OECD's Statistical Information System Collaboration Community workshop 2012

Highlights report

We are pleased to present the OECD's Statistical Information system Collaboration Community workshop 2012 report.

Thank you to everyone who participated in the workshop over the week commencing 12 March 2012 and contributed to it's success.

Feedback was positive from those who completed the online survey and it will help in the planning for next years workshop. We thank you for your valuable feedback.

All presentations from the workshop have been made available on the workshop web site. In addition to the presentations there are a selection of videos available to SIS-CC member organisations. You can also find a selection of photographs taken during the afternoon of day 1 including the group photo as shown.

Key points

- 64 participates representing 25 organisations,
- Theme 'Enabling innovation': how sharing statistical information systems such as .Stat can be a leverage for innovation,
- Goal: increased collaboration opportunities,
- a total of 18 presentations over 3 innovation sessions,
- topics ranged from Open Data, Semantic Web, Mobility and Cloud,
- 5 groups per innovation session, 13 questions discussed and outcomes reported back,
- outcomes prioritised by collaboration partners on day 3,
- we welcomed UNESCO UIS and UoM to the community,
- demonstrated the value in capacity building through the secondment of an ABS staff member,
- further strengthen the position of .Stat as a enabler for other dissemination streams,
- and confirmed and validated the community workplan for the next 12 months.

SIS-CC workshop 2012 participates (Day 1 only)

Overview and outcomes

Day 1 Community network sessions
Objective: Understand the community network

Presentation topics:
- Community vision and key objectives
- Review of the .Stat solution and what has been delivered by OECD and contributions by community members
- What measures are being applied to build capacity within the community network
- Introduction of two new members
- Looking ahead and the roadmaps for OECD, community and a special presentation on the SDMX-RI

The opening and welcome speech was given by Paul Paul Schreyer, Deputy Director, OECD Statistics Directorate who repeated the message from the UN Statistics that all NSO's are facing some common challenges. Paul also highlighted SDMX standard as an important mechanism for facilitating for easier data exchange and standardisation across the statistical community.

A clearly defined vision and framework is a fundamental elment to support the community activities. Lester Rodriques, Head of Information Technology and Network Services presented the vision following on from its validation after the 2011 workshop. Lester enforced the message of the importance of collaboration initiates in this time when NSO's are facing such challenges that further strengthens the case for the community approach being taken by the SIS-CC.

The community is demonstrating that collaboration can work with a number of successful deliveries over the last 12 months, as presented through the 2011/12 review. Major developments included the .Stat browser redesign including the integration of an advanced visualisation component eXplorer, a feature for managing sparse datasets with fully-fill page, unit management, and a number of security enhancements to ensure the .Stat solution remains ahead of new threats and vulnerabilities. In addition to these core developments there were a number of tools and applications that leveraged .Stat including a tool for managing structural metadata and iPad app as presented by the IMF.

Capacity building is critical for the SIS-CC to sustain in the long term and continue to support the developments across the community. A brief overview was given of how member organisations participate in the development work, in a spirit of co-production, and within the broader context of SIS-CC governance framework. This was followed by a case study of the recent secondment to the OECD from ABS to build capacity in developments. This confirmed secondments are needed to build capacity and provide the necessary environment for developing core components. The collaborative environment relies on the community working effectively together to increase cooperation, increase knowledge, build relationships and understanding along with increase staff development.

The 2012 workshop saw the welcoming of two new organisations to the community and a chance for them to present their business case for joining. This further demonstrated how organisations can meet the challenges of today by becoming part of an initiate such as the SIS-CC with key criteria including meets business requirements and strategic directions, budget restrictions, not to reinvent the wheel, community started to deliver, facilitating innovation, leveraging other initiatives i.e. SDMX, .Stat a tool for aggregating data for vertical domains, agile approach to developments, and strong visible project governance.

In the final part of the community network session the global roadmap of the OECD and community was presented along with a special presentation on the planned integration of the SDMX-RI within .Stat.

Day 1-2 Innovation sessions
Objective: Share knowledge, build expertise, and brainstorm collectively

Main topics covered:
- Data Machine-Readability in support of open data strategies
- Data Discovery and the potential of semantic web
- Exploring new technologies for dissemination

Three sessions over 1 ½ days covering areas from open data, standards such as SDMX, DPSL and JSON, common api's, linked data and semantic web, search, mobility, HTML, SVG, cloud and web accessibility. Each session concluded with group work to discuss, answer set questions, and propose a way forward for the community on the given topic.

