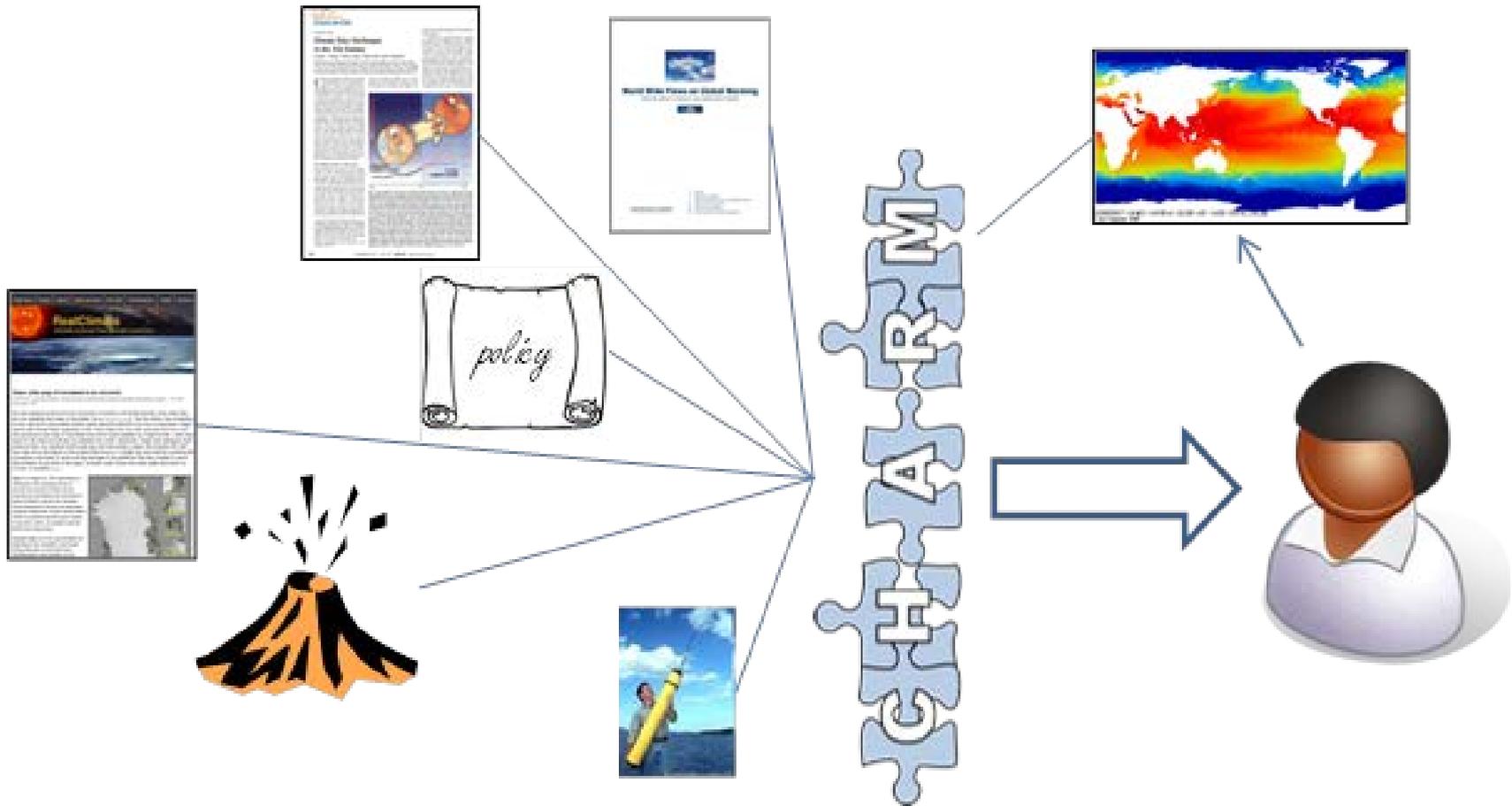


(Adapted from Dowell et al., 2013),  
*“Strategy Towards an Architecture for  
Climate Monitoring from Space”*)

