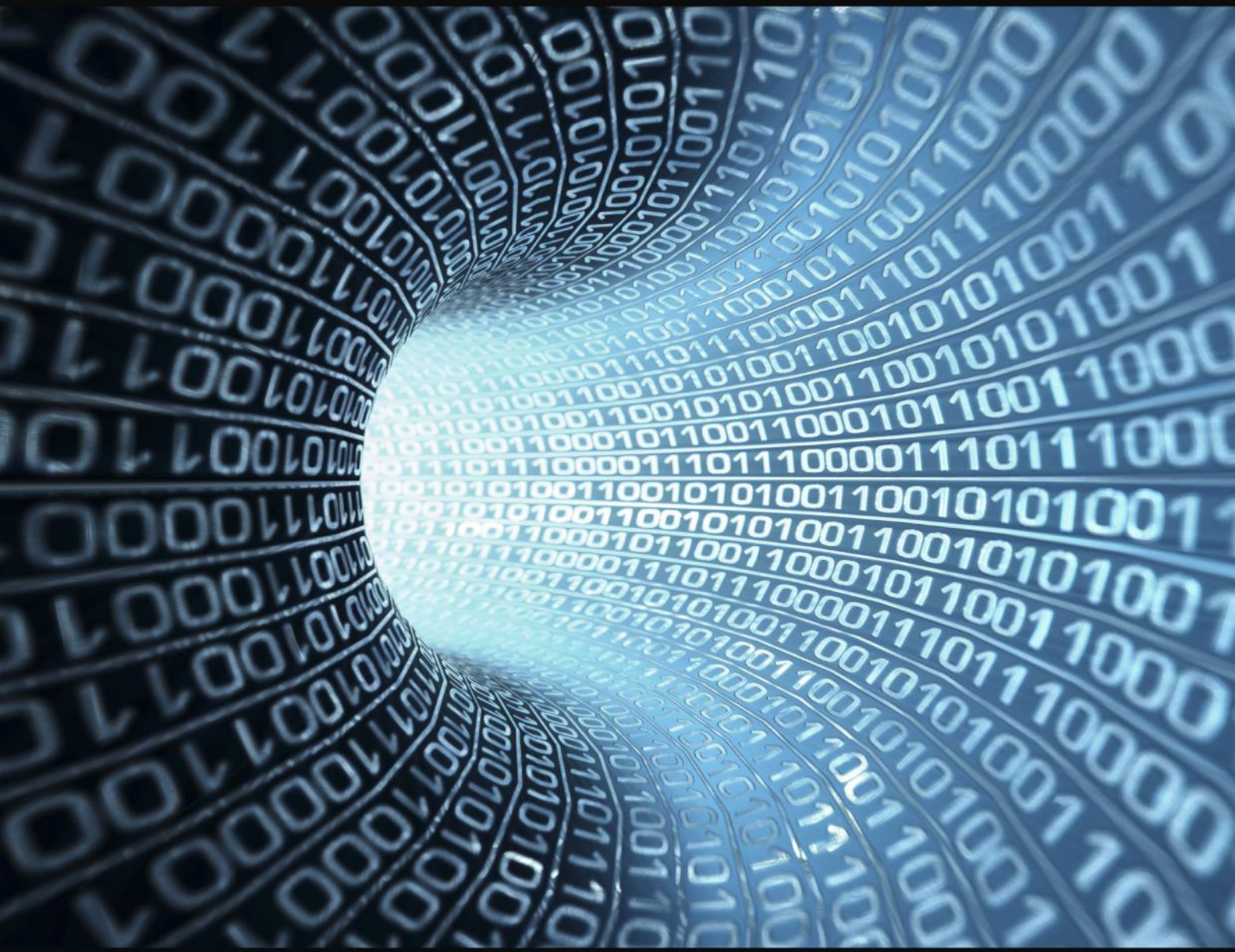
Copernicus Climate Change Service

Climate Data Store (CDS) Workshop Proceedings

3 – 6 March 2015



Funded by the European Union

Implemented by ECMWF

Authors:

Baudouin Raoult, Jean-Noël Thépaut, Dick Dee, Manuel Fuentes (all members of the Workshop organising committee)

Acknowledgements:

We would like to thank all of the Workshop delegates, Presenters and Session Chairs for their contributions to this workshop.

We are very grateful to John Schnase (NASA), Ned Gardiner (NOAA), Stefano Nativi (CNR) and Wim Som de Cerff (KNMI) for chairing the Working Groups and their work to summarise the findings of these groups.

Thanks to Stephan Siemen, Sylvie Lamy-Thepaut, Martin Suttie, Kevin Marsh, Oliver Gorwits, Andrew Brady, Stephen Richards, Samantha Moreby and Simon Witter for their contributions to these proceedings.

Thanks also to Samantha Moreby, Anna Ghelli, Kelly Esteney and other ECMWF Staff for their assistance with this Workshop.

Contents

1 Introduction

This document presents a summary of the “Copernicus Climate Data Store Infrastructure” Workshop which was held at ECMWF between 3rd and 6th March 2015[1].

ECMWF has been entrusted by the European Commission to implement the Copernicus Climate Change Service (C3S). The aim of the C3S is to provide European stakeholders, including public authorities, businesses and citizens, with access to authoritative information about climate change and its impact on society.

The development of the C3S is now underway and one of the first activities involves the design and development of the C3S portal and underlying Climate Data Store (CDS), a distributed facility for providing information about past, present and future climate in terms of Essential Climate Variables and derived climate indicators. The CDS will be at the centre of the C3S software infrastructure, which is also planned to include the, the “Toolbox” (and related “Toolbox” Content), the Catalogue, the Broker and the Web Portal.

2 Workshop description

The aim of the workshop was to discuss the development of the various components of the C3S software infrastructure.

The format of the workshop was a mixture of presentations and working group discussion sessions. The workshop programme itself is given in Appendix 1 and had three broad themes: 'User expectations', 'Existing Climate Service Providers' and 'Industry Perspectives'.

Over 70 participants from European institutions, national meteorological services, research institutes and companies from 19 countries attended. There were over 40 presentations during the course of the Workshop, and 4 of these were made remotely via videoconferencing facilities.

A set of suggested questions were prepared for each of the three days for the 4 Working Groups to discuss, with all of the working groups addressing the same topic on a given day. These were:

Tuesday 3 March: The Catalogue and Portal
Wednesday 4 March: The 'Toolbox'
Thursday 5 March: Content, Standards and Interoperability.

The sets of questions are given in Appendix 2. The summaries of the 4 Working Group discussions are shown in Appendix 3. The presentations made by each Working Group Chairperson are given in Appendix 4.

Note that the summaries contained in this document are as they were presented at the Workshop.

[1] Please note that the views expressed herein are those of the attendees themselves and may not reflect those of any organisation to which they may belong.

3 Summary of Plenary Discussions

One of the main messages voiced by the workshop participants is that user engagement is the key to building a successful Climate Data Store (CDS). For this, an approach for continuous improvement of the portal must be put in place, and the user interface must support a degree of customisation by the user. The portal must also provide different views to different users, depending on their level of expertise and domain knowledge. It should be possible to browse the content of the CDS without login, and any registration requirement should be as simple as possible. Login will be required to get access to actual data, products and services, so that detailed access statistics can be collected, for reporting and capacity planning.

A user forum is also required, along with training for users on how to use the system. A series of use cases should be presented, allowing users to learn by example. Support should be available from a call desk, reachable via the web portal. A “find an expert” facility must be provided by the portal, allowing users to get help from a knowledgeable source on how to interpret data and products from the CDS.

Graphical data must be presented to the users in a consistent manner, to ensure a unified “look and feel”. Presentation of information about uncertainties to non-expert users will need special consideration.

Data in the CDS will primarily be in binary form, although there is a need to support text documents provided they are supplied with adequate metadata. The workshop did not conclude on whether or not socio-economic data should be hosted by the CDS. At this stage the preferred approach would be to provide reference links (URL) to such data.

Data and products suppliers to the CDS will have to follow agreed data management principles. This includes the provision of detailed and accurate metadata information. All data and products should be referenced by a Digital Object Identifier (DOI); this applies both to primary data and any ancillary data that are referenced.

All content of the CDS should be freely available without restriction (Open Data). Support for commercial data and products will be considered at a later stage.

Data and products in the CDS should be made available in various ways. The ability to visualise them directly in the data portal is paramount. The ability to download data must include facilities for subsetting large datasets, re-gridding/re-projecting gridded products, such as fields, and performing format conversions when applicable.

In addition to these basic data transformation facilities, a collection of more advanced tools (the “CDS Toolbox”) will be provided. Users will be able to download these tools, or invoke them from the data portal, by selecting from a list of predefined workflows that can be parameterised. These workflows will operate directly on the data and products in the CDS. When possible, the tools should preserve quality information associated with the input data. Results of the workflows should be available for visualisation in the portal. Providing users with the possibility to upload their own data as input of these tools raises several issues that will have to be considered carefully. Similarly, whether or not the results of these workflows can become part of the CDS should be reviewed at a later stage.

A large number of users will be using the system simultaneously to execute workflows that depend on the users' input parameters. It is therefore of the utmost importance that no one can accidentally (or maliciously) bring the system down by submitting unreasonable requests. A CDS toolbox workflow will run under a scheduler with controlled quality of service based on queues, limits and priorities. Caching of results will also contribute to the performance of the toolbox. The Climate Data Operators (CDO), the NetCDF Operators (NCO) and ECMWF's Metview were mentioned as suitable tools to start building the CDS Toolbox.

Downloading of data and products, as well as invoking the toolbox should not be limited to interactive access. A web-based API should be provided to allow bulk downloads and scripted access to the CDS. The R and Python programming languages were mentioned as good candidates to interact programmatically with the CDS.

While the CDS will primarily support lightweight workflows on small volumes of data, it is likely that some users will require more extensive processing of large amounts of data. To cater for such needs, the use of private, public and hybrid cloud infrastructure must be envisaged.

The CDS must be interoperable with other Copernicus services, as well as with the Infrastructure for Spatial Information in the European Community (INSPIRE), Global Earth Observation System of Systems (GEOSS), the WMO Information System (WIS), the Global Framework for Climate Services (GFCS), and the World Climate Research Programme (WCRP).

The use of standards is essential to ensure interoperability between the CDS and its suppliers, as well as between the CDS and its users. The workshop identified the standards of the International Organization for Standardization (ISO), the Open Geospatial Consortium (OGC), the World Meteorological Organization (WMO) and the Unidata Program from University Corporation for Atmospheric Research (UCAR) as the most relevant to the CDS. Since these standards will evolve over time the C3S should keep a close link to the relevant standardisation committees. SIS users and providers may also be able to provide advice on issues of governance and relevant standards from their own specific domains.

It is vital that the performance of the CDS is continually monitored and assessed. A number of key performance indicators (KPIs) must be defined for this purpose. A bottom-line indicator of the usefulness and success of the system is the year-on-year increase in the number of active users.

An important recommendation of the workshop regarding initial development of the CDS is to start with a set of basic functionalities, allowing early users to provide feedback, and then implement more advanced features over time. To achieve this, ECMWF will have to work in close collaboration with competitively selected contractors to implement the CDS in an iterative fashion. As a result, an agile development methodology is preferred.

4 CDS Workshop Attendees Photograph

Figure 1: Attendees at the ECMWF CDS Workshop, 3-6 March 2015

5 Workshop Resources

Details about the CDS Workshop are available from:

http://www.ecmwf.int/en/copernicus-climate-data-store-workshop

This page also contains links to all of the presentations given at the workshop.

Information about the Copernicus Climate Change Service is available from:

http://www.copernicus-climate.eu/

Information about the Copernicus programme and services is available from:

http://www.copernicus.eu/

and

http://www.copernicus-climate.eu/what-is-copernicus

Appendix 1: CDS Workshop Programme

Tuesday 3 March		
Session 1: User expectations		**Chair: Dick Dee**
09:00-09:20	Welcome / C3S	Jean-Noël Thépaut (ECMWF)
09:20-09:40	Setting the scene	Baudouin Raoult (ECMWF)
09:40-10:00	CLIP-C (Users)	Victoria Bennett (STFC)
10:00-10:20	Use of climate data for EEA activities	André Jol (EEA)
10:20-10:30	Discussions	
10:30-11:00	*Coffee break*	
Session 2: User expectations		**Chair: Manuel Fuentes**
11:00-11:20	Lessons from other dialogues and experiences	Roger Street (UKCIP)
11:20-11:40	Overview on the WMO Climate data and monitoring activities	Omar Baddour (WMO)
11:40-12:00	A carbon cycle dedicated portal (CATLAS) and the needs from the community	Philippe Peylin (LSCE)
12:00-12:20	GFCS and its User Oriented Approach	Lucia Valcarce (GFCS)
12:20-12:40	WMO Information System	Matteo Dell'Acqua (Météo-France)
12:40-13:00	Examples of use of climate data for EEA indicators and Climate-ADAPT	Blaz Kurnik (EEA)
13:00-14:00	*Lunch break*	
Session 3: Existing solutions		**Chair: Roger Street**
14:00-14:20	Exploitation of ECMWF-infrastructures for user-centric innovation: a priority for Europe	Carlos Morais-Pires (DG-CNET) *Remotely presented*
14:20-14:40	From prehistoric ice cores to modern satellite observations - climate data and services from NOAA's National Centers for Environmental Information	Ed Kearns (NCDC) *Remotely presented*
14:40-15:00	Delivering MACC data – what have we learned?	Miha Razinger (ECMWF)

15:00-15:20	User experiences from the Climate Explorer and ECA&D	Geert Jan van Oldenborgh (KNMI)
15:20-15:30	Introduction to the Working groups	Dick Dee & Baudouin Raoult
15:30-16:00	*Coffee break*	
Session 4: Working groups		
16:00-18:00	Working group discussions	
	Close Day 1	
18:00	*Drinks reception*	
Wednesday 4 March		
Session 5: Existing solutions		**Chair: Blaz Kurnik**
09:00-09:20	Perspectives on users expectations from European Policy and through international case studies of Climate Services	Mark Dowell (JRC)
09:20-09:40	GEOSS Common Infrastructure and the GEO DAB	Stefano Nativi (CNR)
09:40-10:00	Insights from 40 years of evolving data services at NCAR's Research Data Archive	Douglas Schuster (NCAR)
10:00-10:20	MyOcean: architecture and data access experience	Sophie Besnard (CLS)
10:20-10:30	Workshop photograph - meet at Reception	
10:30-11:00	*Coffee break*	
Session 6: Existing solutions		**Chair: Stefano Nativi**
11:00-11:20	Drias portal: a climate service	Maryvonne Kerdoncuff (Météo-France)
11:20-11:40	Data Management in support of Climate Services @ MeteoSwiss	Estelle Grüter, (MeteoSwiss)
11:40-12:00	Data Access at EUMETSAT	Harald Rothfuss (EUMETSAT)
12:00-12:20	Operating data services in data.knmi.nl, ECA&D and climate4impact.eu	Wim Som de Cerff (KNMI)
12:20-12:40	The US Climate Resilience Toolkit	Ned Gardiner (NOAA)
12:40-13:00	The EUMETSAT Satellite Application Facility on climate monitoring	Martin Werscheck (DWD)
13:00-14:00	*Lunch break*	
Session 7: Existing solutions		**Chair: Sophie Besnard**

14:00-14:20	The IRI Data Library: enhancing accessibility of climate knowledge	Benno Blumenthal (IRI) *Remotely presented*
14:20-14:40	Climate Analytics as a service	John Schnase (NASA)
14:40-15:00	Climate scenarios for scientists: the Med-CORDEX solution	Marcello Petitta (IC3)
15:00-15:30	Working group discussions	
15:30-16:00	*Coffee break*	
Session 8: Working groups		
16:00-18:00	Working group discussions	
	Close Day 2	
18:00	*Buffet reception*	
Thursday 5 March		
Session 9: Existing solutions		**Chair: Estelle Grüter**
09:00-09:20	CLIP-C	Martin Juckes (STFC)
09:20-09:40	The SPECS experience: climate predictions on the ESGF	Pierre-Antoine Bretonniere (IC3)
09:40-10:00	EUDAT - Open Data Services for Research	Per Öster (EUDAT)
10:00-10:20	European climate modelling infrastructure: accessing climate projections	Sylvie Joussaume (LSCE/IPSL with IS-ENES2 collaborators)
10:30-11:00	*Coffee break*	
Session 10: Industry perspectives		**Chair: Sylvie Joussaume**
11:00-11:20	ECV Data Access in ESA's Climate Change Initiative	Ed Pechorro (ESA)
11:20-11:40	Climate Services: Markets, Systems & Technologies	Chetan Pradhan (CGI)
11:40-12:00	ESA Climate Change Initiative-Soil Moisture (CCI SM): Serving our users – lessons for Copernicus Climate Change Service	Eva Haas (ESA CCI Soil Moisture team) (Geoville)
12:00-12:20	Concepts to Consider in a European Climate Portal and Data Store	Derek Greer (Vega Space)
12:20-12:40	The Multi-sensor Evolution Analysis (MEA) system as a Climate Data Exploitation Platform	Simone Mantovani (MEEO S.r.l.)