The first session concentrated on data machine-readability. .Stat already features data machine-readabilitlity functions but with the need to evolve to support the ambitious Open Data strategies of SIS-CC members. The focus of this session was aimed at directing the community towards the right decisions as to what standards and which implementations to prioritise. It began with a presentation by Xavier Badosa from Idescat, entitled Standards for statistical data dissemination: a wish list. Interesting and thought provoking, Xavier looked at how when standards are made simple it can lead to greater opportunities for reuse across the development community. JSON was introduced as a means to facilitate this as it is much more natural to programmers than XML.

See Xavier's full presentation here and a video of the talk can be viewed here

The presentations that followed continued with the theme for supporting open data strategies and the need to introduce mechanisms to facilitate this. The session also look at real working examples of how this can be applied including how Vincenzo Patruno from ISTAT has developed JSON based api's connected to .Stat, and Omar Benjelloun from Google presented the DPSL standard that is behind the Google Public Data web application enabling wider use of data. Even with this drive to introduce standards that are more developer focused it was clear from the presentations and discussions that followed that SDMX still has a strong part to play and can be a leverage for this.
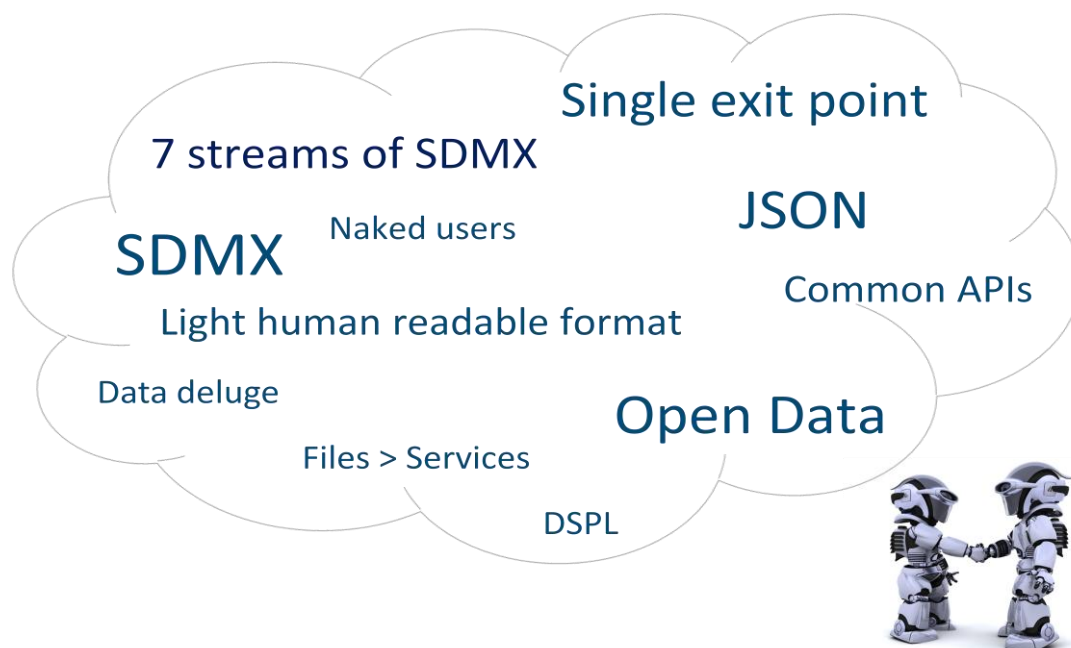
Work group highlights

Q. What standard should be implemented to facilitate open data initiatives including plug and play developments?

- SDMX-ML – not a standard for Open Data but important for exchange of data between organisations
- Odata – implemented in most MS products
- JSON implementation – use cases inc. HTML5, mobility, APIs, easier understandable by developers, con: no validation
- DSPL – based on xml and CSV. All data referenced and easily searchable, simple and great for visualisations
- SDMX too verbose and too complex for open data initiatives
- DDI: UC microdata for reasearchers, complex standards

Q. Should .Stat be a universal data explorer for other data sources?

Identifying the need
- Types of data use:
    - Structured: SDMX, DDI, google (DSPL), etc.
    - Unstructured: much of the open data;
- Eliminate duplication of datasets – always refer to the reference dataset from the data producer will provide users with the highest quality data:
    - no loss of or changes to metadata as datasets are re-processed;
    - Users will have access to most current data;
    - Users will have increased trust in data – using reference source
- Provide users with rich selection of datasets for analysis;
- Help users find most relevant dataset "fit for purpose."