# Where can users go for help?

- Scientific literature
  - Huge, verbose and inaccessible to some communities
  - Not well linked to source data
- Technical reports and conference proceedings
  - Hard to find, scattered or inaccessible
- Data centres
  - increasingly strong at providing some important metadata, but don't usually include community feedback
  - Not all countries and communities have data centres!
- Websites and blogs
  - From CEOS Handbook to a scientist's blog
  - Increasingly useful, but scattered

# “Commentary metadata”



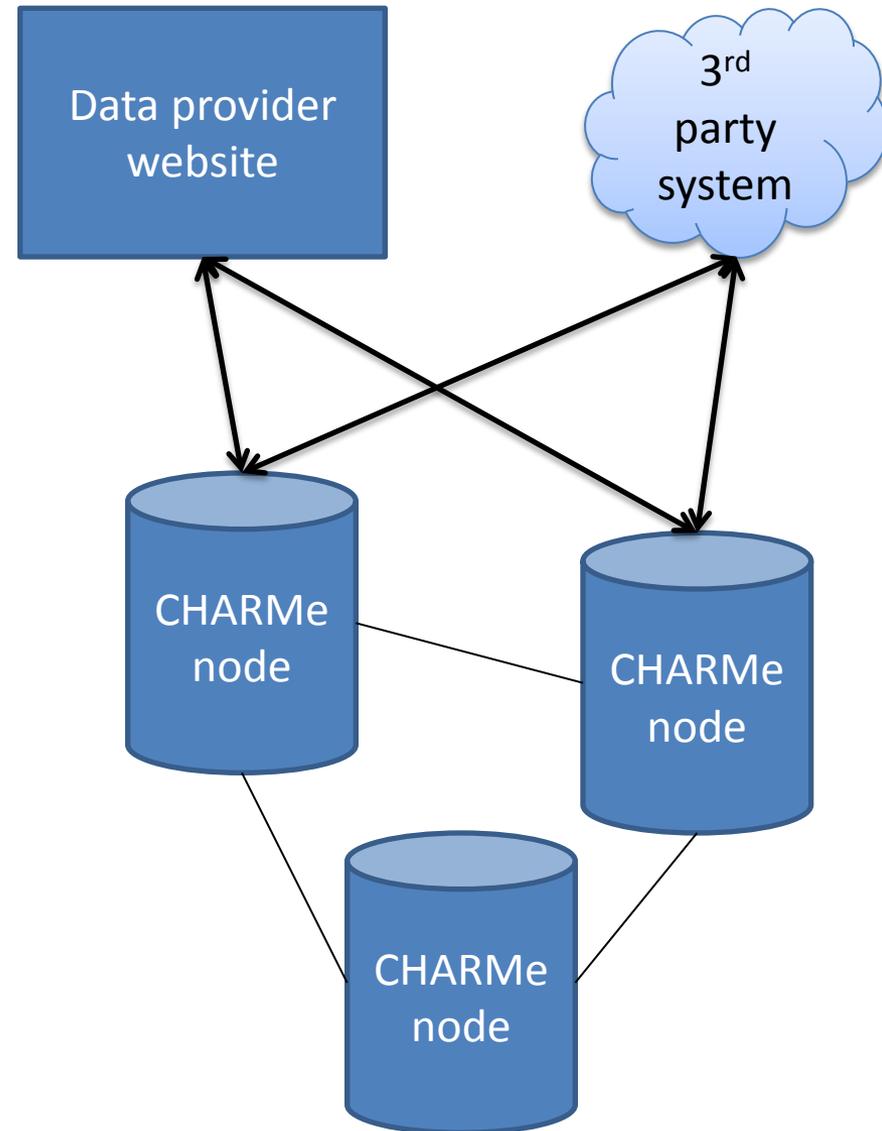
# Examples of commentary metadata

- **Post-fact annotations**, e.g. citations, ad-hoc comments and notes;
- **Results of assessments**, e.g. validation campaigns, intercomparisons with models or other observations, reanalysis;
- **Provenance**, e.g. dependencies on other datasets, processing algorithms and chain, data source;
- **Properties of data distribution**, e.g. data policy and licensing, timeliness (is the data delivered in real time?), reliability;
- **External events** that may affect the data, e.g. volcanic eruptions, El-Nino index, satellite or instrument failure, operational changes to the orbit calculations.