12:40-13:00	EO data hosting and processing – core capabilities and emerging solutions	Andrew Groom (Airbus Defence and Space UK)
13:00-14:00	*Lunch break*	
Session 11: Industry perspectives		**Chair: Martin Juckes**
14:00-14:20	Towards a thriving data-driven economy	Remote: Márta Nagy-Rothengass (DG CONNECT)
14:20-14:40	Case Study and reflections in EO Data Management	Richard Campbell (SERCO)
Session 12: Working groups		
14:40-15:30	Working group discussions	
15:30-16:00	*Coffee break*	
16:00-16:30	Working group discussions	
16:30	Close Day 3	
16:30-18:00	Combined Working Group Chairperson's, Centre Representative and Note taker's meeting	
Friday 6 March		
09:00-10:00	Computer Hall tour/Networking/Working group summaries	
Session 13: Plenary		
10:00-12:00	Plenary meeting	
12:00	Close Day 4	

Appendix 2: Working Group Questions

Tuesday 3 March: The Catalogue and Portal

We have provided the questions below to help stimulate discussions in the working groups. Please note they are not a set list of questions that need to be answered. Additional questions and comments are welcome.

- From a user perspective
  - What methods do you prefer to use to interact with the portal?
  - Does your anticipated workflow include both searching and browsing for datasets?
  - Which existing portals and services do you find most useful?
  - Would you share your experience with other users though collaborative platforms, sharing guidelines, tips, comments, annotations, ratings?
  - Should there be a concept of 'personal basket'?
  - What are your 'operational' requirements that the CDS should meet?
    - Timeliness, robustness, availability, ...
  - Does every component of the CDS have the same operational requirements?
- From a provider perspective:
  - What has been learned from the process of developing existing portals?
  - Which aspects do you feel are particularly successful or useful to users?
  - Does your portal evolved over time through user demand to provide increased functionality?
  - How is user registration and communication handled?
  - How do you collect feedback from users?
  - What statistics are generated for the actual portal monitoring and usage?
    - What tools do you use to analyse statistics of usage?
  - What tools/techniques can be used to maintain adequate service levels?
  - How do you manage different user requirements?
  - What is the decision process to choose to use a particular technology for your portal?
  - How do you keep stakeholders closely engaged?
  - How do you evolve your portal over time, ie, how do you scale it up, add new datasets, add new standards, add new functionality

Wednesday 4 March: The 'Toolbox'

We have provided the questions below to help stimulate discussions in the working groups. Please note they are not a set list of questions that need to be answered. Additional questions and comments are welcome.

- From a user perspective
  - How important is that the portal provides a set of analysis tools that can be invoked from a browser?
  - Which kind of processing tools will be of most benefit to you?
  - Do you need the flexibility to manipulate the maps and graphics provided by the C3S portal to tailor them to suit your own needs?
    - What software packages will users use to interact with maps?
  - Should the toolbox be available as a set of downloadable libraries and tools?
    - What programming language should these tools be in?
  - Should the toolbox be a set of analysis services that can be parameterised and invoked remotely?
    - Assuming that these analysis services are compute intensive, would you be ready to pay the services of a cloud provider?
  - Should the C3S provide means to "bring computations to data"?
    - Should this be limited to a prescribed set of tools?
    - Should the C3S portal allow users to upload their own tools?
    - How could Cloud Computing help us?
- From a portal perspective
  - What kind of interactive analysis tools does your portal offer?
  - Do you offer batch access to analysis tools?
    - What protocols/standards are you using?
  - How do you ensure quality of service (e.g. queues, priorities, limits)
    - How do you guarantee a fair use of the services
    - How do you avoid denial of service attacks (even unintentional)
  - How do you "charge/bill" users for using large amount of resources
  - Do you make use of commercial Cloud Services
    - If yes, do you have any experience you would like to share?

Thursday 5 March: Content, Standards and Interoperability

We have provided the questions below to help stimulate discussions in the working groups. Please note they are not a set list of questions that need to be answered. Additional questions and comments are welcome.

- What form of climate information is most useful to you:
  - Maps
  - Graphs
  - Raw data
  - Text reports
- Should socio-economic data for sectoral applications be accessible via the CDS?
- In what form do you expect to get information about data quality and uncertainties?
- How important is quality and provenance, when considering data, products and/or information in the CDS?
- What about support? Documentation, user guidance, do's and don'ts, best practices?
- How to handle non-binary data (ie, information such as reports in word, excel spreadsheets, PDF, …)
- Do you think it is important that users are able to upload their own data?
- What are your preferred output formats?
- What standards are most relevant to your work?
  - OGC standards (WMS, WCS, ...)
  - Unidata tools (Thredds, OpenDAP, ...)
  - Other (Sectorial system specific,...)
- What data products are of most use to the various sectors? Energy, agriculture, etc.
- What other systems should the C3S be interoperable with?

Appendix 3: Working Group Summary Notes

Working Group 1: Tuesday 3 March 2015 - Topic: The Catalogue and Portal

General comment

- Will the CDS be a "simple" data portal or a "climate service"?
  - Users hope for a service, but this means a well-staffed Helpdesk and good information
  - Very different resourcing
  - Is the CDS only handling data in "physical units" or also on the level of the decision makers?

Operational availability of (components of) the system

- Definition of operational: continuous and sustained delivery
  - Control of all of the processing chain
  - Continuous development of service to keep it state-of-the-art (also technology wise)
  - Operational not like weather forecasting
- How will the update frequency be?
- Expectation is that Copernicus stores (archive?) everything
- External events might trigger urgency
  - Do not mix update frequency with availability
- Keep the catalogue up-to-date and available might be already a challenge
- Some support even their portals on weekends (but with existing resources)
- The requirements will be different for the kind of access
  - Human versus programmatic (batch) access

Search & metadata

- Identified as a keen factor for success
  - For example for WIS it was important to have a good search
  - The failure to provide a good search will be ...
    - Increase the load on the Helpdesk
    - Users are unlikely to come back
- Can we provide a good search?
  - Search will depend on who is searching
    - Researcher, decision maker, general public, ...
    - Can we assume we have "educated" users?
      - We need good product user manual and training material
      - We need a good Helpdesk
  - Faceted search
    - Users can choose in which categories they search
  - Community should be involved in the development of the search functionality
    - Search should "learn" over time what is required
    - Define categories in facet search

- Is it enough to offer only English as search language?
    - Offering more will increase complexity and cost dramatically
    - Offering only Angling might limit audience (especially from the general public)

- How can users assess the quality of the datasets?
    - Data set Metadata needs to be visible and searchable
        - Challenge is to have providers to provide metadata
        - Metadata needs to be well structured
        - Can we use DOI?
        - Should we look at the Open Archive Standard?
    - The users would benefit if the quality and evaluation feedback would be visible to them
        - Maturity matrix for service and data sets
    - Link with projects like FP7 CHARMe to link data sets to annotations
        - Might need moderation!
        - Can Copernicus take over the running of the CHARMe server?
            - Important to find long-term solution - nobody invests if it has no future

Marketplace

- Perhaps comparison to Amazon market place is to simplified
    - Users need some training to use the system
- Registration/sign-on
    - Required to offer user specific services (basket, favourite (searches)...)
    - Required to keep statistics on usage of system and data sets (affiliation, usage, requirements, ...)
    - Might be required for some data sets, because of their data policy
- Should we offer a space for favourites (aka ECMWF's *MyRoom*)
    - Users like it, especially id search is not sufficient
    - Danger is users set-up once and do not look what is new on the system
        - Users could be informed about changes to the system and new data sets advertised
- Should the CDS offer "Standing orders"?
    - "Push" versus "Pull"

Working Group 1: Wednesday 4 March 2015 - Topic: The “Toolbox”

What is a Toolbox?

- Some found the expression confusing at the beginning of the discussion
    - We agreed to see it as a list of functionality provided by the C3S/CDS
    - To decide what tools need to be offered we need to know what CDS should provide
        - The commission wants businesses to build on the infrastructure
        - C3S should not replace existing services and build on what already exits

- Enables to make use of data and should enable wide spread use
    - Diversity is a major goal
    - Basic functionality
        - Like OpenDAP
        - Subsetting, regridding, reformatting

What do existing services provide?

- Basic functionality
    - User survey show no demand for more complex functionality
    - Many provide a (simple) preview
    - Users want to use their tools (R, Python, Matlab, …)
    - Currently most users still download all data and process locally
        - Initiatives like Jasmin can change this

Recommendations

- CDS should provide basic functionality to access data
    - Without preventing access to raw data
    - Should aim to be an IaaS
    - Offer WebAPI to access data and offer the processing options
        - Keep users flexible to build their own services
        - This could the definitions of own workflows
        - Start with simple functionally and grow it with what users request
        - Be agile with service provision
            - Be not be afraid to remove functionality (if you see it is not used)
            - Not only in developments
            - Continuous evolution of services to offer modern interfaces
                - API & programming languages
                - To encourage new generation of developers to get involved

Liability

- Who has liability for a poor service build from a Copernicus data set?
    - This cannot be prevented
        - These are not scientific papers which get peer reviewed
- Clear communication needed what Copernicus and C3S will provide and what they will not
    - Disclaimer for users
- Training and education could help to prevent this
- Certification could be counterproductive since it could give taken at a certificate of the service build
- Wrong use is not the risk – the real risk is that the C3S does not find wide spread usage
    - This comes back that the search is vital for the success of CDS

Working Group 1: Thursday 5 March 2015 - Topic: Content, Standards and Interoperability

- All forms of information were mentioned
  - Depends on user groups
    - Multi-layer approach needed
    - We need to find balance and priorities
      - Spectrum of users is large
  - Will the service create data?
    - ECMWF is also service provider
    - It might be more efficient to some work more centrally
    - Some users ask for raw data some for biased corrected data - some want maps which communicate issues (EEA)
    - NOAA portal are nice example, but still provide raw data
      - Different entry points for different group of users - targeted portals
      - All needs look to be integrated

Metadata

- INSPIRE schema should guide
  - More from Author/creation side
  - Who will give information on processing - user might want to know which bias correction was applied
    - What services/data are impacted by a faulty data set?
    - Provider find own solution which are not interoperable
  - Technology exist and should be harvest
    - WMO WIS, GEOS?, ...

Documentation/Education/Outreach

- Users struggle probability and quality measures
  - Depends on user community
- Continuing surveys to access needs -> part of outreach already
  - Better know your users
  - To get more users together than these workshops
  - People might be tired on surveys

Should users be able to upload their data?

- Demand in Sectorial store - local impact data
  - As a place for publication and processing
- Recommendation not to offer it
  - At least in the beginning
  - Then could be an offering for cloud computing
  - WMS overlay

Interoperability

- Big challenge is reliability (see below)
- Long-standing issue - will be a challenge
- Many to many relationship between data and services

- Deeply engrained usages
    - Sea surface different definition between EO and modellers
- Standards do not guarantee interoperability
    - Machine-to-machine also not enough
    - Grib-NetCDF conversion
        - Convert the usage
    - Make use of existing efforts and feed them into CDS
        - WIS, ...
    - Tools - reference implementations for standards
        - Standard "new language" - helps with learning
        - Not all services implement standards well - reference will help
        - Somebody needs to test from outside
        - Any change in the services can cause failure downstream
        - List of standards -> Best Practices with OGC and develop test
        - Standards versus enforcement on formats
            - Committee can enforce policy
            - Evaluation and Quality control role - could check compliance
    - Verify checksum - for robust download of data
    - Different references how can they be mapped
        - Registries in OGC - Dictionaries
        - Help communications with users
    - Handling of non-binary data (reports, PDF)

Reliability

- Challenge is to define SLA
    - No service cannot run on "best effort"
        - What will failure of service mean? 24/7? 9-to-5 might be enough
        - Measure of failure should be defined (what means 20% are offline)
        - Robust access needs to be planned for ...
            - Most unreliable SLA defines service?
            - Discovery 24/7 - data provision less so
            - Portfolio of SLAs
            - If services are paid KPI needs to be defined
                - Payment might help to drive adoption of standards
        - GEOS over 40 different adaptors and portals
        - Portal will have to cope with services being down
    - Flexibility is challenge for scalability and hard to keep resilient
    - How do you estimate cost of WCS request before execution - combine WCS and WPS
    - Solution some have is to limit to users - might not be a good solution

        - Problems with crawlers and repeated download - poorly written software
        - Large request need to be handled
            - Active versus programmatic (batch, routine) users
            - Hard to plan - very speculative numbers of users
            - Users can wait for large request - need to be informed
        - Processing request is limited market
            - Dangers is it overestimated and invest in unnecessary developments
- ESA portal ITT
    - Request open data services
        - Could be problematic when integration in non-open data policy
        - Packages solutions

Find arbitrary starting point --> AGILE

- Be able to respond to the users
- How operational are the various services, especially the projections
- Incremental or streaming of content
    - Depends on the parameter
    - You want similar (model) data together
    - Frequency and access to latest projections/forecast - less than 5 years
        - Quality control and evaluation has to have its place
    - Versioning
        - For operational integrity
        - Clear documentation of changes

Working Group 2: Tuesday 3 March 2015 - Topic: The Catalogue and Portal

Key Issues Discussed

- The make-up of the room was majority domain experts
- Data policy, open or closed. Open!
- Ease of access to data. Required: interoperability, standards
- Role of the provider: just provide data as is or provide tailored products with added value. Big datasets implies both
- Size of datasets (~PB). Dataset reduction strategies. Need for computer processing (sub-setting, time-series, etc) close to the storage
- Vision of a "distributed" or "federated" system
- Trade-offs of "thin" or "thick" functionality at the portal services (client software). Different use cases need both types
- Inclusion of regional data in the C3S?
- Long term maintenance of datasets. Guaranteed availability forever
- Users types: Researcher/Knowledge Worker/Journalist/Policy Maker
- Machine-to-machine. Anyone can build something new
- Innovation based on diversity of providers and data provided
- Use of the industries standards for downloaded data
- Quick looks and visualisation - pre-canned, tailored and bespoke
- Amazon model of a market does not include sufficient QA of provider's datasets (peer-reviewed)
- Need to ensure scientific quality of data and provision. Certification. Gold standard
- FP7 Matrix for Quality Maturity (CORE-CLIMAX?)
- Quality metadata links intrinsically with the data
- Traceability of data from source to consumption
- Fitness of data for user purposes
- Cost factors
  - Expensive can be achieved by the few
  - Cheap can be achieved by the many
- End user
  - Sectoral Information System
  - Questions
    - How can you browse one petabyte of data?
    - What is the data that I need to have?
    - Where to get expertise required to interpret data?
  - Data providers can be anywhere (including out of Europe)
  - Standards conformance
  - Free data and non-free (should the non-free data be visible in the catalogue)?
  - Common shared user registration and single sign on
  - Example of ESA data as compared with NASA
  - End users will follow the path of least resistance
  - Power users (numerical computation) vs End user (data and visualisation for reports)
  - Easy quick high value data queries will be popular
  - Hard, slow, low value data queries will not
  - Importance of robust, simple single sign on and/or registration (SSO)
  - Low impedance user interface is easy to use

- How should data be organised so that key use cases can be met?
- Possibility of native language
- National vs EU data
- How to maintain credit for datasets

- Who will provide data?
  - ECMWF
  - Hadley Centre?
  - NCEI?
- Data contributors

Specific levels of service

- Data policy: open and accessible
- Data traceability
- Data versioning and management
- Digital (e.g. DOI) tags- maintain provenance (essential for end-user) as well as credit (essential for data provider)
- Origination
- Harmonising data. Ensuring comparability
- Need to reduce the size of data to a point where it can be consumed
- Usage information, feedback, usage analytics

Working Group 2: Wednesday 4 March 2015 - Topic: The "Toolbox"

Data Reduction Functions

- Sub-setting
  - Ability to use place names for sub-setting
  - Geo-political shapes
  - Custom upload shapes
  - River basins
  - Temporal
  - Ecosystem
  - By data values
  - Vertical levels
  - Corridors
  - Vertical profiles
- Downscaling
- Re-gridding (interpolation)
- Aggregation
- Filtering
  - QA/QC/Citation/Algorithms/Metadata

Building Blocks

- Data caching
- Onsite vs offsite
- Core datasets
- Analysis Toolbox - downloadable - what language?

- Where are the Toolboxes?
  - One Toolbox catalogue or many?
  - Sector specific Toolboxes
  - Processing next to the data. Centralised versus distributed
- Visualisation
  - Map api
  - Machine-2-machine
  - Catalogue management
  - Maps production
    - publication quality
    - multiple formats
    - GIS ready
- My local data
  - Upload my custom data into client app for comparison to core produced data visualisations
  - Transformation of my data using core tools into comparable forms
- Data store structure compatible with software engineering (map reducible processing)
- Documentation
- Tutorials (simple and detailed)
- Plugins
  - Example R-usage
- Distinguish tools for common use, basic, experts and etc.

