

The logo for the Wolfram Technology Conference 2004 is displayed on a blue background with a grid of white squares. The text "WOLFRAM" is in white, "TECHNOLOGY" is in orange, and "CONFERENCE 2004" is in white.

WOLFRAM TECHNOLOGY CONFERENCE 2004

Navigating the Data Universe: The Future of Import /Export

*Dale R. Horton, Mike Pilat, Flavio Smirne
Software Technology Group
Wolfram Research, Inc.*

New in *Mathematica* 5.1 and 6.0:

Scientific Data, Spreadsheets, Databases, 2D and 3D Graphics, Multimedia, Bioinformatics,
Geographic Information Systems, Document Formats, and much more ...

Initialization

WOLFRAM TECHNOLOGY CONFERENCE 2004

Overview of Import and Export

There are two main ways to transfer data into and out of *Mathematica*.

- *MathLink* (including J/Link, .NET/Link, Database Access, Web Services)
- File/Stream operations (including Import /Export, Read /Write)

This makes Import /Export a major component allowing *Mathematica* to interact with the "outside world". Some of the many uses include:

- Transferring data to or from other applications.
- Manipulate and process data to generate reports.
- Publishing documents. (HTML, $\text{T}_{\text{E}}\text{X}$, XML, etc.)
- Generating web*Mathematica* graphics and typesetting.
- Manipulating raster graphics (Digital Image Processing).

Example

WOLFRAM TECHNOLOGY CONFERENCE 2004

Data Formats

XLS: Microsoft Excel spreadsheet format
DIF: Lotus Data Interchange format
HDF5: NCSA Hierarchical Data Format
MAT5: Matlab matrix format
DBF: dBase database format
Table, TSV, CSV: ASCII table formats (speedups)

Others in *Mathematica* 5.0: FITS, HarwellBoeing, HDF, MTX

XLS Example

HDF5 Example

WOLFRAM TECHNOLOGY CONFERENCE 2004

Graphics

SVG: Scalable Vector Graphics

PCX: ZSoft PC Paintbrush format

TGA: Truevision Targa bitmap format

Support for Transparency

Others in *Mathematica* 5.0: BMP, DICOM, EPS, GIF, JPEG, PNM, PNG, PICT, TIFF, WMF, XBitmap

Transparency

WOLFRAM TECHNOLOGY CONFERENCE 2004

3D Graphics

3DS: Discreet 3D Studio MAX scene format
BYU: Brigham Young University format
JVX: JavaView XML format
Maya: Alias Maya ASCII scene format
OBJ: Wavefront OBJ format
OFF: Geomview object file format
PLY: Stanford polygon format
POV: POV-Ray tracing format
RIB: Pixar Renderman entity format
X3D: Web3D XML format
DXF: AutoCAD drawing interchange format (improvements)
STL: Stereolithography format (improvements)

PLY

OBJ

WOLFRAM TECHNOLOGY CONFERENCE 2004

Multimedia Formats

MIDI: Musical Instrument Digital Interchange

MusicXML: Musical data XML format

MP3: MPEG-3 Audio format

OGG/Vorbis: Xiph.org audio codec (in OGG transport)

OGG/Theora: Xiph.org video codec (in OGG transport)

FLAC: Free Lossless Audio Codec

AVI: Audio Video Interleave format

Others in *Mathematica* 5.0: AIFF, Animated GIF, AU, SND, WAV

AVI Example

A blue banner with a grid pattern and the text "WOLFRAM TECHNOLOGY CONFERENCE 2004" in white and yellow.

WOLFRAM TECHNOLOGY CONFERENCE 2004

Bioinformatics: Genomic Sequence Formats

EMBL: EBI nucleotide database

FASTA: NCBI Fast Alignment Search Tools Anything

GenBank: Genetic Sequence Data Bank

Simple Cistron Data Structure (FASTA File)

More Complex Cistron Data Structure (GenBank File)

WOLFRAM TECHNOLOGY CONFERENCE 2004

Geographic Information Systems

MIF/MID: MapInfo

SHP/DBF: ESRI Shapefile Bundle

E00: ArcInfo Export

GeoTIFF: Geographical TIFF

TIGER: Topologically Integrated Geographic Encoding and Referencing

SDTS: USGS Spatial Data Transfer Standard (improvements)

Support for GPS data

School Districts – Elementary (DuPage County – Illinois)

Urban Areas (Champaign County – Illinois)

2000

City Map

Rivers

Land

Maur (3 Layers)

Pont (3 Layers)

SDTS Data Visualisation

WOLFRAM TECHNOLOGY CONFERENCE 2004

Document Formats

PDF: Portable Document Format

RTF: Rich Text Format

RSS: Really Simple Syndication XML format

TeX: TeX typesetting language (improvements)

Others in *Mathematica* 5.0: HTML, MathML, NB, NotebookML, XHTML+MathML, XML

PDF

TeX

WOLFRAM TECHNOLOGY CONFERENCE 2004

Infrastructure

Very fast binary i/o operations.

Support for Importing from URLS. (HTTP, FTP)

Compression and Archiving. (GZIP, TAR, ZIP)

Access to file format elements like raw data or metainformation.

Visualization and Data Manipulation Tools.

Integration with Connectivity Tools. (WebServices, Database Access, etc.)

Binary I/O

URL/Compression

File Format Elements

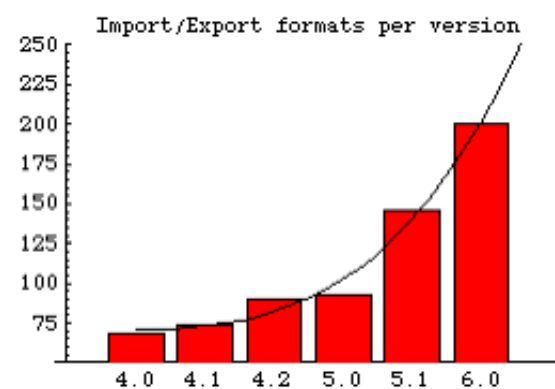
WOLFRAM TECHNOLOGY CONFERENCE 2004

Navigating the Data Universe ...

As demonstrated above, we have a strong commitment to improving and expanding Import /Export into the future.

Not just by adding new formats, but also by adding new features and fundamental tools.

The growth of formats and features has been impressive, and will continue to be so.



But we're fully aware that there's much more work to do. You can help us identify new areas to pursue.

We'd like to hear any suggestions you might have. If you don't get a chance to discuss them with us, you can send them to support@wolfram.com.

A blue banner with a grid pattern and the text "WOLFRAM TECHNOLOGY CONFERENCE 2004" in white and yellow.

WOLFRAM TECHNOLOGY CONFERENCE 2004

Appendix: New in 5.1

New Formats:

XLS, DIF, HDF5, MAT5

PCX

AVI

Binary Formats (Bit, Byte, Integer16, Real32, etc.)

Improved formats:

TeX

Table, TSV, CSV (speedups)

Binary I/O (BinaryWrite, BinaryRead, BinaryReadList)

New infrastructure:

URL Support for Import

GZIP Compression