Examples of universal data explorers already available
- IMF presented their generic SDMX browser (flat file SDMX files & registry referenced SDMX files);
- Google presented Public Data Explorer for DPSL files;
- Generic:
    - Wolfram Alpha
    - Quicky
- Non-statistics:
    - Scrapers:
        - Shopping sites: pricegrabber.com (finds low priced electronics)
        - Travel/tourism sites
        - News sites

User benefits
- Single gui/tool to learn;
- Accessing reference data source means users will have immediate access to data revisions/updates;
- With effective search tool, users will find most appropriate datasets for their needs;
- Introduce data transparency, improve quality by increasing ability to compare/contrast datasets

**What is required to achieve this, at business and technical levels**
- To work, data needs to be "available", "findable", "consumable", "comparable"
- Implement a "data search" engine as the entry point to a universal data browser;
- Modify .Stat explorer to view an SDMX dataset located on an external site;
- Extend .Stat explorer to support comparing datasets
- Extend the search engine & .Stat explorer to support new standards, DDI, GSDL, and restrict comparisons to files of same format, at least initially.

The second session focused on the converging worlds of semantic web and statistical metadata, in a way that should enable new ways of linking data and content. Very much an exploratory session looking at the potential value add this convergence represents and the directions for future implementations in .Stat.

Paul Murphy from the University of Manchester opened the session with a presentation entitled 'The Missing Link: Giving Statistical Data Meaning'. The UK has very much been leading the way in regard to linked data and despite being years since it was first discussed a clear business case is yet to be defined. Paul presented some exploratory examples of what Mimas have been doing at the University and the potential linked data could bring if done right.

See Paul's full presentation here and a video of the talk can be viewed here

Stéphane Varin from the OECD gave a further presentation on linked data and linking narrative content. Chris Nelson, Metadata technology talked about RDF technologies and presented the BIS sandbox, a SDMX based data portal. The session then moved onto search and the work to introduce a single point of search integrating all dissemination channels across an enterprise. This included a presentation of a PoC by Tony Breen from SNZ as well as how work is underway to introduce a connection between a 3rd party engine and .Stat to allow for this enterprise wide search capability.
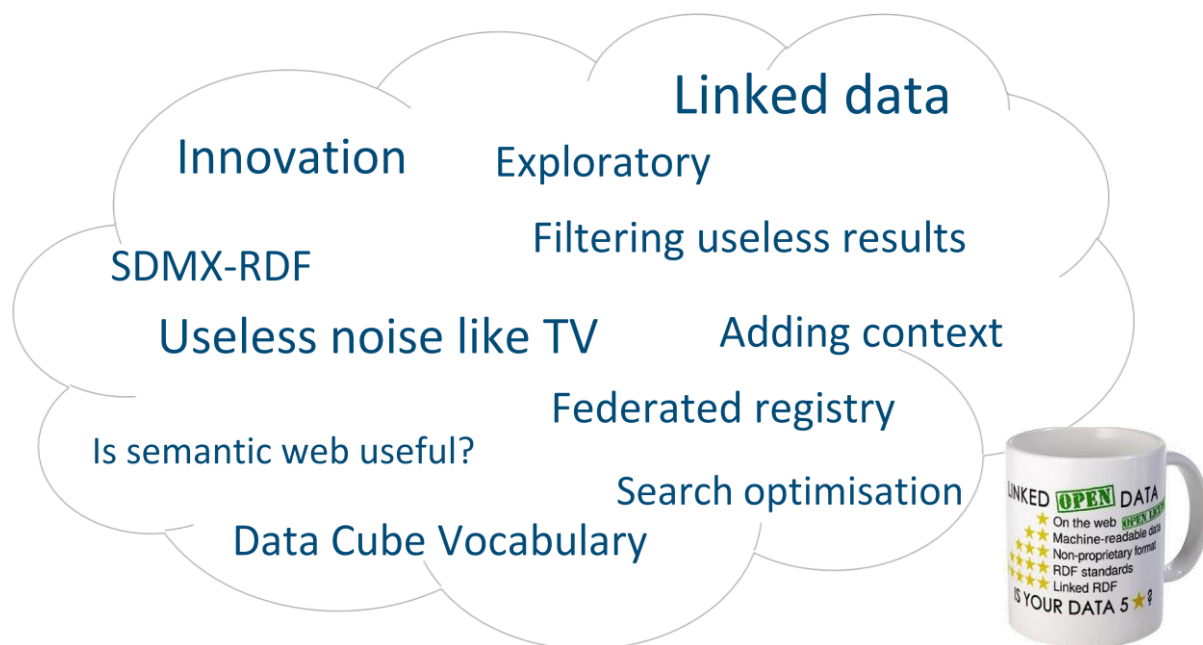
Work group highlights