Advanced Analytics

- All statistics
  - Histograms
  - Extreme values
  - As graphs
  - As data (ASCII)
  - EOFs, PCAs
  - Spectral, wavelength
  - Differences (distributed data sets)
  - Correlations, regression
  - Ranking (smallest, warmest, etc)
  - Confidence intervals
  - Points and grids
- Hovmueller
- SVP?
- Cross sections
- Upload own code
- Open source
- Ensemble statistics
- Verification and validation

Qualities

- Citable references
- Free and open access

- Transparent
- Educational
- Traceability
- Quality control of data metadata and tools
- Fitness for purpose
- Numerical uncertainty estimates
- Consistency of tools across federated datasets
- Ownership and credit
- Control mutuality
- 24x7 uptime
- speed response interactivity

Collaboration

- Bottom up idea generation methods
- Supporting end users
- Work with non-climate datasets
  - socio-economic data integration
- Usage levels
  - Level 0 – sub-setting data
  - Level 1 - diagnostic calculations, graphs, charts
  - Level 2 - Expert - complex statistical, cross-discipline or across data provider
- Virtual research environment
  - Upload algorithm
  - Share it
  - Run it
- Capturing user feedback on every level and timescale
  - Include during design phase
- Promote communication between agencies and data providers

Sectoral Information Systems

- Tailored, easy-to-use, focussed, build upon CDS service
- Added value for sectors
- Different Toolboxes
- Skill sets needed for sectors will be diverse across sectors
- Organic or spontaneous SIS
- Theme teams to create bespoke/tailored portals with consistent UX/design but targeted content.
  - Content expert (sector)
  - Medium expert (web)
  - Technology expert (CDS)
- Training and tutorials
- Expert Support

Working Group 2: Thursday 5 March 2015 - Topic: Content, Standards and Interoperability

Focus on Sectorial Information needs

- Several categories of users:
  - Interest in warnings information (need of thresholds)
  - Decision support (needs of reports)
- Experienced users will need raw data to support their own analysis
  - Learning curve
- Some products will be created after that
- Some of this products could then be integrated back to the CDS
- Some will stay proprietary and confidential
- Daily support is very important for the success
  - Even if the data is not good, the support should be good)
- Not climate driven, just climate information is needed
- Value added loop
- Value add chain
  - Data source
  - Add value in Sector
  - Submit back to data store
  - Curation/QA
  - Catalogue
  - New data available
- Traceability of data and processing steps (provenance)

Focus on General Standards

- Popularity of Netcdf
- Interoperability standards
  - TDS, OPeNDAP, ftp, WMS, WMC, OGC, XML– all of which will evolve; important to be at the table with the governing bodies so that as data store grows, it is possible help the data model scale accordingly
  - Data provider and C3S who is responsible for enforcing standards?
- Maintaining user relationships. Sectoral End User + Sectoral Expert + Providers
- Spectrum of complexity
  - Simple -> Complex
    - Simple:
    - Easy to achieve
    - Cheap to achieve (or not)
    - Generic and applicable to everyone
    - Web user interface
  - Complex
    - Hard to achieve
    - Expensive to develop
    - Expensive to run
    - Possibly bespoke and applicable to few
    - Scientific possibly or algorithmic
    - May be distributed across datasets

- Computer resources needed close to data

Functional Architecture

- Distributed and centralized architectures
  - Dynamically updated metadata
  - Usage
  - Provenance
  - Traceability
- Processing steps
  - Re-use of popular analyses
  - Harvest user behaviour
  - Cache results to provide them quickly
- How to serve proprietary data requests
  - Confidentiality
  - Security
- Engage SIS representatives in governance and implementation

---

Working Group 3: Tuesday 3 March 2015 - Topic: The Catalogue and Portal

Introduction

- Perspectives from users and providers
- Users are both interactive (human) and non-interactive (scripts)

Functionality

- User interface customisation can be unwieldy and confusing. Bookmarks may be sufficient
- Present different options according to class of user – public/novice/power/expert
- Code templates/examples that make use of the service API
- Wizards to guide a user from an unstructured question to specific results
- A Climate Guide to explain the concepts and terminology (glossary)
- Metadata gaps should be reported by the portal (quality of coverage, rather than content)
- Delta downloads for large datasets (many challenges with this were noted)
- User to user discussion/interaction, facilitation groups, peer support

Policy

- Make login (necessary for EC audit) as easy as possible

Operational Requirements

- INSPIRE compliance; response times of requests
- Quantity of data that can be provided via broker, versus a redirect to data provider
- How to provide service desk, monitoring, operational support, user support (& SLA)
- Documented change management process

Stakeholder Engagement

- Be friendly, timely, and relevant!
- A model of continual service improvement
- KPI for self-assessment on customer/user satisfaction
- Accept user feedback and suggestions – react to this so that users feel engaged
- User training embedded in the portal (rather than a separate site)
- Agile approach – prototypes offered to the community for feedback

Working Group 3: Wednesday 4 March 2015 - Topic: The "Toolbox"

Value-Added Features

- Processing targeted at each sector (terminology, standards/formats)
- Present different options according to class of user – public/novice/power/expert
- Cloud-burst model for more intensive workloads (intensive in terms of…?)
- Charging for more intensive workloads. What is the business model?
- User could take contract directly with cloud provider, but workload provided by CDS

- Organisations wish to recoup investment in libraries and algorithms. IP concerns?
- Visualisation is sometimes more interesting than useful
- Wizards to guide a user from an unstructured question to specific results

Desirable Processing Services

- Sub-setting in space and time domains
- Co-ordinate reference system mapping
- Data format conversion
- Average, standard deviation, min/max (and other examples from NASA)
- Some parameterised options such as thresholding
- Max Planck Institute's Climate Data Operators tools
- Parameterise as much as possible
- Questions scientists wouldn't ask, such as queries based on political boundaries
- Guidelines on what queries make sense (and not)

Data Provenance

- Track fix-ups applied to data within the metadata
- Record the raw source
- Provenance is vital for reporting errors
- Commercial users may use provenance to trace licensing
- Ability to publish processed data-set back to the portal ("community catalogue")
- Highlight to a user that another user has processed a dataset in some way, and how

Computation Specification

- Many votes for Python integration (useable API or code upload). Some requests for R
- OCG processing (WCPS, WPS) and similar
- Clearly defined and published constraints on computing facility (SLD)
- Some commercial sectors will not wish their use/activity to be made public
- User's own virtual machine image avoids licensing/IP issues for CDS provider

Cloud Computing

- Group was very interested in access to a cloud system for processing
- A form of elasticity not available in the core CDS service

Working Group 3: Thursday 5 March 2015 - Topic: Content, Standards and Interoperability

Data Formats

- Animations are good for outreach, but tables of time-series data are essential

- Some debate of what is meant by “raw data” (=> “can be processed”, i.e. not a map)
- What is the real tool for decision-makers?
- Maps, graphs, raw data and text reports
- A process should be in place to establish new requirements through the service lifetime
- Consistency in colour schemes, style sheets (across data providers/sources)
- Should branding of data products be permitted?
- Establish and incorporate best-practice and good-practice (regional/sectorial)

Sectorial Applications (socio-economic data)

- Seen as incredibly challenging
- May not be required at the start, but must still be possible if planned as a second phase
- A hot topic in the WMO – weather warnings and climate data need context
- Alternatively, a different service for socio-economic data and its integration
- Include reference points so that data can be added as an overlay
- Risk of not reaching non-research/scientist audience if these data are not integrated
- Climatologists cannot necessarily determine these requirements

Data Qualities and Uncertainties

- Use standard practices wherever possible
- Some data sources (IPCC) represent this graphically but not in the raw data (?)
- Somewhat subjective – must be well described so users know how to interpret
- Depends on audience – scientific/research vs. decision-maker/commercial (high confidence)
- Providers might not want someone else assigning scores to their data
- Pre-define quality indicators that the providers must include with their data (SLA)

Support, Documentation

- Knowledgebase; best practices; user guides
- Social ecosystem – experts sharing knowledge and experience with a community
- An “ask the expert” (consultancy) service
- Champions (“ambassadors”) and power-users

Non-Binary Data

- Ignore generic documents, this is not a document management system
- Weather data plus a small number of open standard formats
- Support open formats such as PDF that can contain metadata and be future-proof
- Digital Object Identifiers

Data Formats

- NetCDF/CF, CSV, ASCII
- CF is good for grid and time-series, but too permissive
- Versions supported should be agreed and published
- A true archive would not alter data formats; if the data itself isn't changed this is acceptable

Interoperability

- INSPIRE, WIS, GEOSS

Working Group 4: Tuesday 3 March 2015 - Topic: The Catalogue and Portal

The discussion centred on what the CDS data portal should provide to the users and how to achieve that.
The following set of topics focus on the development and management of the portal.

Development of the portal

Météo-France works with contractors/industry through procurements

- Web data portal development is performed using agile methods
- Only high level requirements are specified

Monitoring usage of the data portal

The following suggestions were made on how best to monitor the usage of the portal

- Implement a user authentication process
- Make use of system logs and web statistics
- Record data volumes downloaded
- Follow how many users are accessing the portal on a weekly basis
- Track which products are most and least used, this could help in deciding whether to withdraw a product from the portal
- Send a questionnaire to users on a yearly basis covering their usage of the portal

Suggested improvements to existing portals

The following points were suggestions by providers on how they could improve their existing portals.

- Perform “A/B” testing for finding user preferences. Implement two “methods” of something, e.g. user registration, and see which one is used most by users
- Perform task oriented testing with real users
- Make it easier to publish new datasets, improve catalogue management
- Add useful information, e.g. climate information bulletins, extreme events, attribution information
- Some governments are working on portals to include more themes, e.g. climate, health,… (mini-Copernicus at national level)
- Split user interface from infrastructure needed to provide data, so that the user interface can evolve more frequently

User interaction with the portal

- Users generally know where to find the data. However, users like to get notification of a new product (even if the user is knowledgeable)
- Layering of users, with different user experience is suggested
  - non-expert: need guidance, examples, case studies
  - decision-makers: need set of tools to help evaluate the data
  - scientists: like to have the raw data, but appreciate basic browsing and visualisation capability
- Two types of access were identified:
  - Daily updates for many users (free, open, …)
  - Large datasets (associated data policy, registration, authentication, …)

User and stakeholder engagement

- Hold user meetings on at least an annual basis, if not more frequently
- Users to communicate via the service Help Desk for operational products
- Experiences so far from CLIPC
  - User workshops and dialogues, performing tests with “fresh”/”novice” users gives very useful feedback
  - Good to have short cycles of interaction with users, perform user interviews
- WMO perform User forums with Nat. Met services to provide advice to users and gather user feedback

The following set of points reflect the discussions on data policy issues and requirements for the data portal.

Data policy

- Should Copernicus contain restricted data? (high quality observations restricted by data provider)
- Recommendations:
  - Copernicus data should be “open”
  - If providers do not want their data to be openly available, they should not contribute (However, there was a recommendation at the plenary session on the final day of the workshop to not exclude the possibility of a provider uploading a proprietary data set to the CDS.)
  - There should preferably be a single licence

Operational requirements

- Climate data takes time and effort to quality check
- Science related to climate is changing every day: we may require to have different versions of the same dataset, having different methodologies applied, and to include more data
- Documentation of quality is very important: users should have visibility to the process by which the data was generated
- Should we flag quality/maturity of products? (Related projects: QA4ECV, CHARMe)
- Copernicus should make use of DOIs for data and derived products
- Quality information is very important, not all needs to be exposed to the user, but we need to have the details available if required

Metadata requirements

- Several types of metadata seem necessary:
  - Discovery metadata (WIS-type)
  - Descriptive metadata (WIGOS-type)
  - Quality of the data/products
- What metadata? Maybe can’t tell until we get users
  - Start with searchable metadata
  - Later, if required, provide more descriptive metadata
  - Copernicus service should provide information on quality, not whether a product is fit-for-purpose - the user will decide this

- If Copernicus service provider is liable for the quality and data must be openly available: how to balance that? If quality standard is too high, how can we make data freely available?

Working Group 4: Wednesday 4 March 2015 - Topic: The "Toolbox"

This discussion was on the kind of tools that should be available to the users through the CDS to enable them to explore, analyse and make use of the data available through the CDS.

Provision of browser-based tools

- Different users will need different Toolboxes
- User categories in CLIPC are: climate scientists, impact scientists, boundary workers, policy makers
- Be careful to also cater for public users, if not they will go elsewhere for their information
- C3S should avoid conflict with services already provided by National Met Services, C3S should cover the gaps
- NMS need to be climate communicators and may need to act swiftly, therefore browser-based analysis tools would be useful (browse, visualise, first-order analytics)
- Visualisation is the first way to explore the data (could be static), dynamic visualisation available depending on computing resources
- Toolbox could be useful for handling and hiding different data types
- Key requirement for the Toolbox: flexible and adaptable to new users

Provision of remotely invoked tools

- Many Toolboxes could be foreseen both on the service side and the client side
- There are many downloadable Toolboxes available
- Perform a review of the various downloadable Toolboxes tools that already exist (preferably open source tools)
- There should be a preference for standardised Toolboxes

Handling of data quality in the Toolbox

- Would expect all data provided by C3S are thoroughly validated and of a known level (preferably high) of quality
- Type of data will determine what kind of quality information should be provided
- Should the user be limited to only using tools on certain quality controlled data? Better to provide good quality metadata and training to inform the user rather than limiting their use of the data
- The Toolbox should take quality information into account
- ESA CCI has recently issued a procurement to develop a Toolbox for satellite data ECVs, it may be useful for C3S to take a look at this (since the meeting we have had access to the procurement text)

Paying for compute services

- This could be an important issue, the data may be freely available but compute costs need to be covered
- Academia and some government services rarely want to pay for services
- We could distinguish between service levels; free computation could take a long time, paid-for service delivered more quickly

General comments on the Toolbox

- Do not confuse the Toolbox with the basic functionality expected (e.g. reformatting, re-gridding, sub-setting, simple visualisation)
- A “playground” for users to experiment with tools would be useful, the C3S could also learn from such a playground.
- Community will develop resources/tools based on data availability through C3S
- Scientific users will probably request higher volumes of data, users who are more targeted (impact scientists, boundary workers) will want more data reduction before use
- Initially C3S should offer a basic service for data analysis and then learn and evolve
- Don't try to over-protect bad usage of the data, learn as the service develops
- Make sure the resources are available to perform the necessary post-processing
- Batch jobs can be run more cheaply, but may not be suitable for some users, e.g. policy-makers
- How to handle machine-machine access, requesting a lot of data:
  - need to manage by applying limits, user identity needed for heavier users to have opportunity to control access
- Some aspect of pre-calculation may be required, e.g. to fulfil requirements from sectorial information system
- Provide annual updates of plots of climate variables for policy makers, high quality data sets for climate scientists
- Calculation methods should be open for users to see
- The Toolbox should be open with the possibility of users to peer-review and/or comment
- Case studies and use cases should be employed to guide the users and help avoid pitfalls - encourage good practice (very strong point)
- Don't forget the people providing the support, it's not all about technology

Working Group 4: Thursday 5 March 2015 - Topic: Content, Standards and Interoperability

Provision of climate information

- Raw data definitely required
- Graphs and maps needed for data discovery
- Not so much interest for text reports, unless they are output from SIS: text reports could be value added products generated by downstream users

- Many different users need different types of maps; global, regional, local

Provision of socio-economic information

- Socio-economic information comes from many different sources, the granularity of socio-economic data can be quite different from climate data (different scale of time steps)
- Could be a very large effort to include in CDS
- Should probably provide links to this data, if included in CDS then CDS is responsible for the quality control. Better to link to sources of socio-economic data, e.g. EUROSTAT

Data quality and uncertainties

- There should be a clear distinction between quality and uncertainty
- If sectioning data, quality data provided should be sectioned too
- Standards for quality? skill scores, etc like in meteorology
- How should data quality be conveyed to the policy maker? Is it the responsibility of the intermediary?
- Different terminologies exist between different communities, e.g. satellite ECVs and climate modelling - ESA CCI is making a contribution to mapping/harmonising these terminologies
- Commentary on how useful the data is should be provided; strengths and weaknesses
- Users should be invited to comment and give feedback on the quality of data
- Quality and provenance are perceived to be of the utmost importance

Handling non-binary data

- Store non-binary data and make them searchable through internet search engines
- Knowledge management systems are available to handle non-binary data
- Non-binary data should be linked to the data
- Non-binary data should have appropriate metadata associated with it
- Spreadsheets, auxiliary information, etc should have a DOI

Support to users

- Use multi-media, walkthroughs recommended
- Employ an issue tracking system for user support
- Use of the following; user forum area, self-help, user community help, discussions on data set level, commentary metadata (CHARMe)
- Sharing platforms (social media sites could be of help here)

Output formats

- For raw data the consensus was to use NetCDF
- Output format may depend on the type of information, e.g. indicators are not in NetCDF
- For sectorial information, industry standard formats may be specific by sector

- It would be useful to have scripted access to CDS data, e.g. using "R" or Python
- Users will ask for some simple formats (excel, csv, kmz) for small data volumes
- What formats suit policy makers and education?

Standards

The following standards were suggested without too much debate.