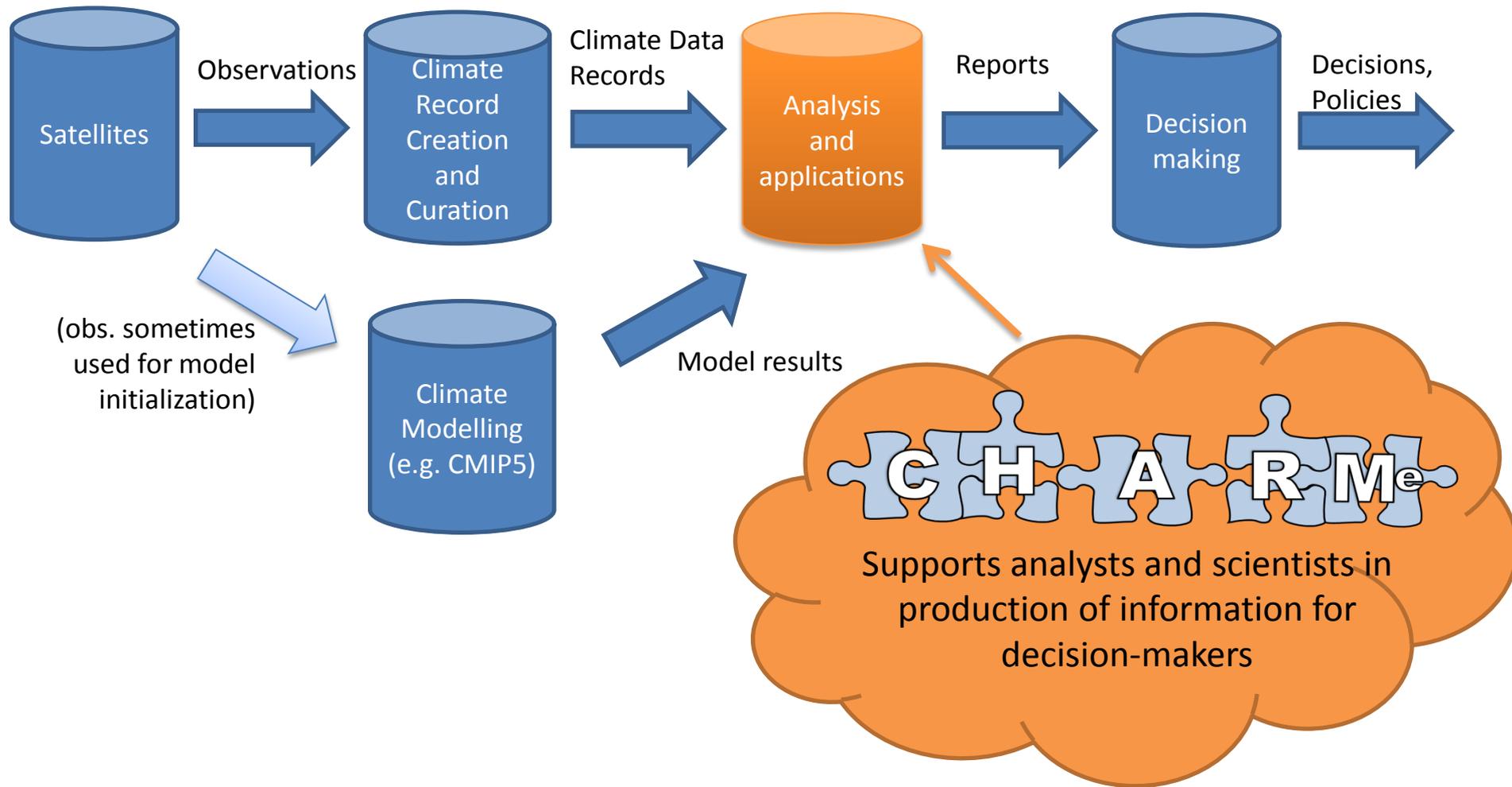
General rule: information originates from **users or external entities**, not original data providers