Q. What are the scenarios for SDMX linked data implementation and what are the alternatives?
- Implement a data cube vocabulary
- Convert from SDMX to RDF data vocabulary
- Triples store
- Not many tools available to support this
- Determine the usefulness of this
- Federated registry

Q. Discuss ways to improve search through semantic enrichment of .Stat
- Propose related links and information
- Identify target groups
- Provide exact result with link to discover more
- Improve discovery of data
- Learn from failure of results i.e. misspelt words
- Search subscription service
- Index visualisations

The final innovation session covered Mobility, Cloud, HTML5 and how multiple radical technology evolution are driving the market, calling for renewed design of dissemination systems. Whether it is at the level of the application interface or more broadly, application architecture is fast moving and it is clear we need to be prepared to key pace. This session was all about assessing the value add that .Stat could bring in this new context and identifying SIS-CC priorities

Publishing interactive statistics visualisation for mobile devices using HTML5 by Patrik Lundblad from NComVA AB, provided the lead into this session showing how they are adapting to this ever changing environment with the introduction of HTML5 based vislets.

See Patrik's full presentation here and a video of the talk can be viewed here

This session gave many more practical examples than the previous ones and included a well received presentation by Eric Dery from IMF entitled Statistics On the GO, a mobile application built in just 10 weeks that utilised .Stat behind. There was much interest on sharing this application and highlighted the need to think about this from the concept phase of any developments.

The focus then moved onto the cloud and the potential it has to offer with a need to reduce costs but maintain the same level of support and service to our users. Jens Dossé from OECD highlighted the key criteria that a cloud based solution would need to support statistical processes and needs.

The final presentation of the innovation session turned to web accessibility with Merry Branson from ABS providing some insights in what is required to meet the AA standard especially in this fast moving technological world. Innovation is sure to play a key role in helping achieve this.
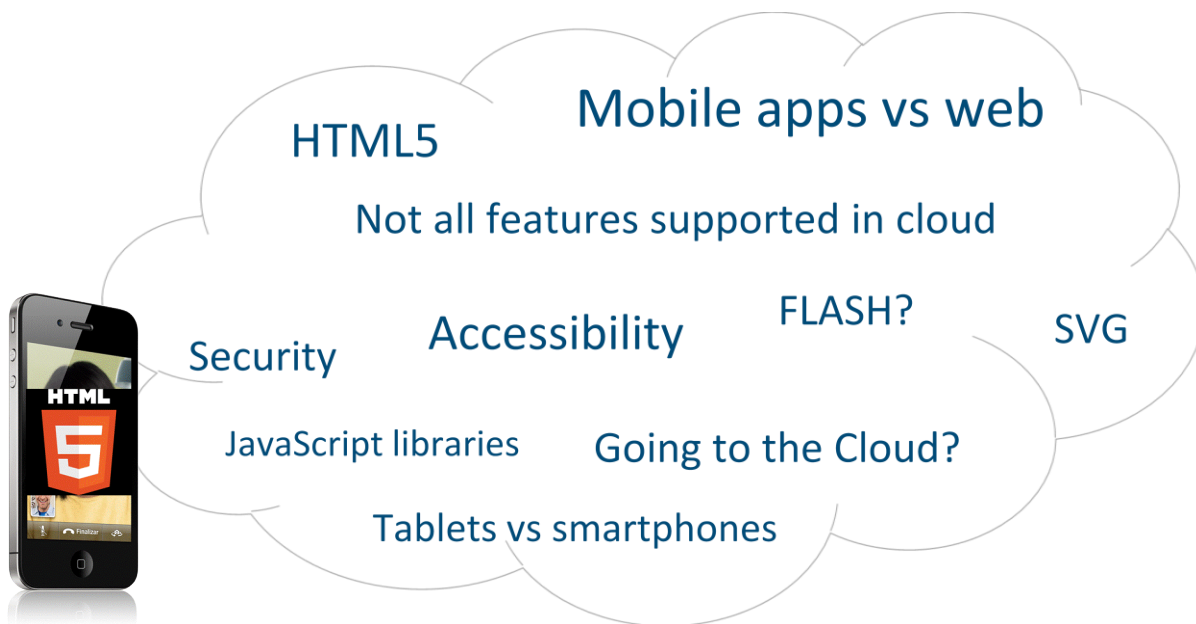
Q. What technology should be the focus to provide graphical features within .Stat, and plug and play applications?
- Flash is mature with rich features
- Flash is not mobile friendly and properiatary format
- HTML5 (Not mature) and JavaScript (Mature) with rich features
- Browser support for HTML5 is poor today but future looks good
- SVG mature open standard. Peformance can be questionable on mobile devices though