- INSPIRE, WMO Core Metadata Profile, WIS
- OGC standards, Unidata tools, JSON for web users
- NetCDF CF for raw data
- Which standards apply to Sectorial Information?
- Good to have the capability to map metadata to different metadata standards

Interoperability

- C3S should be interoperable with GEOSS, WIS and any repository of climate data from Nat. Met. Services
- ESA implemented HMA (heterogeneous multi-mission accessibility) - could C3S do something analogous?
- We should seek input from users searching for or browsing the data
- User should not need to know about interoperability
- Access should be easy and transparent
- Scalability for numbers of users is necessary; limits on numbers of requests, policy clear to users at the beginning, implement a "throttle" mechanism
- Some level of separation between "popular" products and research-based products

Appendix 4: Working Group Plenary Presentations

Working Group 1: Plenary Summary

WG1 – The Catalogue and Portal
Recommended Topics for Key Decisions

Support

- Is CDS a Data Portal?
  With technical support for discovery and access …
- Is CDS a Climate Service?
  With well-staffed help desk, decision support for a diverse audience …

Availability

- What data will be included?
  Inclusion of national datasets has implications …
- Update frequency
  Policies and processes fordataset updates and emergency response …
- Continuous development
  To ensure state-of-the-art technologies and science practices…

Discovery

- Effective mechanisms are critical
  - *Catalogue currency*
  - *Rich, abundant metadata*
  - *Robust, multimodal search*
- Language policies
- Community feedback

Personalization

- Registration policies
  Anonymous user vs registered user capabilities …
- "MySpace" support
  A convenience to the user and great way to acquire metrics …
- Tailored push/pull delivery of products and notifications

WG1 – The Toolkit
Key Decisions and Recommendations

Objectives

- What is a Toolkit?
  Toolkit = Functionality = C3S CDS Services …
- What are the metrics for success?
  Primary goal is to mobilize climate data to meet EC objectives: promote job creation, market growth, security …

Risks

- Who's liable for incorrect or malicious use of C3S CDS data or services?
  Probably not something we can control …
- Mitigation strategies
  Disclaimers, explanations, training, education …
- Greatest risk is failed opportunity
  Need new attitude about liability – the greatest risk is that C3S CDS data do not get widely used …

Provisioning

- Need a fundamentally new and modern approach to service deployment that promotes technical innovation and marketplace build-out of advanced capabilities
- Key architectural elements:
  - *Core C3S CDS functions that focus on data mobilization: subsetting, regridding, reformatting and workflows …*
  - *A C3S CDS Web API that exposes core functions and promotes private-sector application development and innovation …*
- Adopt agile development and deployment approach

  Develop and deploy core capabilities as simple, discrete, circumscribed, low-overhead operations that can be easily tested, discarded, or adopted in response to customer needs …

WG1 – Content, Standards, and Interoperability
Key Decisions and Recommendations

Content

- An adaptive, generalizable process is needed
  - *Identify the attributes driving content decisions: size, type, state (raw, refined), SLAs (upstream and downstream), …*
  - *Establish an reasonable initial policy …*
  - *Enable a mechanism to solicit and respond to customer feedback …*
- Data uploads not needed now

Standards

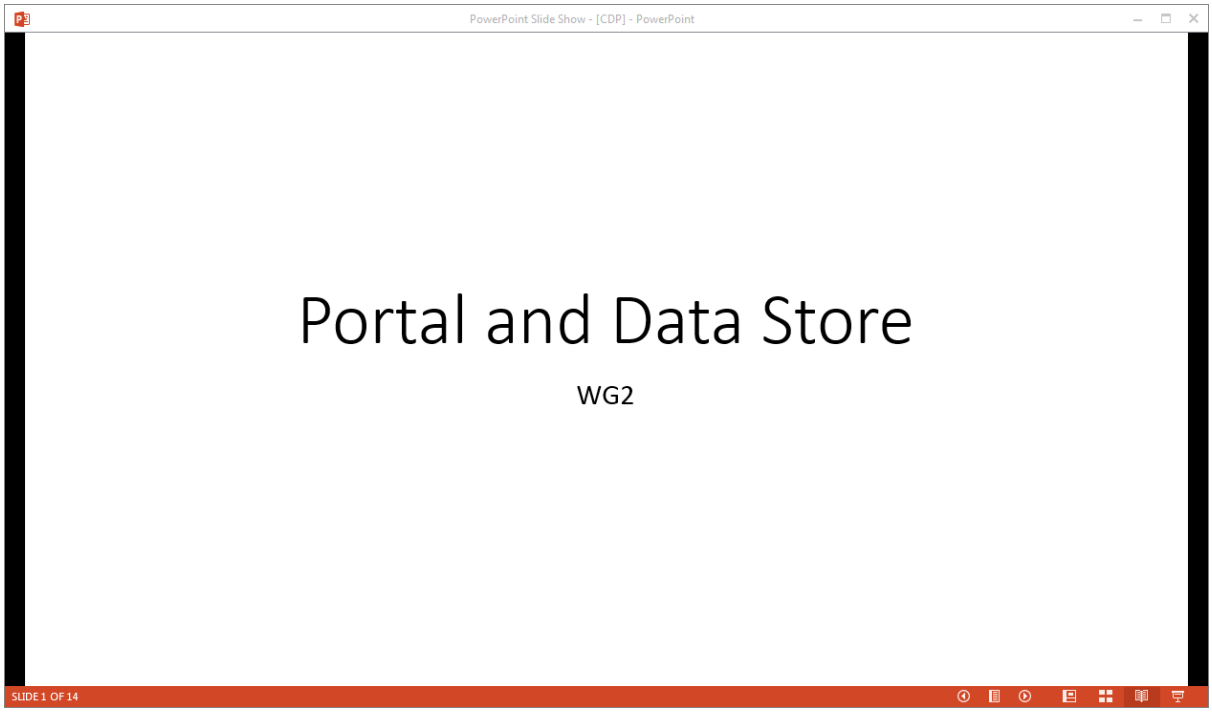
- See above
- Provide reference implementations on existing standards and formats
- Include standards review entity as part of the C3S CDS governance model

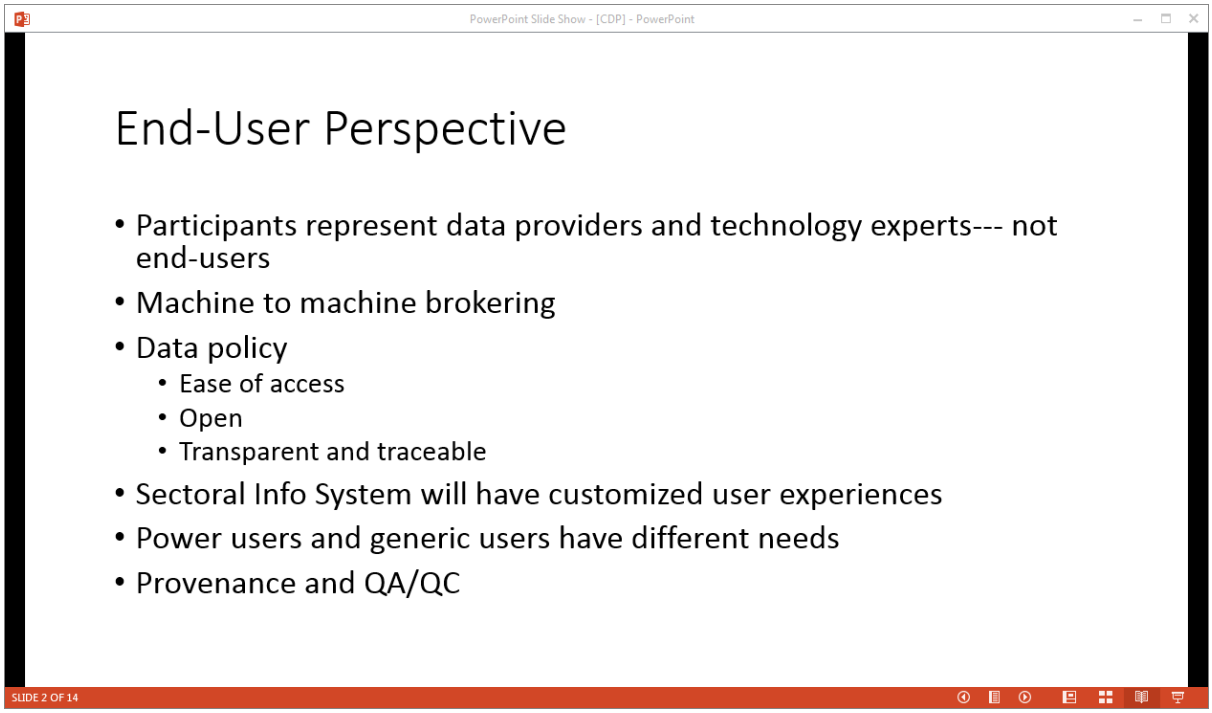
Interoperability

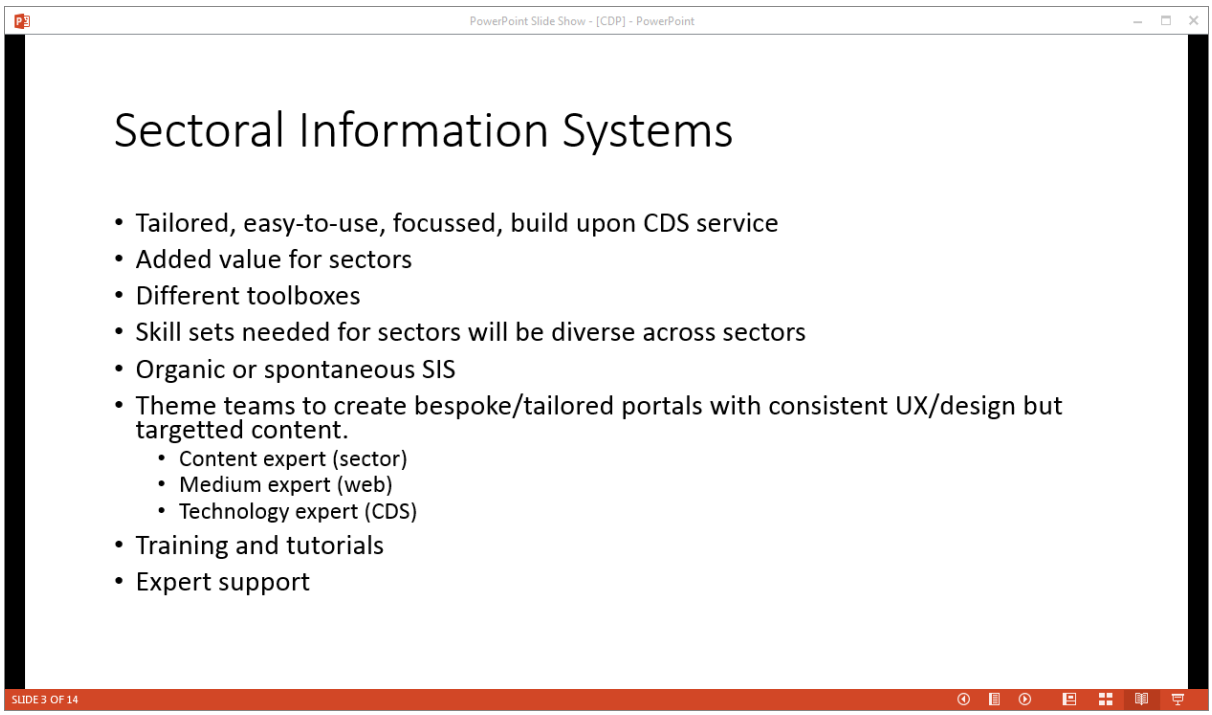
- Many dimensions to interoperability
  - Many-to-many relationships among data, tools, services
  - Machine2Machine, Application2Application, Community2Community (e.g. expression of probabilities, different ontologies, etc.)
  - *Need an adaptive, generalizable process like above*
- The complexities of interoperability put reliability and quality at risk

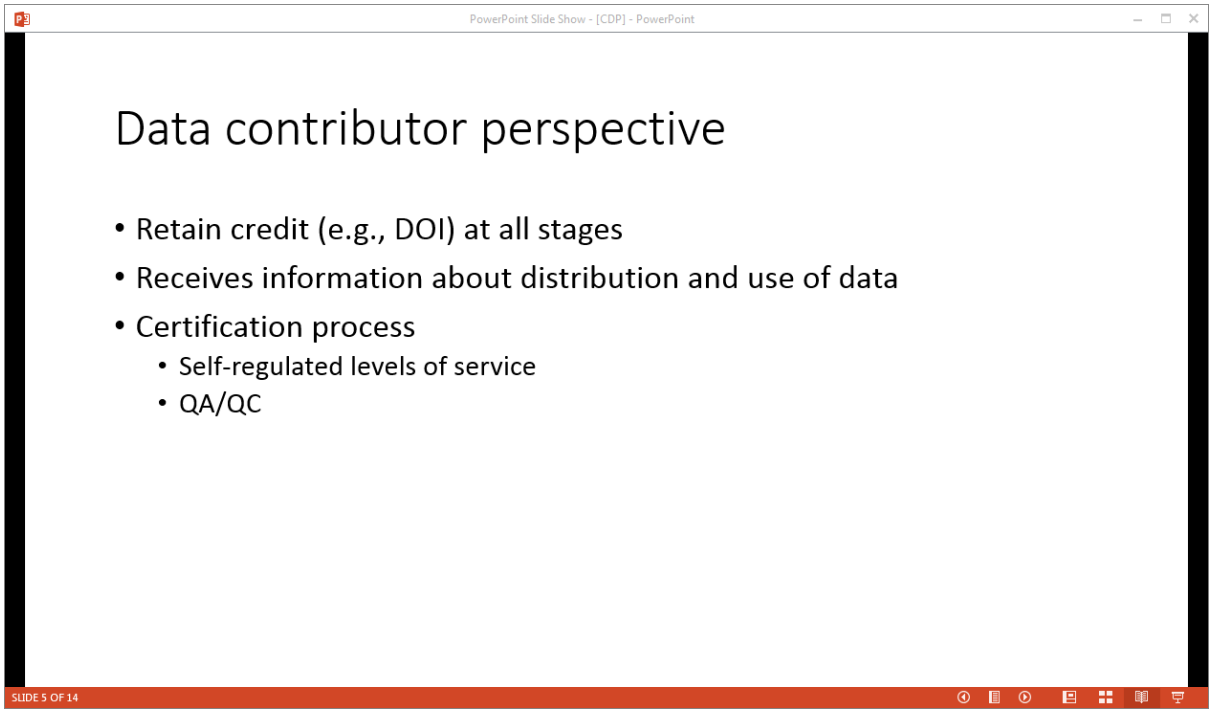
  Portal cannot prevent it, but it can do things to mitigate risk: routine external testing of services, contractual enforcement of quality controls, e.g. checksums, guidance through best practices guidelines and reference implementations…