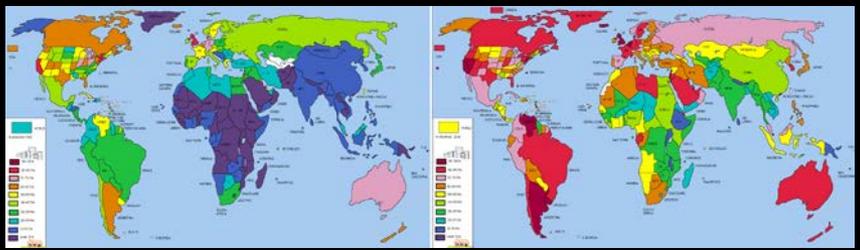
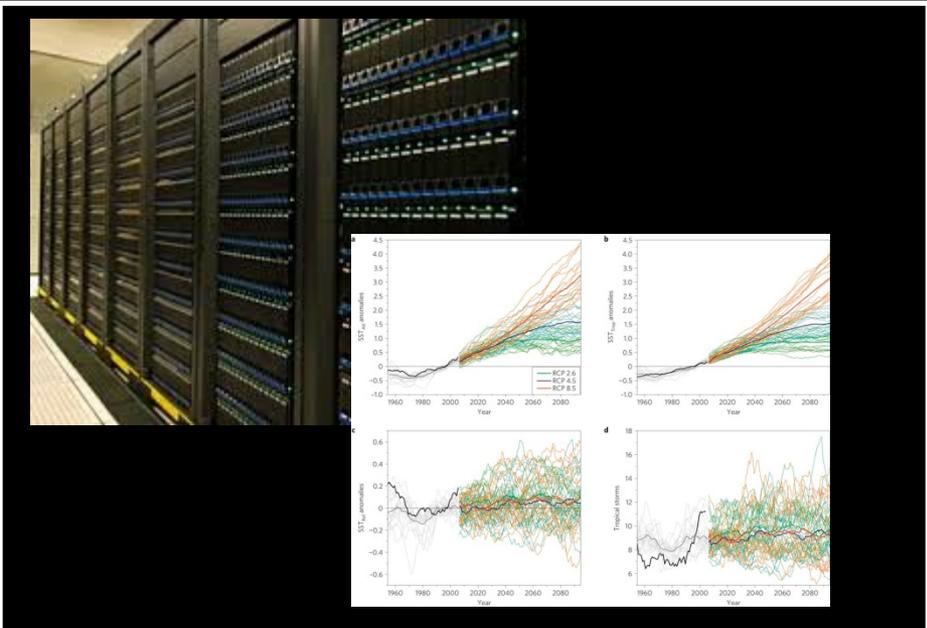
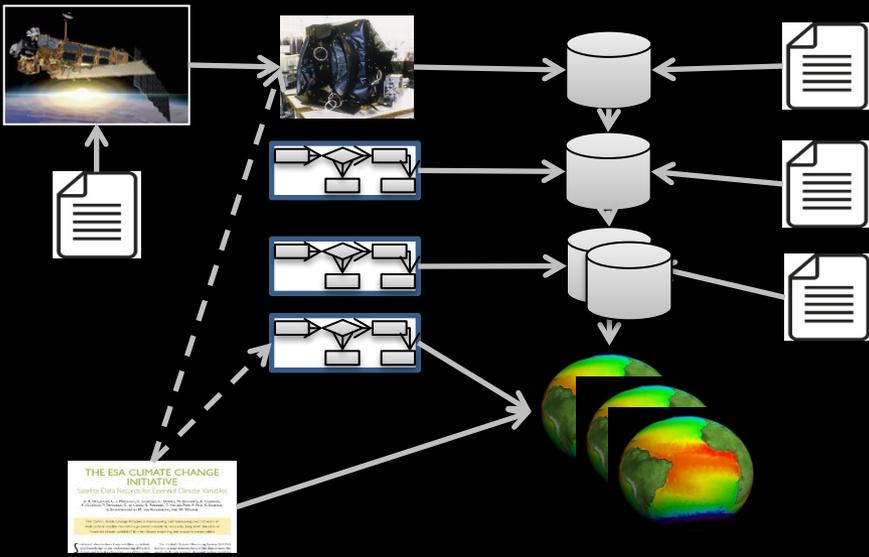
# How will this be done?