Q. What is needed to enable .Stat to support a cloud architecture, and is the cloud ready for statistical processes?
- Disseminate public data in the cloud
- An ETL that allows for a level of control on what can be published with controlled access
- Concerns include location specific regulation and service provider solidity

Full list of presentations and available videos

Presentation	Presenter	Organisation	Session
2011 achievements.ppsx	Jens Dossé	OECD	Community activities
Community vision and key objectives.pptx	Lester Rodriques	OECD	Community activities
Case study of the recent secondment to the OECD from ABS.ppt	Merry Branson	ABS	Community activities
Community capacity building and process improvements.ppsx	Jonathan Challener	OECD	Community activities
1.Workplan 2012.pptx	Jens Dossé	OECD	Community activities
2.OECD_SDMX-RI.ppt	Bengt-Ake Lindblad	EuroStat	Community activities
3.SDMX-RI applied to .stat.ppt	Francesco Rizzo	ISTAT	Community activities
1.SIS-CC UNESCO UIS Business Case.ppt	Brian Buffett	UNESCO	Community activities
2.SIS-CC Mimas Business Case.pptx	Paul Murphy	University of Manchester (Mimas)	Community activities
1.The Missing Link - Giving Statistical Data Meaning.pptx Video: The Missing Link - Giving Statistical Data Meaning - Highres.wmv	Paul Murphy	University of Manchester (Mimas)	Data Discovery
2.Linking Data And Narrative Content Through Semantic Web Technologies.pptx Video: Linking data and narrative content through semantic web.wmv	Stéphane Varin	OECD	Data Discovery
3.OECD_Embracing RDF-Based Technologies for Statistical Data.pptx Video: Embracing RDF-based technologies for Statistical data.wmv	Arofan Gregory	Open Data Foundation	Data Discovery
4.Building a Data Portal with SDMX - Sandbox.ppt Video: Building a Data Portal with SDMX.wmv	Chris Nelson	Metadata Technology	Data Discovery
5.Single point of search for dissemination tools.pptx Video: Single point of search for dissemination tools.wmv	Tony Breen	SNZ	Data Discovery
6.Improving the search capabilities of dotStatnew.pptx Video: Improving the search capabilities of .Stat.wmv	Casper Meyer	OECD	Data Discovery
1.xbadosa-standards4statdatadissemination.pptx Video: Standards for statistical data dissemination a wish list.wmv	Xavier Badosa	Idescat	Data Machine-Readability
2.Connecting with .Stat through APIs.pptx	Vincenzo Patruno	ISTAT	Data Machine-Readability
3.DSPLOpenData.pdf	Omar Benjelloun	Google	Data Machine-Readability
4.Building Systems and Tools with Common Components.ppt	Chris Nelson	Metadata Technology	Data Machine-Readability
5.OECD_DDI_SDMX.ppt.pptx	Arofan Gregory	Open Data Foundation	Data Machine-Readability
7.Interfacing to SDMX-RI.ppt	Francesco Rizzo	ISTAT	Data Machine-Readability
2.Statistics on the GO.pptx Video: Statistics On the GO.wmv	Eric Dery	IMF	Web Dissemination
3.Using SVG potential for graphical representation of statistical data.pptx	Henri Kayali	OECD	Web Dissemination
4.Intro to Shared-Services.pptx	Ruchikar Dalela	TCS	Web Dissemination
5.Joining the cloud formation - Suggested architecture.pptx Video: Joining the cloud formation.wmv	Jens Dossé	OECD	Web Dissemination
1.NComVA March 2012 - OECD.pptx Video: Publishing interactive statistics visualisation for mobile.wmv	Patrik Lundblad	NComVA AB	Web Dissemination