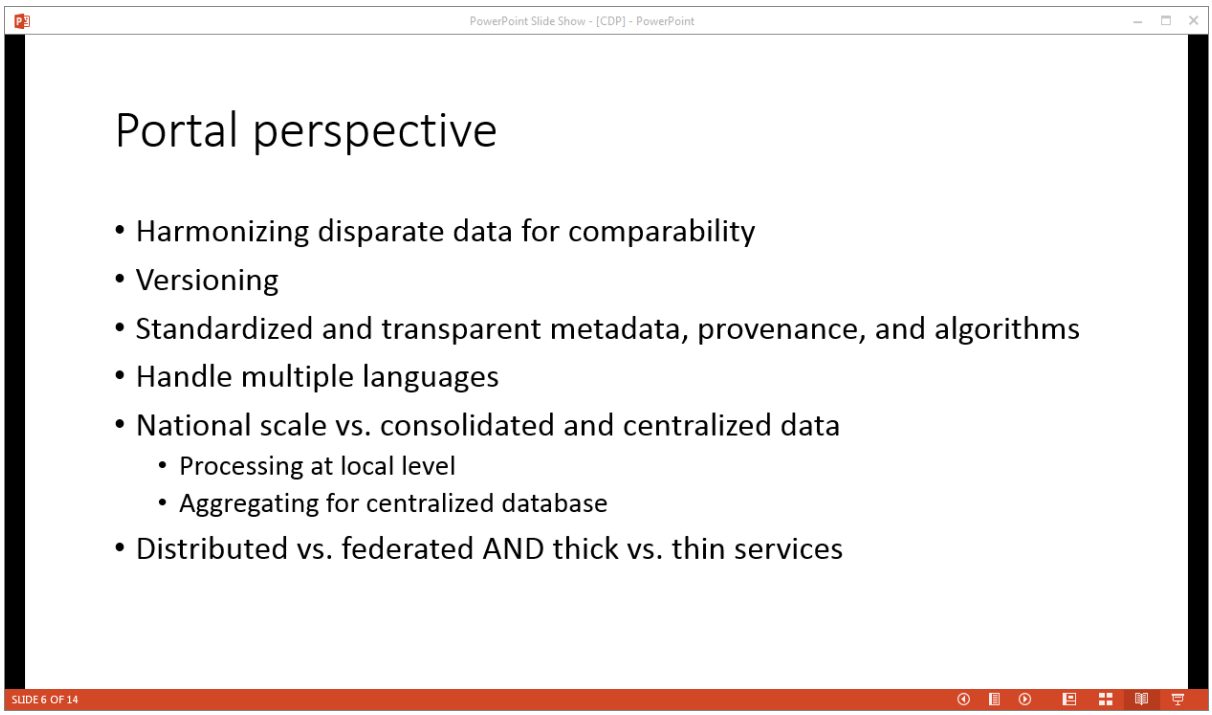
Working Group 2: Plenary Summary

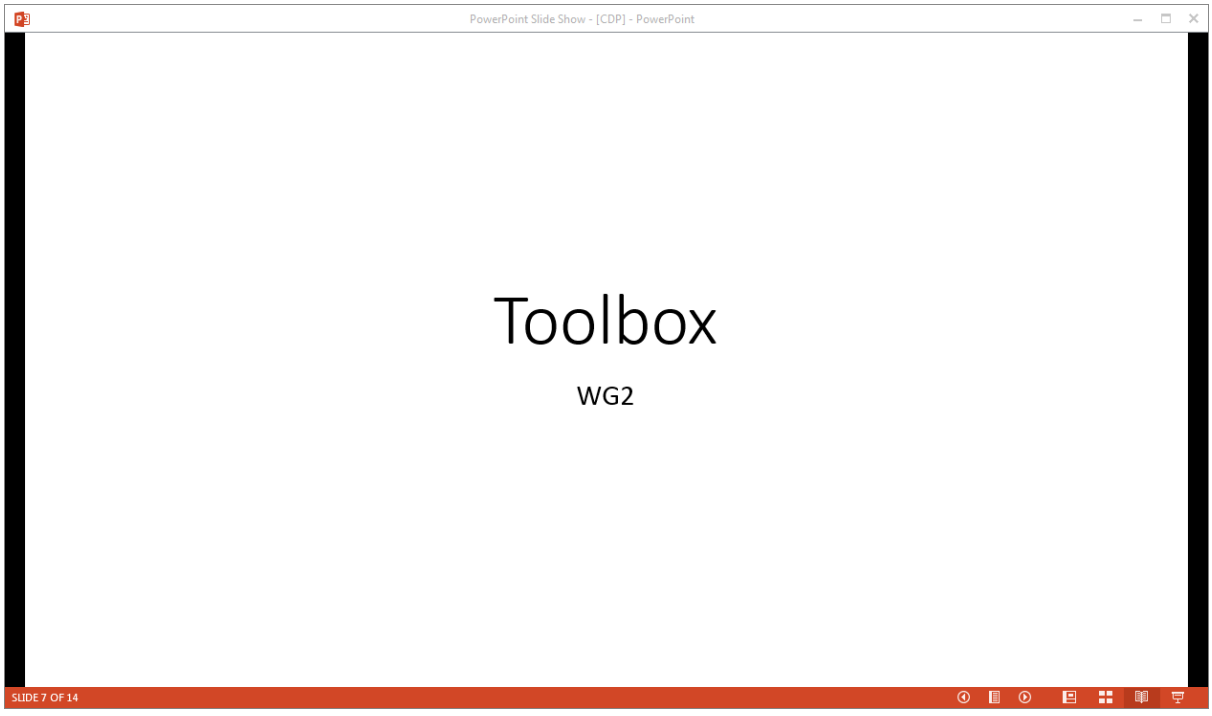


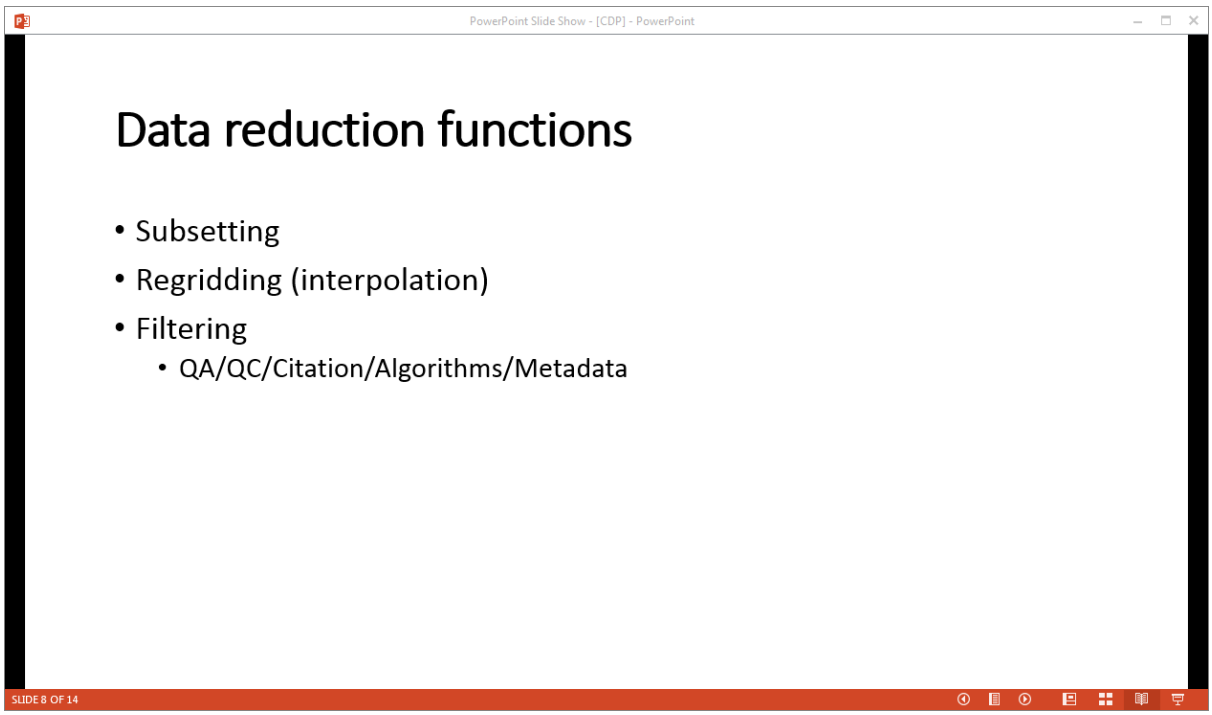


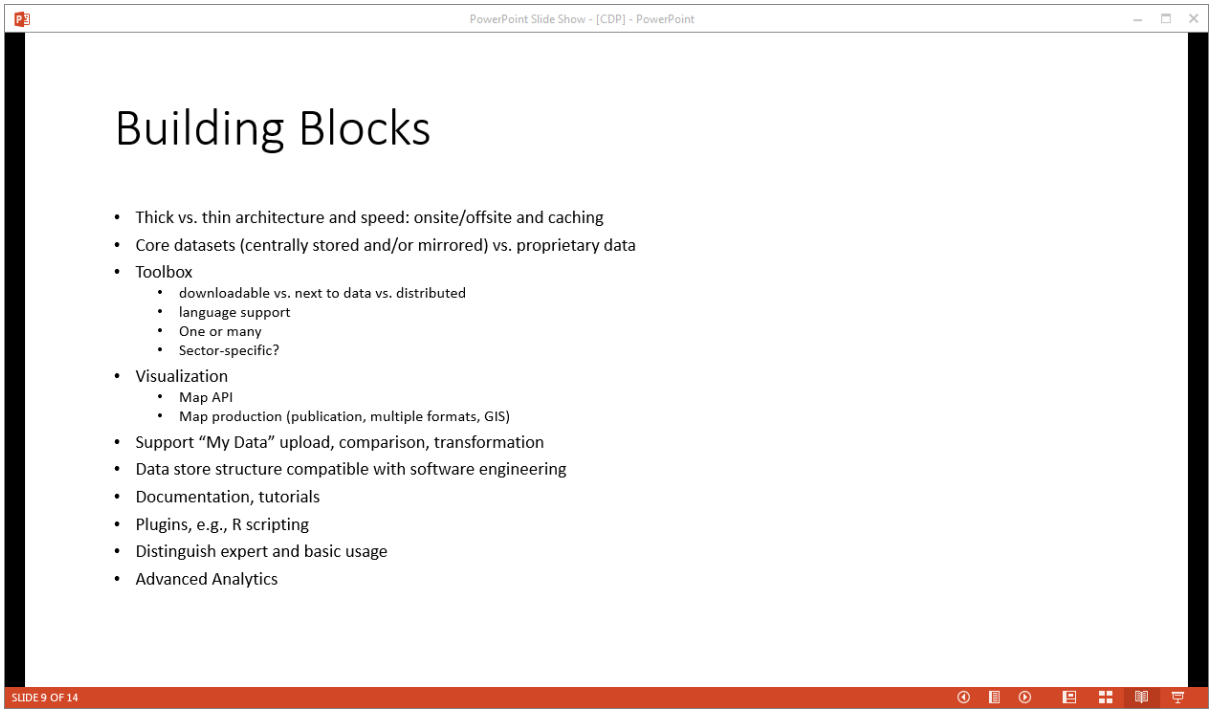

PowerPoint Slide Show - [CDP] - PowerPoint
Sectoral Information Systems
• Tailored, easy-to-use, focussed, build upon CDS service
• Added value for sectors
• Different toolboxes
• Skill sets needed for sectors will be diverse across sectors
• Organic or spontaneous SIS
• Theme teams to create bespoke/tailored portals with consistent UX/design but targetted content.
  • Content expert (sector)
  • Medium expert (web)
  • Technology expert (CDS)
• Training and tutorials
• Expert support
SLIDE 3 OF 14

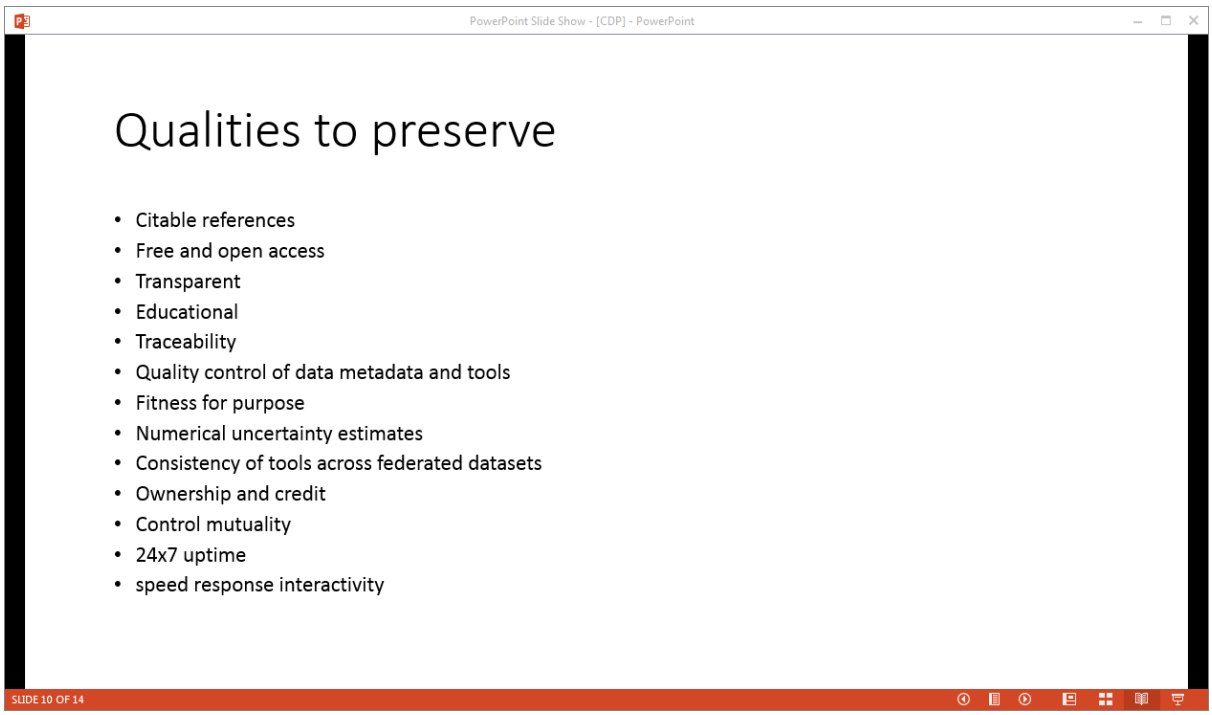

PowerPoint Slide Show - [CDP] - PowerPoint
Data contributor perspective
• Retain credit (e.g., DOI) at all stages
• Receives information about distribution and use of data
• Certification process
  • Self-regulated levels of service
  • QA/QC
SLIDE 5 OF 14


PowerPoint Slide Show - [CDP] - PowerPoint
Portal perspective
• Harmonizing disparate data for comparability
• Versioning
• Standardized and transparent metadata, provenance, and algorithms
• Handle multiple languages
• National scale vs. consolidated and centralized data
• Processing at local level
• Aggregating for centralized database
• Distributed vs. federated AND thick vs. thin services
SLIDE 6 OF 14


PowerPoint Slide Show - [CDP] - PowerPoint
Toolbox
WG2
SLIDE 7 OF 14

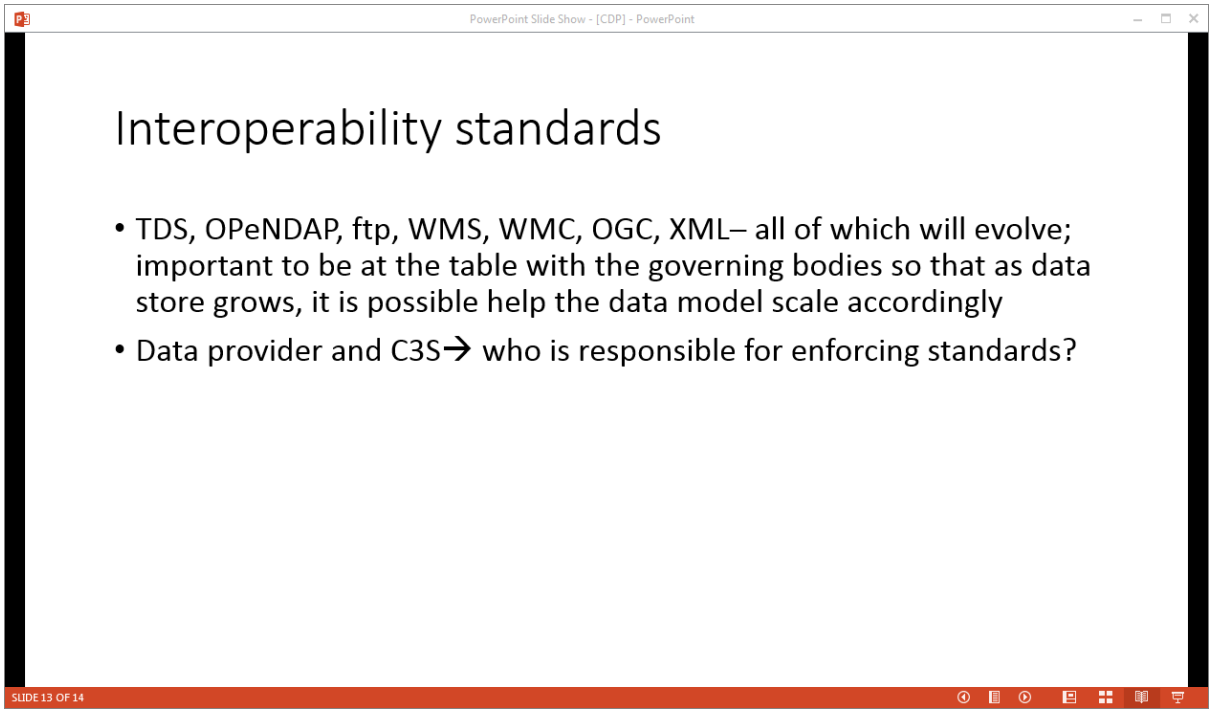

PowerPoint Slide Show - [CDP] - PowerPoint
Data reduction functions
• Subsetting
• Regridding (interpolation)
• Filtering
• QA/QC/Citation/Algorithms/Metadata
SLIDE 8 OF 14