- CHARMe will create connected repositories of commentary information
  - Stored as triples in “CHARMe nodes”,
- Information can be read and entered through websites or Web Services
- Using principles of Open Linked Data

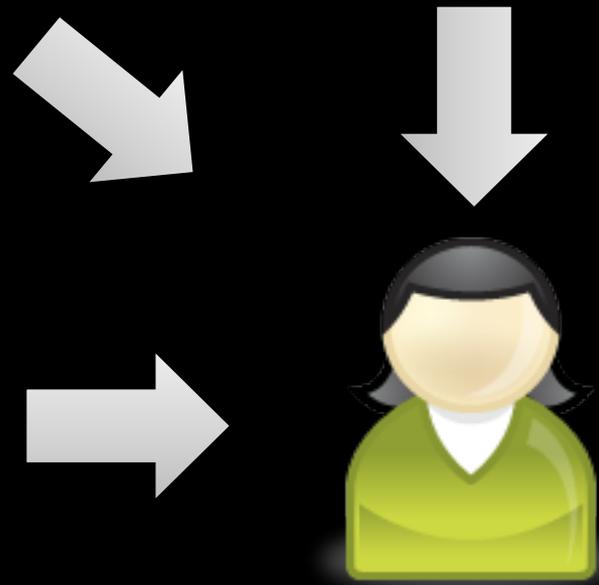


# Where does CHARMe fit in?





[www.fooddeserts.org](http://www.fooddeserts.org)



# Challenges

- Ensuring community adoption:
  - We are “injecting” CHARMe capabilities into **existing websites** used in the community
  - We will “seed” the CHARMe system with information to attract users (e.g. links between publications and datasets)
- Ensuring quality of commentary metadata
  - **Moderation** is a strong user requirement
  - We will provide guides to creators of commentary – what makes a comment helpful?
- Technical challenges:
  - Ensuring that datasets (and missions, instruments...) have identifiers
  - Deciding what a dataset is!

# What CHARMe will enable (some examples)

## **Users:**

- “Find me all the documents that have been written about this dataset”
  - “... in both peer-reviewed journals and the grey literature”
  - “... and specifically about precipitation in Africa”
  - “... in both CEDA’s and Astrium’s archives”
- “What factors might affect the quality of this dataset?”  
e.g. upstream datasets, external events
- “What have other users already discovered about this dataset?”
- “I want to find information related to the dataset I’m looking at”

## **Data providers:**

- “Who is using my dataset and what are they saying about it?”
- “Let me subscribe to new user comments and reply to them”

# What CHARMe will not enable

- “Give me the best dataset on sea surface temperature”
  - The “best” dataset depends on the application
- CHARMe will not provide a new “quality stamp” for datasets
  - But will be able to link to such things if other people publish them, e.g. CORE-CLIMAX Application Performance Matrix, QA4EO certification
- CHARMe will not provide **access** to actual data
  - (but will help with discovery of data)
- Not planning to create “one-stop shop” for information
  - We want the information to appear where users are already looking

# How will CHARMe be used?

The screenshot displays the CHARMe Annotations interface. On the left, a table lists data providers with the URL 'badc.nerc.ac.uk'. The main window, titled 'CHARMe Annotations', shows a specific annotation for the dataset '/view/badc.nerc.ac.uk\_\_ATOM\_\_dataent\_12330719779627095'. The annotation content includes:

- Text annotations:**
  - Ground truthing in the Arctic is cold! 22/11/2013 16:16pm | Andrew Henry
  - It is cold in the arctic when doing ground truthing 22/11/2013 16:16pm | Andrew Henry
- Publications:**
  - Unable to retrieve metadata [http://www.aslo.org/lo/toc/vol\\_45/issue\\_...](http://www.aslo.org/lo/toc/vol_45/issue_...) 22/11/2013 16:16pm | Maurizio Nagni
  - Bottenheim, J. W., Gallant, A. G., & Brice, K. A. (1986). Measurements of NO<sub>y</sub> species and O<sub>3</sub> at 82° N latitude. Geophysical Research Letters, 13(2), 113-116. Wiley Blackwell (John Wiley & Sons). doi:10.1029/GL013i002p00113

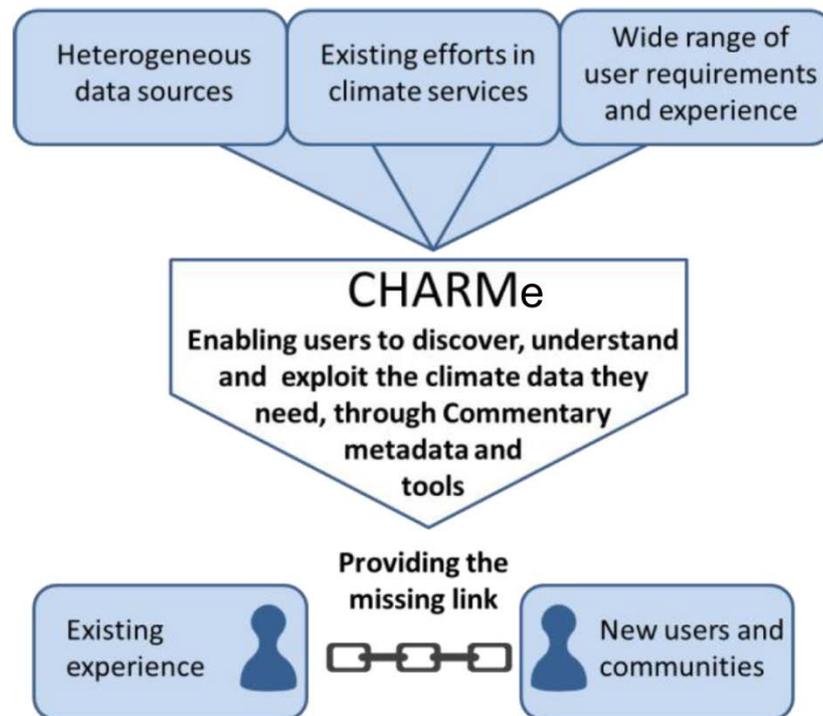
At the bottom of the main window are buttons for 'Sign In', 'New Annotation', and 'Done'. On the right side of the interface, a vertical sidebar contains a 'CHARMe Annotations' header and a series of puzzle-piece icons, each with a 'C' inside, representing the CHARMe button. A red box highlights this sidebar.

Stratosphere (ITLS)

- Search for datasets on existing data provider website
- Use the “CHARMe button” to view or record commentary on a particular dataset
- See [http://www.youtube.com/watch?v=F5k\\_FQt5eb0](http://www.youtube.com/watch?v=F5k_FQt5eb0) for a video of latest prototype

# Summary

- CHARMe is using Linked Data techniques to help users of climate data to connect with all the experience in the community
  - Techniques are not specific to climate data
- CHARMe is about the **links**: other sources provide (most of) the content
- So far we can record documents and free-text annotations about datasets, but can be used to make any kind of links
- Commentary information can be injected into existing websites



**Thank you!**

victoria.bennett@stfc.ac.uk

<http://www.charme.org.uk>