Final list of participates

Mr	ZIVOKIVIC	Andy	Australian Bureau of Statistics
Ms	BRANSON	Merry	Australian Bureau of Statistics
Mr	HOWARTH	Rob	Australian Bureau of Statistics
Mr	BADOSA	Xavier	Statitistical Institute of Catalonia
Mrs	SILLAJÕE	Tuulikki	Statistics Estonia
Mr	CHROSCICKI	Marek	European Commission
Mr	WRONSKI	Adam	Eurostat
Mr	LINDBLAD	Bengt-Åke	Eurostat
Mrs	KIOMALL	Maria	Eurostat
Mr	HUTTUNEN	Markku	Statistics Finland
Mr	MEDLA	Michael	Ifo Institute
Mr	SALOU	Gerard	International Monetary Fund
Mr	ZHU	Gangti	International Monetary Fund
Mr	DERY	Eric	International Monetary Fund
Mr	PICHE	René	International Monetary Fund
Mr	CORONADO IRUEGAS	Abel	INEGI
Mr	BONNAUD	Jean-Pierre	National Institute of Statistics and Economic Studies
Mr	YON	Christophe	National Institute of Statistics and Economic Studies
Mr	HENNEQUIN	Bernad	National Institute of Statistics and Economic Studies
Mr	DUFFES	Guillaume	National Institute of Statistics and Economic Studies
Mr	PATRUNO	Vincenzo	Italian National Institute of Statistics
Mr	RIZZO	Francesco	Italian National Institute of Statistics
Mr	CAMOL	Dario	Italian National Institute of Statistics
Mr	SCALZO	Domenico	Italian National Institute of Statistics
Mr	DE FRANCISCI	Stefano	Italian National Institute of Statistics
Ms	BENETTI	Laura	Banca d'Italia
Dr	BALZERANI	Luca	Banca d'Italia
Ms	MIYAKOSHI	Mitsuyo	Ministry of Internal Affairs and Communications
Ms	AKUTSU	Fumika	Ministry of Internal Affairs and Communications
Mr	ŠPEH	Tomaž	Statistical Office Slovenia
Mr	DEVLIN	Jay	Statistics New Zealand
Mr	BREEN	Tony	Statistics New Zealand
Mr	BROOK	Jonathan	Statistics New Zealand
Mr	MORILLO GALVEZ	Carlos	Spanish National Institute of Statistics
Mr	GROSSENBACHER	Armin	Swiss Federal Statistical Office
Mr	BUFFETT	Brian	UNESCO Institute for Statistics
Mr	ABBAS	Duraid	UNESCO Institute for Statistics
Mr	DYMOND-GREEN	Rob	University of Manchester (Mimas)
Mr	MURPHY	Paul	University of Manchester (Mimas)
Mr	SYROTIUK	Nick	University of Manchester (Mimas)
Mr	BIANCHI	Ron	US Department of Agriculture

Mr	NELSON	Chris	Metadata Technology Ltd
Mr	NELSON	Matt	Metadata Technology Ltd
Mr	BENJELLOUN	Omar	Google
Mr	LUNDBLAD	Patrik	NComVA AB
Mr	DALELA	Ruchikar	Tata Consultancy Services
Mr	KAYALI	Henri	Info3
Mr	RODRIQUES	Lester	OECD
Mr	ANVAR	Eric	OECD
Mr	DOSSÉ	Jens	OECD
Mr	CHALLENER	Jonathan	OECD
Mr	MEYER	Casper	OECD
Mr	RIVIERE	Bertrand	OECD
Mr	KAARTINEN	Petrus	OECD
Mr	SCHREYER	Paul	OECD
Mr	FLETCHER	Trevor	OECD
Mr	BARRACLOUGH	David	OECD
Ms	MITTON	Terri	OECD
Mr	VARIN	Stephane	OECD