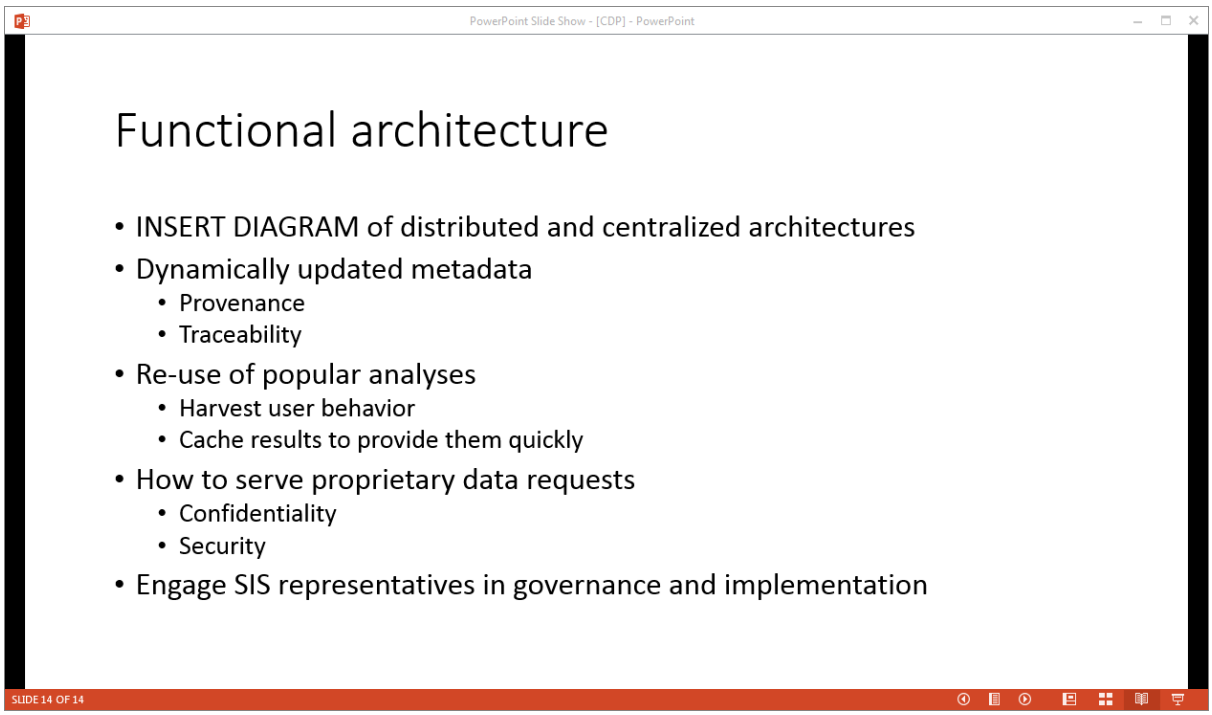

PowerPoint Slide Show - [CDP] - PowerPoint
Building Blocks
• Thick vs. thin architecture and speed: onsite/offsite and caching
• Core datasets (centrally stored and/or mirrored) vs. proprietary data
• Toolbox
• downloadable vs. next to data vs. distributed
• language support
• One or many
• Sector-specific?
• Visualization
• Map API
• Map production (publication, multiple formats, GIS)
• Support "My Data" upload, comparison, transformation
• Data store structure compatible with software engineering
• Documentation, tutorials
• Plugins, e.g., R scripting
• Distinguish expert and basic usage
• Advanced Analytics
SLIDE 9 OF 14


PowerPoint Slide Show - [CDP] - PowerPoint
Qualities to preserve
Citable references
Free and open access
Transparent
Educational
Traceability
Quality control of data metadata and tools
Fitness for purpose
Numerical uncertainty estimates
Consistency of tools across federated datasets
Ownership and credit
Control mutuality
24x7 uptime
speed response interactivity
SLIDE 10 OF 14

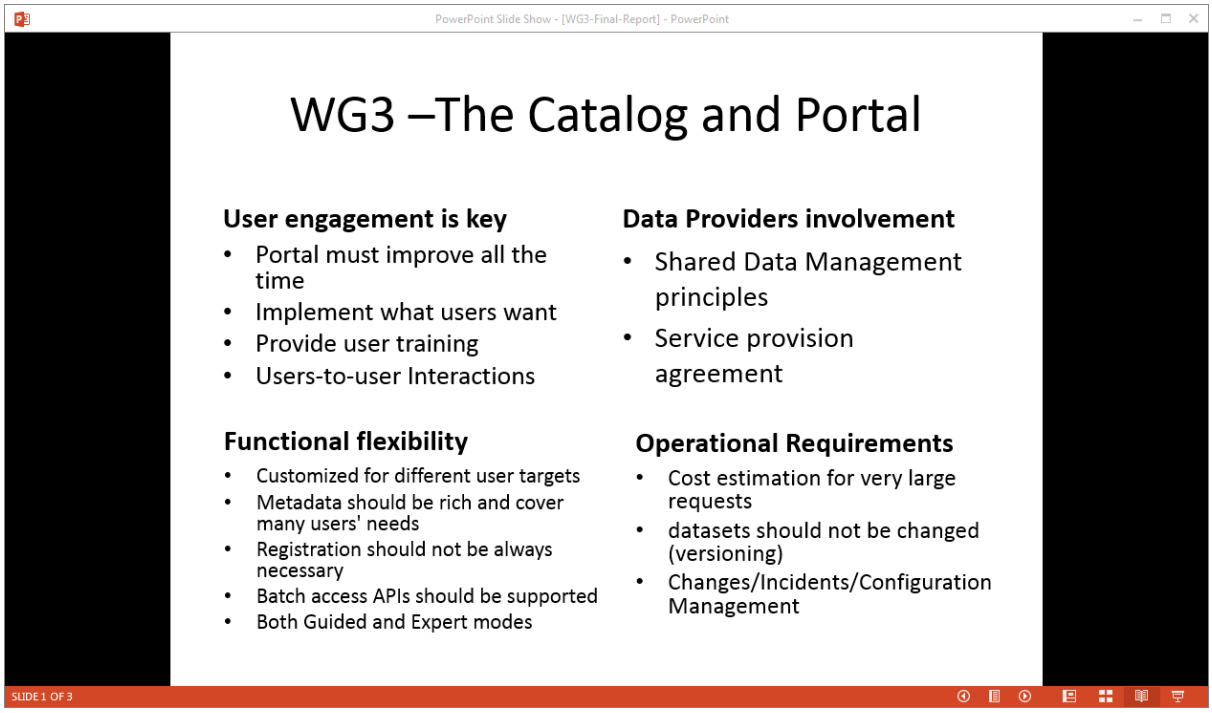

PowerPoint Slide Show - [CDP] - PowerPoint
Collaboration
Bottom up idea generation methods
Supporting end users
Work with non climate datasets
socio-economic data integration
Usage levels
Level 0 - subsetting data
Level 1 - diagnostic calculations, graphs, charts
Level 2 - Expert - complex statistical, cross-discipline or across data provider
Virtual research environment
Upload algorithm
Share it
Run it
Capturing user feedback on every level and timescale
Include during design phase
Promote communication between agencies and data providers.
SLIDE 11 OF 14

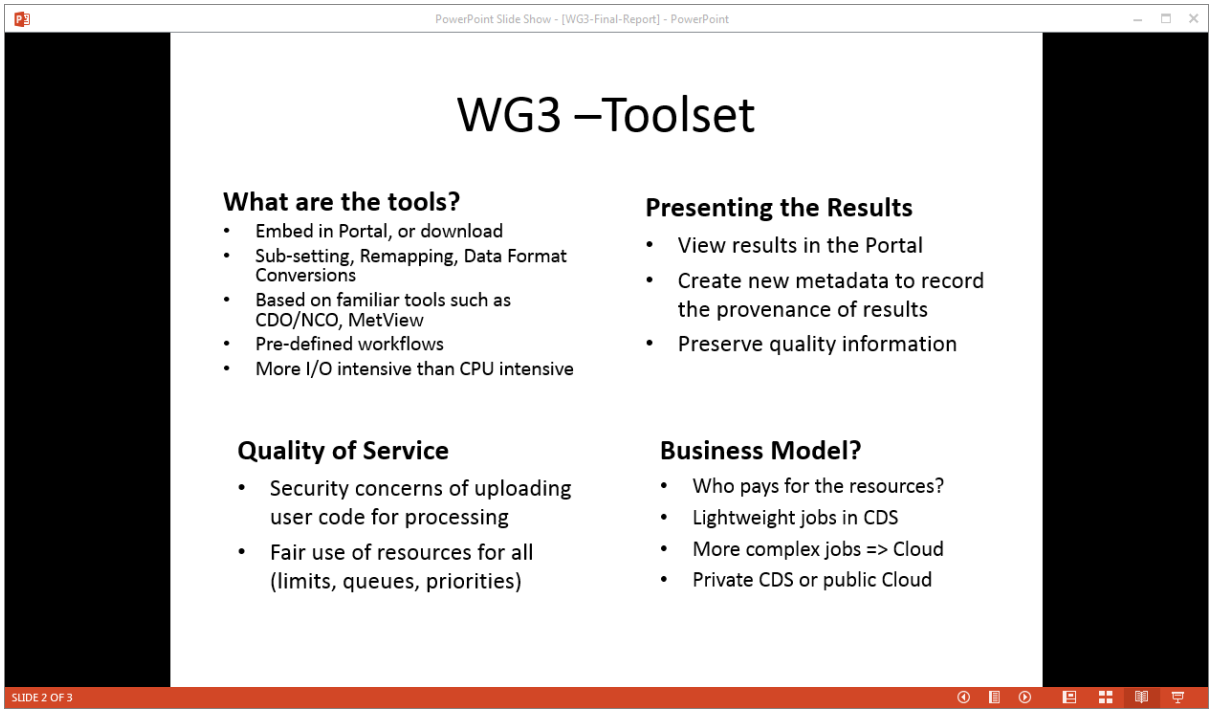

PowerPoint Slide Show - [CDP] - PowerPoint
Content Standards and Interoperability
WG2
SLIDE 12 OF 14

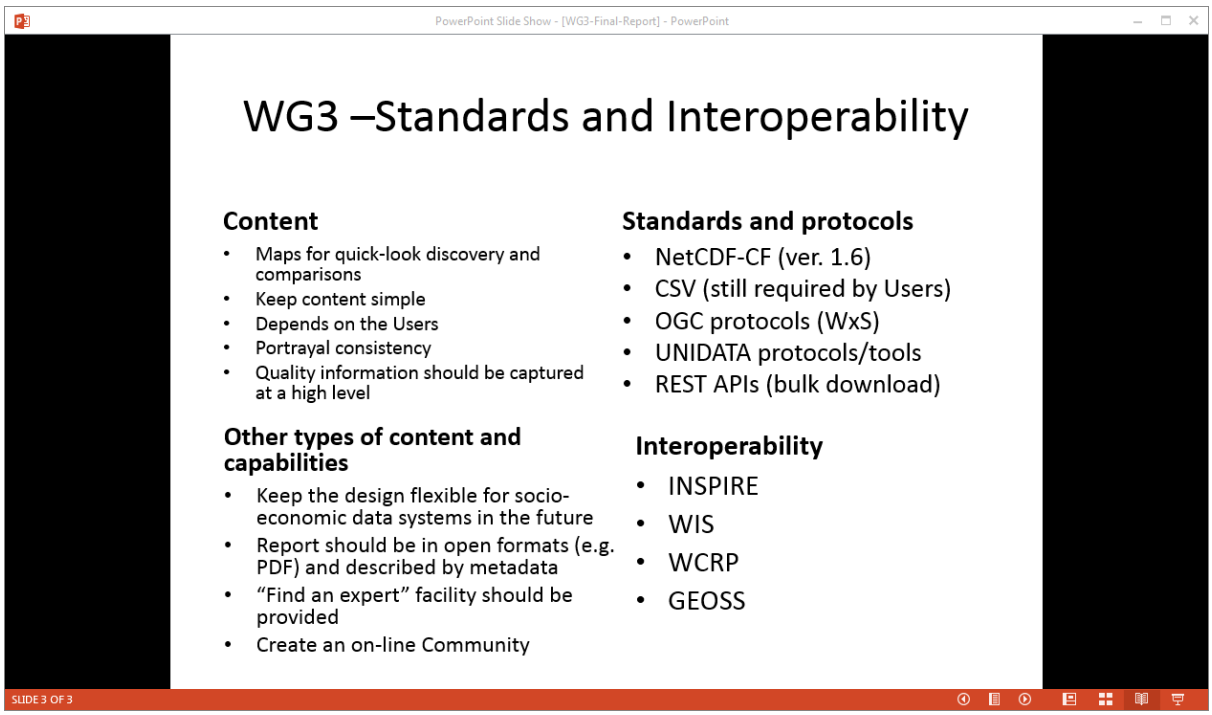

PowerPoint Slide Show - [CDP] - PowerPoint
Interoperability standards
• TDS, OPeNDAP, ftp, WMS, WMC, OGC, XML– all of which will evolve; important to be at the table with the governing bodies so that as data store grows, it is possible help the data model scale accordingly
• Data provider and C3S→ who is responsible for enforcing standards?
SLIDE 13 OF 14


PowerPoint Slide Show - [CDP] - PowerPoint
Functional architecture
• INSERT DIAGRAM of distributed and centralized architectures
• Dynamically updated metadata
• Provenance
• Traceability
• Re-use of popular analyses
• Harvest user behavior
• Cache results to provide them quickly
• How to serve proprietary data requests
• Confidentiality
• Security
• Engage SIS representatives in governance and implementation
SLIDE 14 OF 14

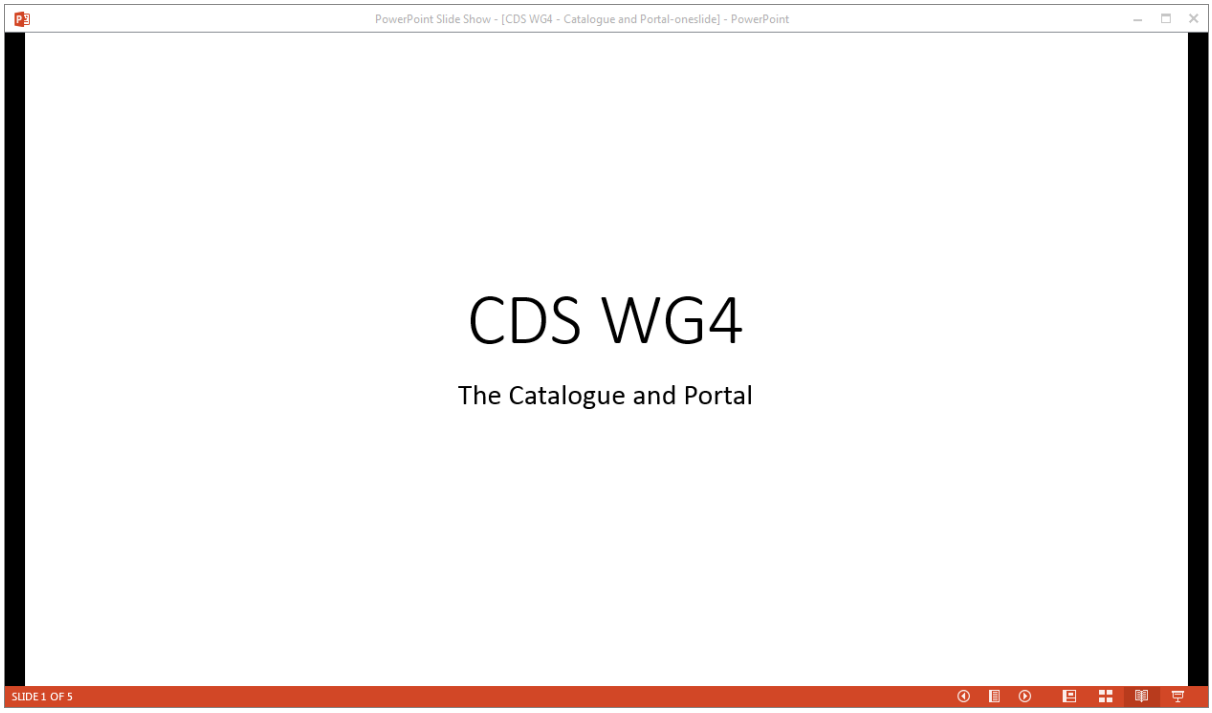
Working Group 3: Plenary Summary

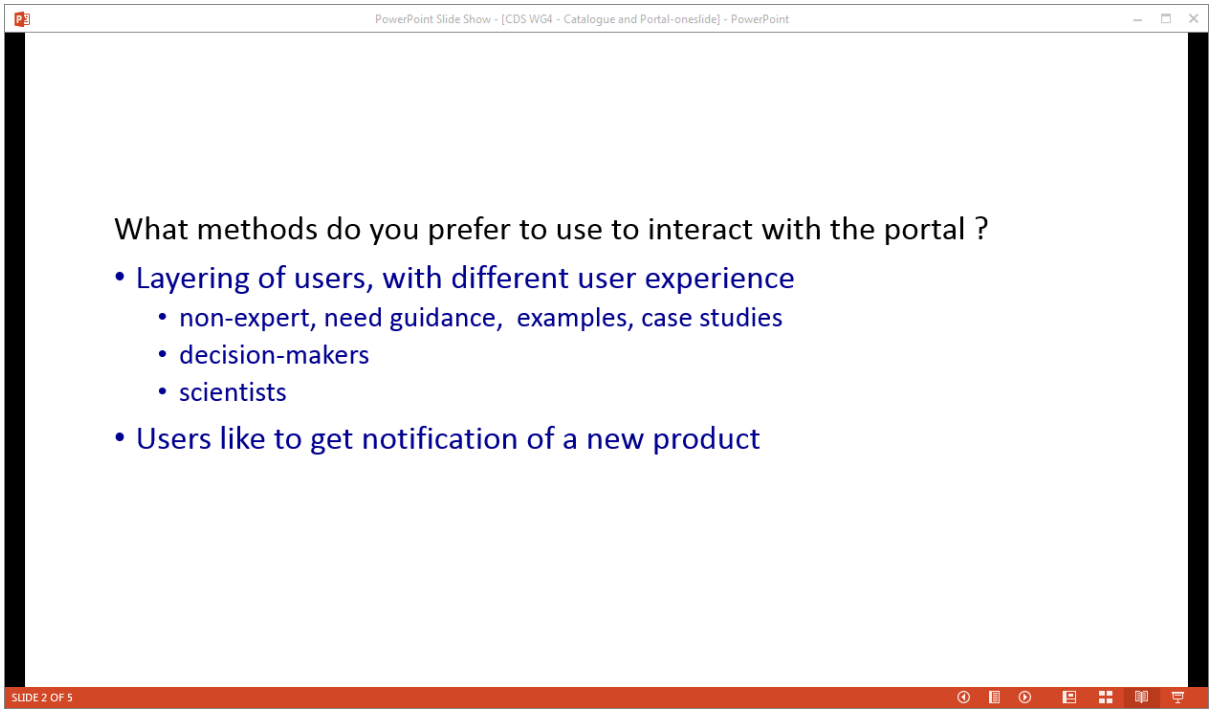


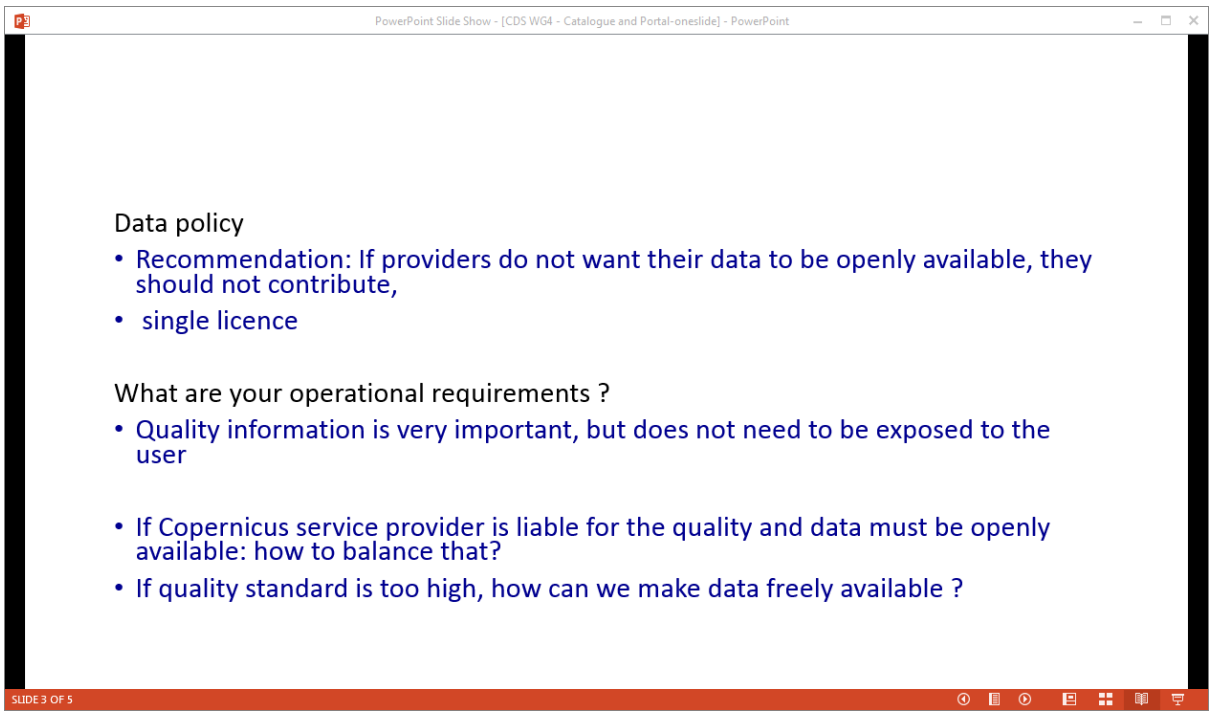


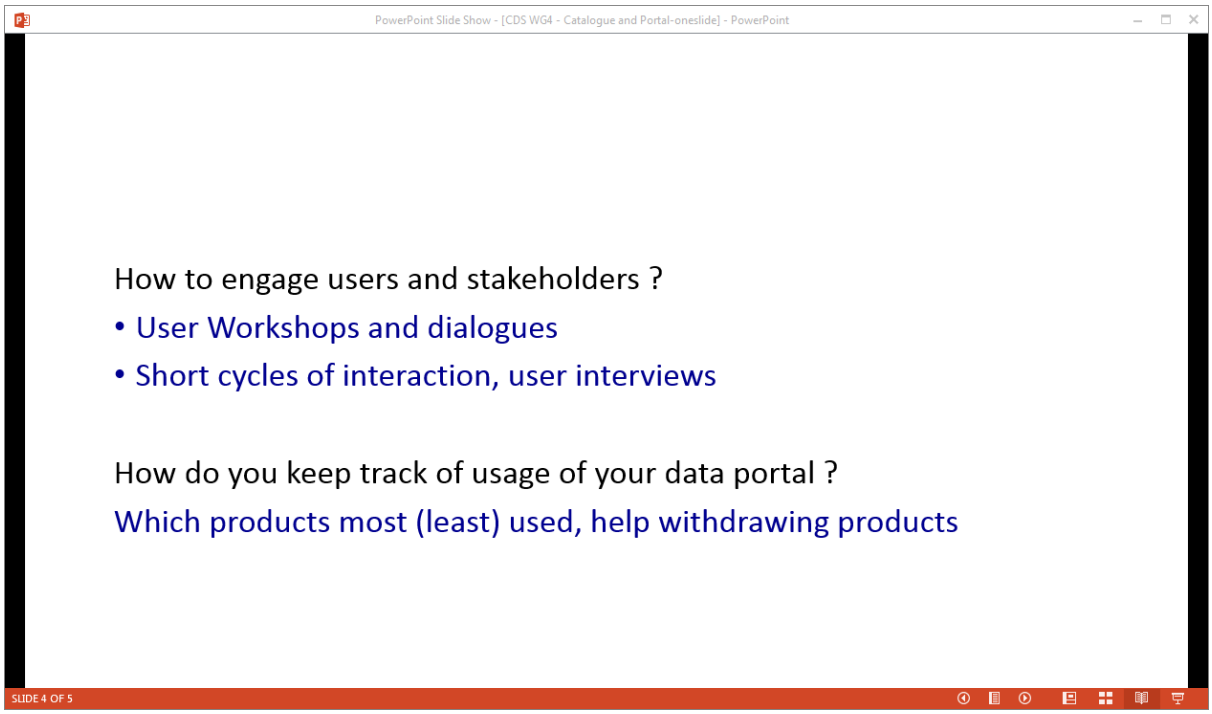

PowerPoint Slide Show - [WG3-Final-Report] - PowerPoint
WG3 –Standards and Interoperability
Content
Maps for quick-look discovery and comparisons
Keep content simple
Depends on the Users
Portrayal consistency
Quality information should be captured at a high level
Other types of content and capabilities
Keep the design flexible for socio-economic data systems in the future
Report should be in open formats (e.g. PDF) and described by metadata
"Find an expert" facility should be provided
Create an on-line Community
Standards and protocols
NetCDF-CF (ver. 1.6)
CSV (still required by Users)
OGC protocols (WxS)
UNIDATA protocols/tools
REST APIs (bulk download)
Interoperability
INSPIRE
WIS
WCRP
GEOSS
SLIDE 3 OF 3

Working Group 4: Plenary Summary






PowerPoint Slide Show - [CDS WG4 - Catalogue and Portal-oneslide] - PowerPoint
Data policy
• Recommendation: If providers do not want their data to be openly available, they should not contribute,
• single licence
What are your operational requirements ?
• Quality information is very important, but does not need to be exposed to the user
• If Copernicus service provider is liable for the quality and data must be openly available: how to balance that?
• If quality standard is too high, how can we make data freely available ?
SLIDE 3 OF 5


PowerPoint Slide Show - [CDS WG4 - Catalogue and Portal-oneslide] - PowerPoint
How to engage users and stakeholders ?
• User Workshops and dialogues
• Short cycles of interaction, user interviews
How do you keep track of usage of your data portal ?
Which products most (least) used, help withdrawing products
SLIDE 4 OF 5

PowerPoint Slide Show - [CDS WG4 - Catalogue and Portal-oneslide] - PowerPoint

How to improve your existing data portal

- A/B testing: Implement two “methods” of something, eg registration, and see which one is used most by users
- Task oriented with real users
- Make it easy to publish new datasets,
- Improve catalogue management
- Split user interface from data infrastructure

- Recommendation: Web data portal development using agile methods

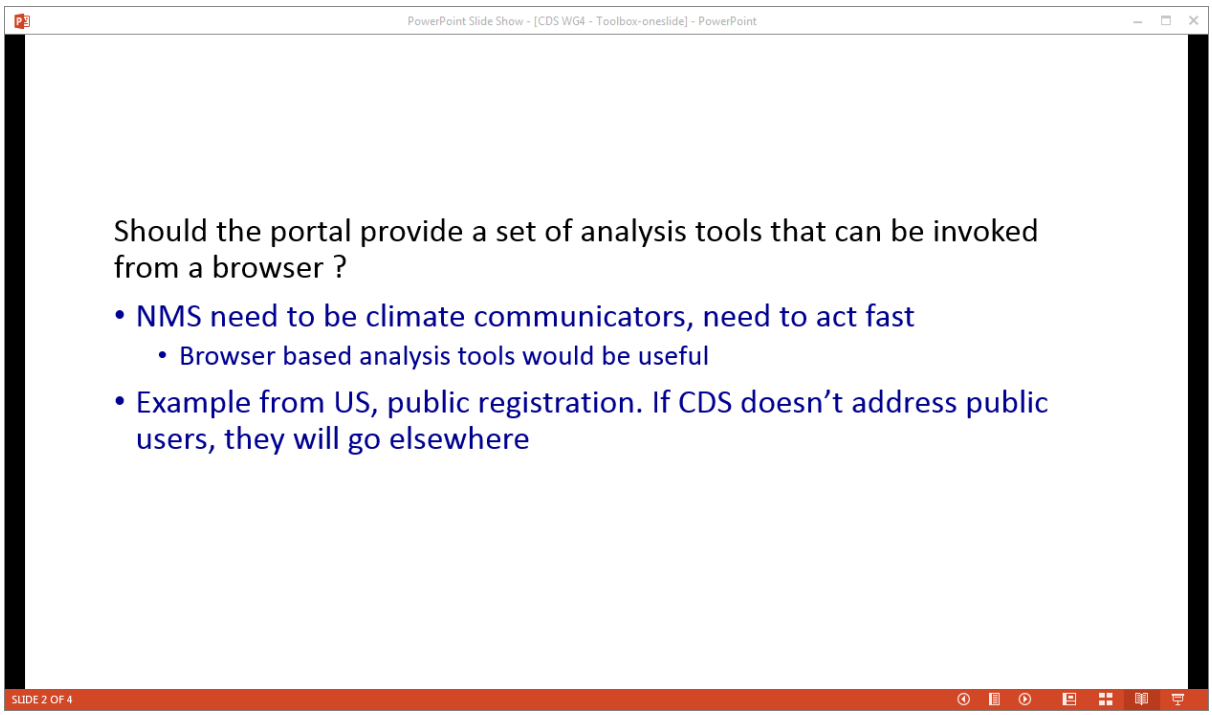
SLIDE 5 OF 5

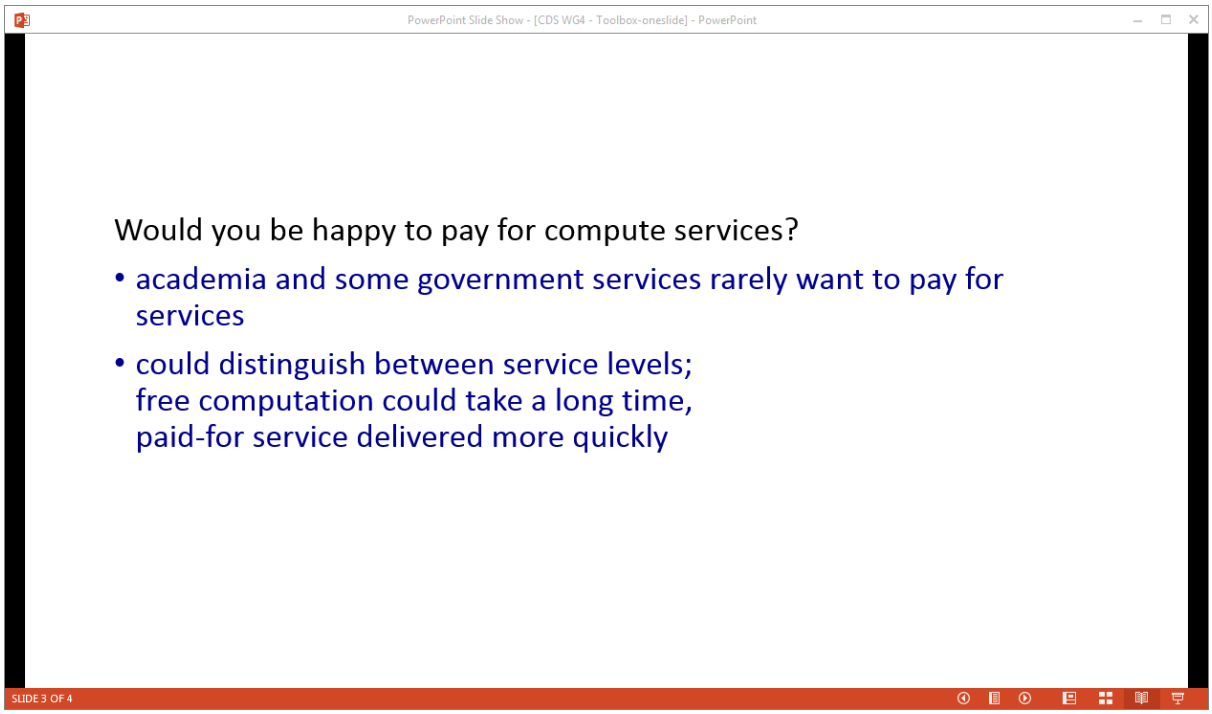
PowerPoint Slide Show - [CDS WG4 - Toolbox-oneslide] - PowerPoint

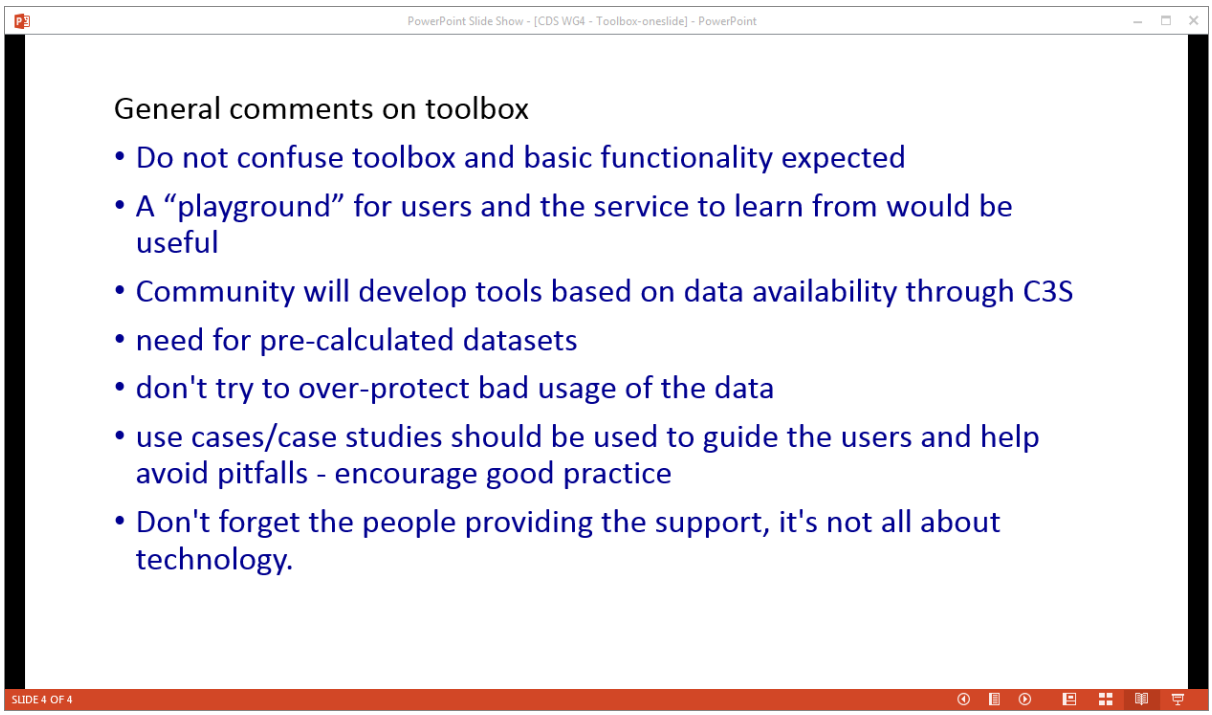
CDS WG4

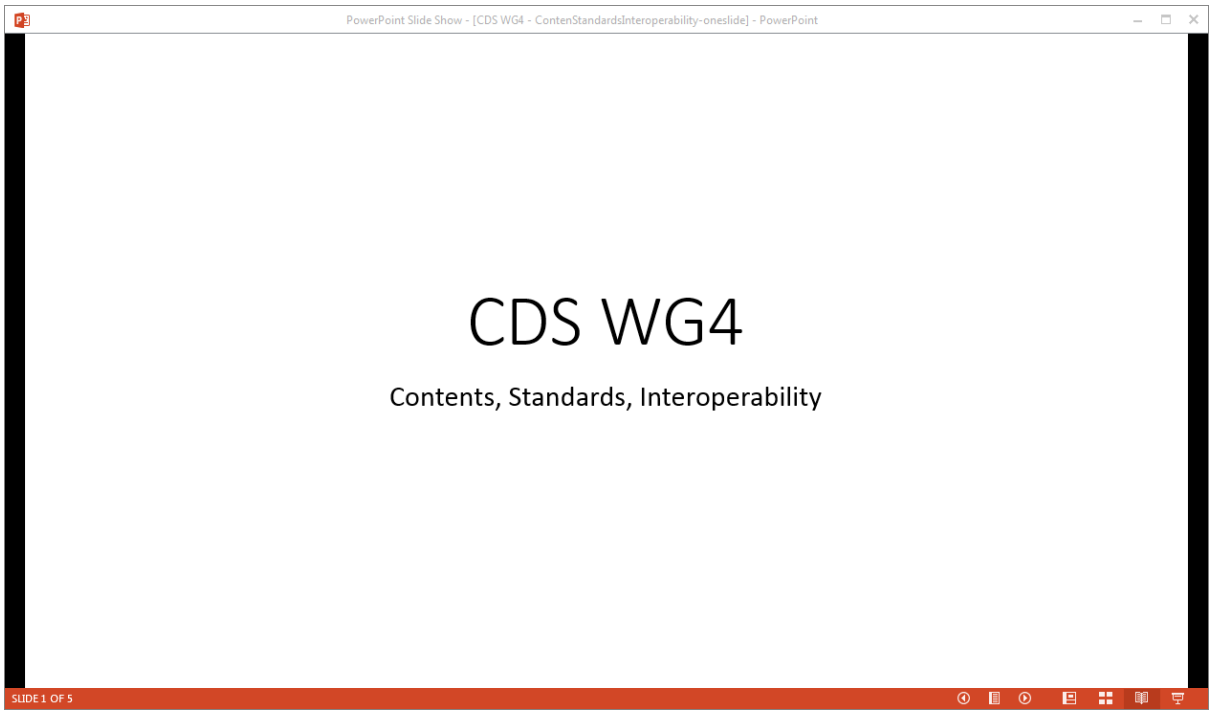
The Toolbox

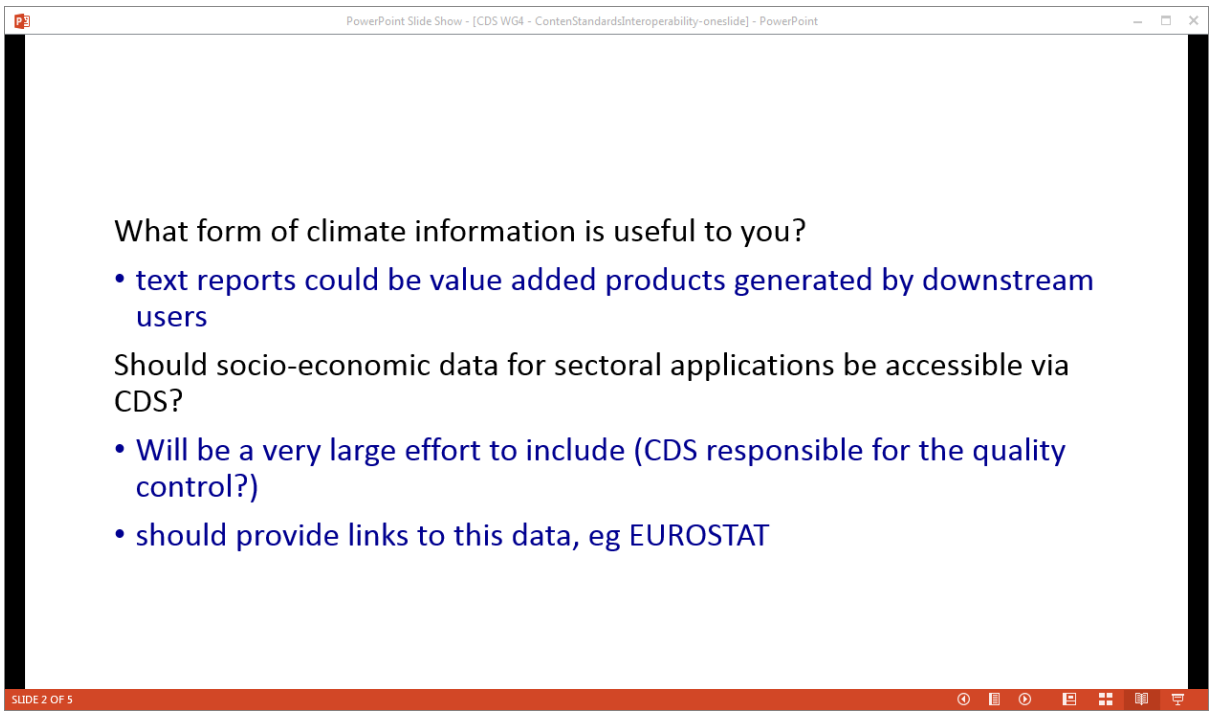
SLIDE 1 OF 4

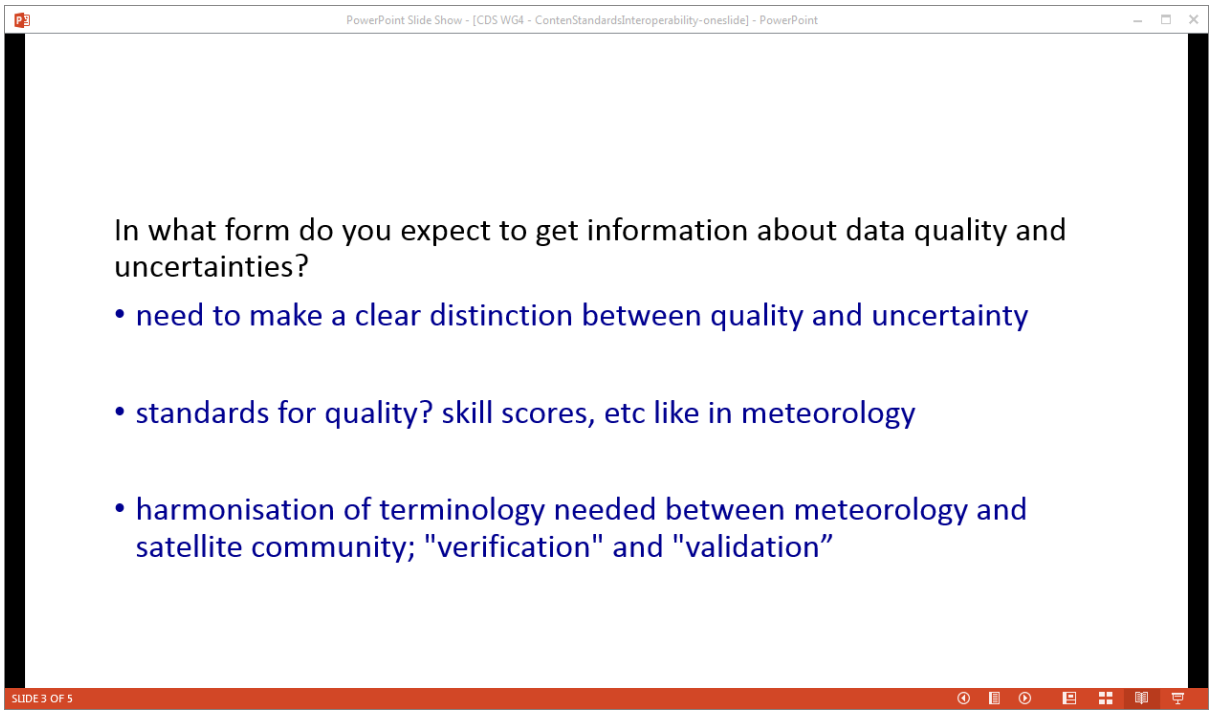

PowerPoint Slide Show - [CDS WG4 - Toolbox-oneslide] - PowerPoint
Should the portal provide a set of analysis tools that can be invoked from a browser ?
• NMS need to be climate communicators, need to act fast
• Browser based analysis tools would be useful
• Example from US, public registration. If CDS doesn't address public users, they will go elsewhere
SLIDE 2 OF 4

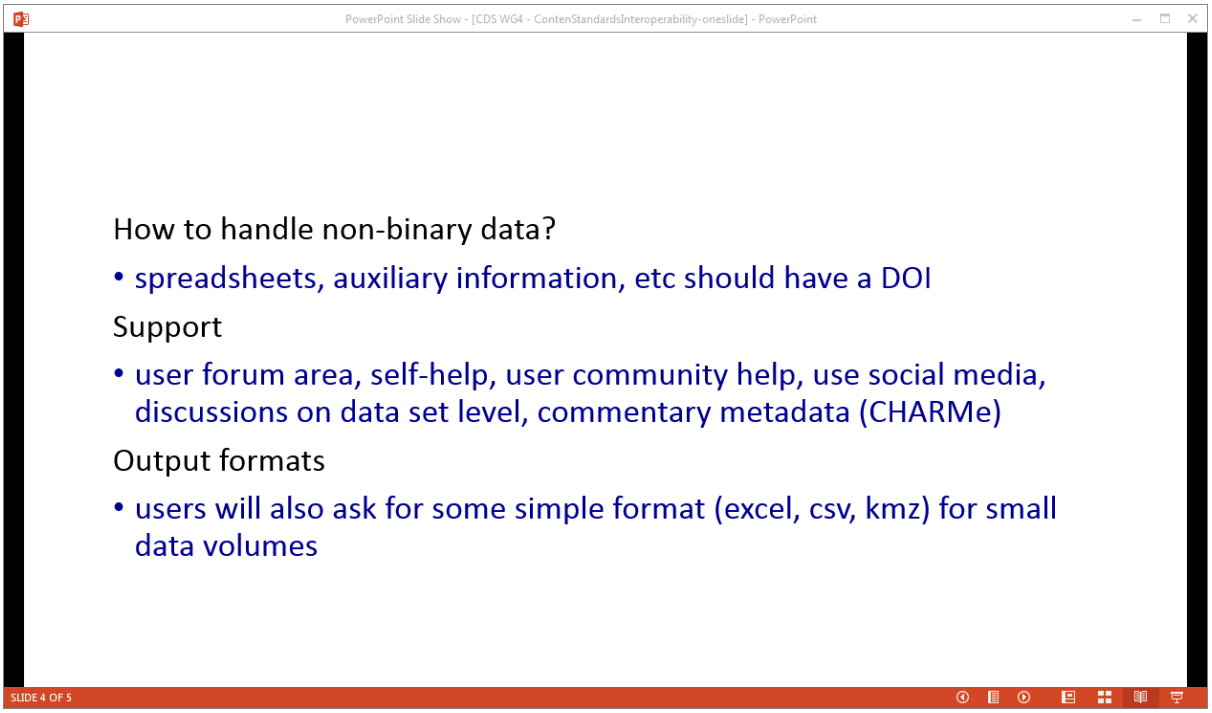

PowerPoint Slide Show - [CDS WG4 - Toolbox-oneslide] - PowerPoint
Would you be happy to pay for compute services?
• academia and some government services rarely want to pay for services
• could distinguish between service levels;
free computation could take a long time,
paid-for service delivered more quickly
SLIDE 3 OF 4

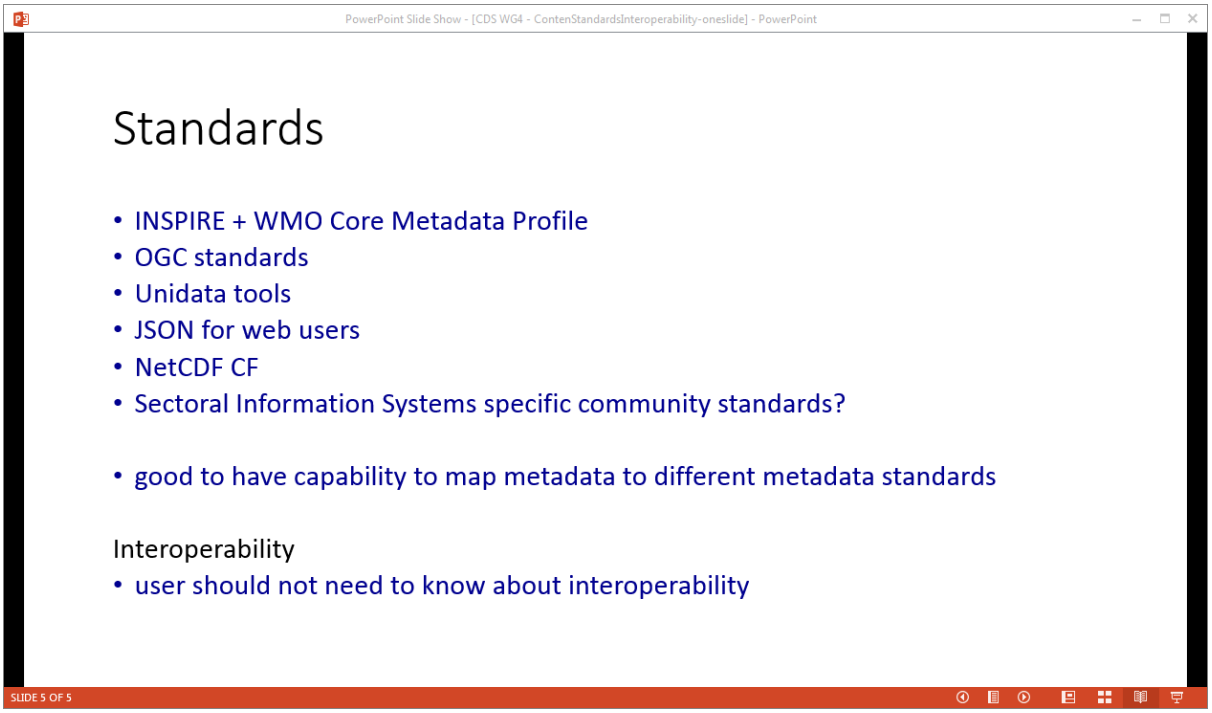

PowerPoint Slide Show - [CDS WG4 - Toolbox-oneslide] - PowerPoint
General comments on toolbox
• Do not confuse toolbox and basic functionality expected
• A “playground” for users and the service to learn from would be useful
• Community will develop tools based on data availability through C3S
• need for pre-calculated datasets
• don't try to over-protect bad usage of the data
• use cases/case studies should be used to guide the users and help avoid pitfalls - encourage good practice
• Don't forget the people providing the support, it's not all about technology.
SLIDE 4 OF 4


PowerPoint Slide Show - [CDS WG4 - ContenStandardsInteroperability-oneslide] - PowerPoint
CDS WG4
Contents, Standards, Interoperability
SLIDE 1 OF 5


PowerPoint Slide Show - [CDS WG4 - ContenStandardsInteroperability-oneslide] - PowerPoint
What form of climate information is useful to you?
• text reports could be value added products generated by downstream users
Should socio-economic data for sectoral applications be accessible via CDS?
• Will be a very large effort to include (CDS responsible for the quality control?)
• should provide links to this data, eg EUROSTAT
SLIDE 2 OF 5


PowerPoint Slide Show - [CDS WG4 - ContenStandardsInteroperability-oneslide] - PowerPoint
In what form do you expect to get information about data quality and uncertainties?
• need to make a clear distinction between quality and uncertainty
• standards for quality? skill scores, etc like in meteorology
• harmonisation of terminology needed between meteorology and satellite community; "verification" and "validation”
SLIDE 3 OF 5


PowerPoint Slide Show - [CDS WG4 - ContenStandardsInteroperability-oneslide] - PowerPoint
How to handle non-binary data?
• spreadsheets, auxiliary information, etc should have a DOI
Support
• user forum area, self-help, user community help, use social media, discussions on data set level, commentary metadata (CHARMe)
Output formats
• users will also ask for some simple format (excel, csv, kmz) for small data volumes
SLIDE 4 OF 5


PowerPoint Slide Show - [CDS WG4 - ContenStandardsInteroperability-oneslide] - PowerPoint
Standards
• INSPIRE + WMO Core Metadata Profile
• OGC standards
• Unidata tools
• JSON for web users
• NetCDF CF
• Sectoral Information Systems specific community standards?
• good to have capability to map metadata to different metadata standards
Interoperability
• user should not need to know about interoperability
SLIDE 5 OF 5